

COMMENTARY

質量分析法と配列データベースを利用する
タンパク質同定法Mass Spectrometry-Based Protein Identification
by Correlation with Sequence Database吉野健一^{a), b)}・大城紀子^{a), b)}・徳永千春^{a), b)}・米澤一仁^{a), b)}
Ken-ichi YOSHINO, Noriko OSHIRO, Chiharu TOKUNAGA,
and Kazuyoshi YONEZAWA

(Received April 9, 2004; Accepted April 13, 2004)

The progress in genome sequencing projects of a large number of organisms and the advance in mass spectrometry of protein analysis have been significant driving forces in the formation of the field proteomics. The advance in protein mass spectrometry includes development of computer algorithms that use mass spectrometric data to identify proteins in sequence databases. The computer algorithms for protein identification by correlation with sequence databases rely on the availability of constraining parameters that distinguish specific matches from all the other sequences in the database. They can be categorized into three strategies. One of strategies is called "peptide mass fingerprinting," which is based upon correlating measured masses of peptides derived from digestion of proteins by a residue-specific protease with theoretically calculated peptide masses derived from proteins registered in sequence database. Two strategies for protein identification using tandem mass spectrometry (MS/MS) data are distinguished by demand for interpretation of product ion mass spectra. Product ion mass fingerprinting using uninterpreted MS/MS data of peptides is conceptually similar approach to peptide mass fingerprinting. SEQUEST[®] and MS/MS Ions Search in MASCOT[®] are the most widely used algorithms for protein identification by searching sequence database using uninterpreted product ion mass spectra. Another strategy using MS/MS data employs the search algorithm by using parameters, such as "peptide sequence tag," found by manual inspection of product ion mass spectra.

1. はじめに

1995年7月、*Science*誌に“Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd”と題した1編の論文が発表された¹⁾。この論文には1.83 Mbのゲノムサイズをもつインフルエンザ菌（インフルエンザウィルスとは異なる）のゲノムの全塩基配列を決定したプロジェクトの成果が記されており、全生物種の中で初めてゲノムの全容が明らかにされた歴史に残る記念すべき業績である。この論文を発表したのは、後に、公的機関による国際共同ヒトゲノムシーケンスプロジェクトに対して競争を挑んだ米国のバイオベンチャー企業、セララ・ジェノミクス社の社長を務めた Venter が率いる The

Institute for Genomic Research (TIGR) の研究グループであった^{2), 3)}。インフルエンザ菌を皮切りに、さまざまな生物種のゲノムシーケンスプロジェクトが遂行され、2003年4月14日、ヒトゲノム完全解読宣言という一つのマイルストーンに至った。現在もなお多くの生物種のプロジェクトが進行しているが、今日までの多種にわたる多様な生物種のゲノムシーケンスプロジェクトの成果によって、今や我々は質的にも量的にも飛躍的に充実した配列データベースを利用することが可能となった。現在の生命科学は、ゲノムの塩基配列とその解析によって得られた膨大な情報を利用しながら研究を進めるポストゲノム（シーケンス）時代に突入している。

質量分析法は、現在、ポストゲノム時代における生命科学の重要な研究領域として注目されている「プロテオミクス」の基盤技術として位置づけられている。しかし、質量分析法を利用してタンパク質、ペプチドのアミノ酸配列を解析し、タンパク質を同定することが可能となったのは、それほど昔のことではない。1世紀近い質量分析法の歴史の中で、ペプチド、タンパク質の構造解析に質量分析が汎用され始めたのは1980年代に入ってからのことである。

1980年以前でも、電子イオン化 (Electron Ionization: EI) 法や化学イオン化 (Chemical Ionization: CI) 法を用い

*a) 神戸大学・パイオシグナル研究センター (〒657-8501 神戸市灘区六甲台町 1-1)

Biosignal Research Center, Kobe University (1-1 Rokkodai-cho, Nada-ku, Kobe 657-8051, Japan)

b) (独)科学技術振興機構、戦略的基礎研究推進事業、領域名「たんぱく質の構造・機能と発現メカニズム」(〒332-0012 川口市本町 4-1-8)

Research Area "Protein Structure and Functional Mechanisms" in Core Research for Evolution Science and Technology (CREST), Japan Science and Technology Agency (4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan)

た直接導入による解析や、ガスクロマトグラフィー質量分析法 (Gas Chromatography/Mass Spectrometry: GC/MS) によってアミノ酸配列を解析することは可能であった^{4)~6)}。しかし、EI 法、CI 法を利用する場合、不揮発性のペプチドをイオン化させるために、あらかじめ揮発性の誘導体に変換する必要があり、誘導体化処理の煩雑さや、処理過程での測定試料の損失などの問題があった。EI 法、CI 法は、分子量が 1000 を超える化合物を解析すること自体が難しく、また、現在用いられている種々のソフトイオン化法とは異なり、イオン源においてフラグメンテーションが起こるイオン化法であるために得られたマススペクトルからペプチドのような複雑な化合物の構造を解析することは容易ではなかった。以上のようなさまざまな理由から、EI-MS、CI-MS を利用したペプチドの構造解析は、比較的分子量の低いペプチドでのみ適応可能な方法であった。通常は、一度の解析でせいぜい 5~6 残基のアミノ酸配列を決定することができる程度で、応用範囲は極めて限定されていた。

1970 年代に汎用された電界脱離 (Field Desorption: FD) イオン化法は、いわゆるソフトイオン化法の黎明期を担ったイオン化法である⁷⁾。このイオン化法の開発によって分子量 2000 程度の大きな、しかも不揮発性のペプチド分子を、誘導体化することなくイオン化させることが可能となった。「ソフトイオン化法」の名のとおり、EI 法や CI 法で観測されるイオン源でのフラグメンテーションは少なく、分子量情報の獲得に直接役立つ、いわゆる分子量関連イオンとして、専らペプチドのプロトン化分子が観測されるマススペクトルを得ることができる。この FD-MS を利用してタンパク質の 1 次構造解析も行われていたが^{8)~10)}、FD イオン化法は、試料を塗布するエミッターの取り扱いが極めて難しく、加えてエミッター上に塗布された試料の状態や性質によって、さまざまなイオン化モードによって分子がイオン化されてしまうことなどから、技術的には熟練を要する不安定なイオン化法であり、信頼できるデータを得るためには「魔術的」と称されるほど相当な技能が必要とされた¹¹⁾。

より簡便なソフトイオン化法の開発が待たれていたが、1981 年、Barber らが高速原子衝撃 (Fast Atom Bombardment: FAB) イオン化法を開発して以来^{12), 13)}、質量分析法を用いたタンパク質、ペプチドの構造解析法の状況が劇的に変化した。タンパク質の翻訳後修飾基の解析を中心に、タンパク質、ペプチドの 1 次構造解析や遺伝子の塩基配列から推定されたアミノ酸配列のバリデーション、化学合成ペプチドやリコンビナントタンパク質の品質管理など、タンパク質、ペプチドの構造解析に FAB-MS が盛んに利用されるようになり、質量分析法の有用性が多くの生命科学の研究者に認知されるようになった^{14), 15)}。

現在のようにタンパク質や核酸の配列データベースが充実していなかった当時、質量分析法を用いたタンパク質、ペプチドのアミノ酸配列の決定や、タンパク質の同定を行うためには、衝突誘起解離 (Collision-induced Dissocia-

tion: CID) によってペプチドのプロトン化分子を開裂させ、得られたフラグメント (プロダクト) イオンマススペクトルを「解読」し、アミノ酸配列を決定することが必要であった^{16), 17)}。エドマン分解処理後や、ペプチドのアミノ末端 (N 末端) から逐次アミノ酸を遊離させるアミノペプチダーゼ、もしくはカルボキシル末端 (C 末端) から逐次アミノ酸を遊離させるカルボキシペプチダーゼを用いて消化を行った後の断片ペプチドを質量分析することによって断片の質量差からアミノ酸配列を決定することも可能であった^{15), 18)}。現在、質量分析法を利用するタンパク質同定法の主流となっている質量分析データをダイレクトに配列データベース検索に用いる方法は、原理的にはアミノ酸配列を決定することなしに、タンパク質を同定しているのに対し、従来の質量分析法を利用したタンパク質同定法は、質量分析データからアミノ酸配列を決定し、その結果に基づいてタンパク質を同定していた。この点で、質量分析法を利用する従来法のタンパク質同定の基本的な戦略は、当時、タンパク質 1 次構造解析法の主流であったエドマン分解法を利用する方法と同じであった。現在 “*de novo* Sequencing” と呼ばれている質量分析のデータから直接アミノ酸配列を解読する従来法は、ペプチドが混合物であっても解析が可能である点や、エドマン分解の進行を妨げるアセチル基やピログルタミン酸などの N 末端の修飾の有無にかかわらずアミノ酸配列の決定が可能である点など、エドマン分解法に対して優れた点も数多くもっていた。しかし、アミノ酸配列がプロダクトイオンマススペクトルの質に影響すること、データの解釈、アミノ酸配列の解読にある程度の経験が必要とされること、エドマン分解法を利用してペプチド、タンパク質のアミノ酸配列分析を行うための機器であるプロテインシーケンサーに対して、質量分析計が高価であったことなどの問題点もあり、当時はエドマン分解法を凌駕するまでには至らなかった。

多種多様な生物種を対象としたゲノムシーケンシングプロジェクトの進行に伴う配列データベースの充実と、エレクトロスプレーイオン化 (Electrospray Ionization: ESI) 法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) 法の開発、および飛行時間型質量分析計 (Time-of-flight Mass Spectrometer: TOFMS)、イオントラップ型質量分析計 (Ion Trap Mass Spectrometer: ITMS) の発達に牽引されたプロテインマススペクトロメリーの発展は、これまでのタンパク質の同定法を一変させた^{19)~24)}。生命科学領域における質量分析法の応用例が急激に増加し、生化学や分子生物学を中心とした数多くの生命科学の研究者が質量分析法に対して高い関心を示すようになってきた。現在のように、タンパク質を同定する手段として質量分析法が主流となった大きな要因は、質量分析計自身の発達もさることながら、タンパク質や核酸の配列データベースが充実したこと、さらにそれを利用する検索ソフトウェアの発達によって、質量分析のデータからアミノ酸配列を解読することなく、タンパク質を同定することが可能となったため

である。本稿では、現在プロテオミクスに代表されるタンパク質を基盤とした生命科学においてタンパク質同定法の主流となっている、質量分析法と配列データベースを利用するタンパク質同定法を概説する。

2. 概 要

現在、質量分析法と配列データベースを利用してタンパク質を同定するために用いられている一般的なスキームは以下のとおりである。

- 1) アミノ酸残基特異的なタンパク質分解酵素（プロテアーゼ）を用いてタンパク質を消化し、断片化する。
- 2) プロテアーゼ消化物を ZipTip[®] などの逆相高速液体クロマトグラフィー用のカラム充填剤が先端に固定されたピペットチップを用いて脱塩、濃縮を行う。（逆相高速）液体クロマトグラフィー（Liquid Chromatography: LC）と質量分析法を直結させた LC/MS 法によって解析する場合はこの過程を省略できる。
- 3) 消化後の断片ペプチド混合物のマススペクトルを測定する。または、タンデム質量分析法（Tandem Mass Spectrometry, Mass Spectrometry/Mass Spectrometry: MS/MS）によって断片ペプチド由来のイオンを開裂させ、プロダクトイオンマススペクトルを測定する。
- 4) 得られたマススペクトルデータを用いて、タンパク質や核酸の配列データベースに対して検索を行い、データベースに登録されているタンパク質（核酸）の中でマススペクトルデータと最も適合するアミノ酸配列を探し出し、タンパク質を同定する。

上記のスキームに従ってタンパク質を同定するためには、データベース検索を行うソフトウェア、すなわちサーチエンジンが必要となる。そのため、現在の質量分析法を利用するタンパク質同定法の方法論は、サーチエンジンのアルゴリズムに大きく依存している。これまで数多くの研究者によって、独自のアルゴリズムを用いたさまざまなサーチエンジンが開発されてきたが、質量分析法と配列

データベースを利用するタンパク質同定法の方法論としては、3 種のカテゴリーに区分することができる (Fig. 1)。

まず MS/MS を利用するか、利用しないかで二分することができる。MS/MS を利用しない場合 (①) が、Peptide Mass Fingerprinting (PMF) 法である^{25)~29)}。

一方、MS/MS を利用する方法はさらに、データベース検索の前に、観測されたプロダクトイオンマススペクトルの解読を行うか、行わないかで二分することができる。この場合のプロダクトイオンマススペクトルは、MALDI-TOFMS を用いたポストソース分解 (Post-source Decay: PSD) によって得られたプロダクトイオンマススペクトルを含んでいる^{*}1。プロダクトイオンマススペクトルの解読は、主としてペプチドの部分アミノ酸配列を *de novo* で解析すること、すなわちデータベースに依存せずにペプチドの部分アミノ酸配列を「新規に (*de novo*)」読み解き、決定することを指している。

MS/MS を利用し、データベース検索の前にプロダクトイオンマススペクトルを解読しない方法 (②) は、現在、LC/MS/MS を利用するタンパク質の同定において最も頻用されている方法である。代表的なソフトウェアの名前を挙げると、Yates らが開発した SEQUEST[®] 30) や Cottrell らが開発した MS/MS Ions Search (MASCOT[®]) 31), 32) がこのカテゴリーに含まれる。このカテゴリーに含まれるアルゴリズムは現在、かなり汎用されている方法であるにもかかわらず、驚くべきことに、PMF (Peptide Mass Fingerprinting) 法のように、同じカテゴリーに含まれるすべてのアルゴリズムを包括する一般的な総称が定着していない。このカテゴリーの基本原理は、プロダクトイオンマススペクトルを解読することなく、プロダクトイオンマスリストの情報を直接データベースに対して検索し、タンパク質を同定することにある。筆者らは最近、“Methods in Molecular Biology” シリーズの成書に掲載したプロトコールの中で、このカテゴリーに含まれる方法を “Product Ion Mass Fingerprinting” 法という総称を与えて紹介した³³⁾。

もう一つのカテゴリーは、MS/MS を利用し、データベース検索を実施する前に、実測されたプロダクトイオンマススペクトルを解読することによって得た情報を検索に利用する方法 (③) である。プロダクトイオンマススペクトルからのペプチドの部分アミノ酸配列の解読、いわゆる *de novo* Sequencing によって得た情報や、低質量領域に観測されるインモニウムイオンの情報からペプチドに含まれるアミノ酸残基の種類を解読し、プリカーサーイオンの

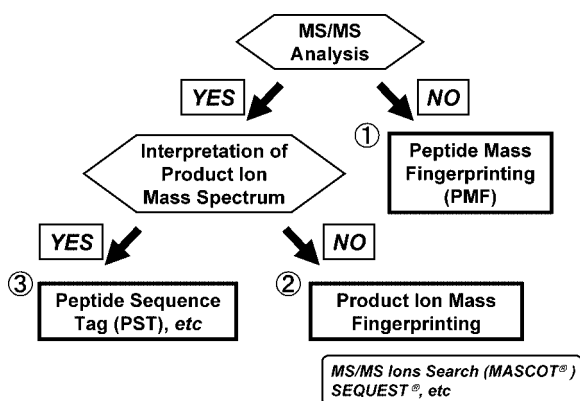


Fig. 1. Three categories of algorithms for mass spectrometry-based protein identification by correlation with sequence database.

^{*}1 PSD 法は、MALDI 法において、レーザー照射によって生成されたイオンが、イオン源の加速場領域を離れて検出器に到達する間に、主としてイオン自身もつ過剰な内部エネルギーによって分解する現象である。MALDI-TOFMS を用い、この現象を利用してプロダクトイオンマススペクトルを得ることは可能であるが、2 台の質量分離部を用いていない点や、解析可能なプロダクトイオンマススペクトルを得るためには測定試料に高い純度が求められる点などから、原理的には通常の MS/MS のカテゴリーには入らない。PSD と MS/MS を混同してはいけない。

Mascot Search

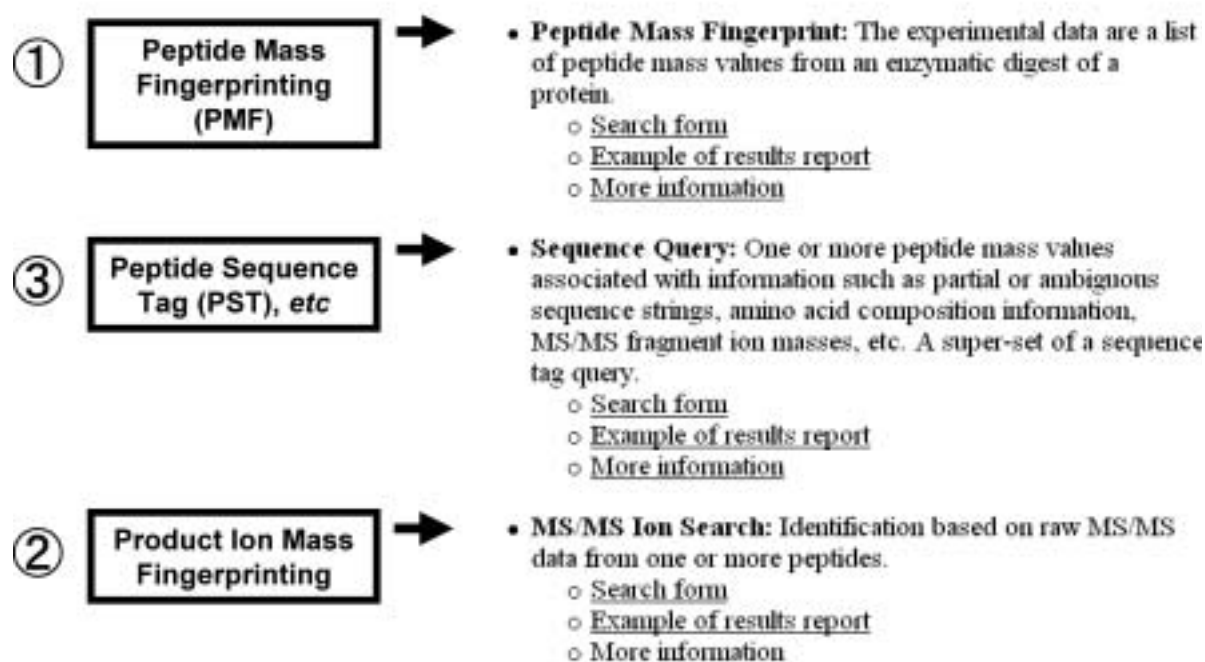


Fig. 2. First page of the MASCOT[®] search engine (Ver. 1.9).

実測質量に加えて、これらの解釈結果を用いて検索が行われる。de novo Sequencing は、マニュアルで行わなくても de novo Sequencing 用のソフトウェアを用いてペプチドの部分アミノ酸配列を推定することは可能である。このカテゴリーに含まれる方法は、データベース検索の前に、プロダクトイオンマスペクトルの解釈が必要という点で、マスペクトルのデータを直接使用する PMF 法や Product Ion Mass Fingerprinting 法とは明確に異なる方法である。当然ながらハイスループットな方法ではない。このカテゴリーに含まれる代表的な方法は、Mann と Wilm によって考案された Peptide Sequence Tag (PST) 法である³⁴⁾。

英国 Matrix Science 社が開発した検索エンジンパッケージ MASCOT[®]³⁵⁾、および米国カリフォルニア大学サンフランシスコ校 (UCSF) が開発したツールパッケージ ProteinProspector³⁶⁾ を用いれば、この 3 種のカテゴリーに属するそれぞれのアルゴリズムを個別に採用している 3 種類のソフトウェアを利用することができる。カテゴリー分類の理解を助けるために、MASCOT[®] 検索を利用する際の初期画面を Fig. 2 に示した。

3. Peptide Mass Fingerprinting (PMF) 法によるタンパク質の同定

3.1 アミノ酸残基特異的タンパク質分解酵素を用いたタンパク質の断片化

大阪府箕面市の土壌から分離された *Achromobacter lyti-*

cus M497-1 株が産生するプロテアーゼ、*Achromobacter* Protease I は、ポリペプチド鎖中のリジン残基のカルボキシル基側においてのみペプチド結合の加水分解を触媒する^{37), 38)}。同様に、脊椎動物の十二指腸でのタンパク質の消化に重要な役割を担っているトリプシンは、ポリペプチド鎖中のリジン残基もしくはアルギニン残基のカルボキシル基側のペプチド結合の加水分解だけを触媒する。一方、トリプシンと同じ消化酵素であっても、脊椎動物の胃内のタンパク質消化を担うプロテアーゼ、ペプシンは、トリプシンのような高い基質特異性はなく、かなり広い基質特異性を有している。これまでに数多くのプロテアーゼが報告されているが、その中で、トリプシンや *Achromobacter* Protease I に代表される、ペプチド鎖中の特定のアミノ酸残基を認識し、特異的に当該アミノ酸残基が関与するペプチド結合の加水分解を触媒するプロテアーゼ (Table 1) を用いてタンパク質を消化した場合、ランダムに断片ペプチドが生成されるわけではない。例えば、上述の *Achromobacter* Protease I (Lysyl Endopeptidase[®]; LEP) を用いた場合、生成される断片ペプチドの C 末端は、タンパク質本来の C 末端ペプチドを除いて、理論的にはすべてリジン残基となる。逆に、一部の例外を除いて、リジン残基が断片ペプチドの C 末端以外の場所を占めることはない。それゆえ、タンパク質のアミノ酸配列が既知であれば、消化によって生成されるすべての断片ペプチドのアミノ酸配列を理論的に知ることができる。

以下、ヒト由来のプロテインキナーゼ C- α タイプ

Table 1. Amino Acid Residue-specific Endoproteases That Are Commercially Available

Name of enzyme	Specificity	Source	EC-number
Trypsin	Lys-X, Arg-X	Bovine or Porcine Pancreas	EC 3.4.21.4
<i>Achromobacter</i> Protease I (Lysyl Endopeptidase [®])	Lys-X	<i>Achromobacter lyticus</i> M497-1 strain	EC 3.4.21.50
Endoproteinase Lys-C	Lys-X	<i>Lysobacter enzymogens</i>	EC 3.4.21.50
Endoproteinase Asp-N	X-Asp	<i>Pseudomonas fragi</i> mutant	EC 3.4.24.2
<i>Staphylococcus aureus</i> Protease (V8 Protease, Endoproteinase Glu-C)	Glu-X, Asp-X ^{a)}	<i>Staphylococcus aureus</i> V8 strain	EC 3.4.21.19
Arginylendopeptidase (Mouse submandibular protease)	Arg-X	Murine Submaxillary Gland	Unregistered
Endoproteinase Arg-C	Arg-X	<i>Clostridium histolyticum</i>	EC 3.4.22.8
Asparaginylendopeptidase	Asn-X	Jack Bean	Unregistered
Proline Specific Endopeptidase	Pro-X	<i>Flavobacterium meningosepticum</i>	EC 3.4.21.26

^{a)} Specificity for Glu-X is achieved in ammonium bicarbonate buffer (pH 7.8) or ammonium acetate (pH 4.0).

MADVPPGNDSTASQDVANRFARKGALRQKN 30
VHEVKDHKFIARFFKQPTFC SHCTDFIWGFGK 60
GKQGFQCQVC CFVVKRCHE FVTFSCPGAD 90
KGPDTPDDPRSKHKFKIHTYG SPTFCDHCGSL 120
LLYGLIHQGMKCDTCMDNVHKQCVINVPSSL 150
CGMDHTEK 158

Fig. 3. Amino acid sequence of N-terminal region (1–158) of PKC α .

Table 2. Amino Acid Sequences and Monoisotopic Mass Values of the Theoretical LEP Fragments of PKC α (1–158)^{a)}

Fragment (residue Nos.)	Amino acid sequence	Monoisotopic mass value (MH ⁺)
1– 28	MADVPPGNDSTASQDVANRFARK	2497.19
24– 29	GALRQK	672.41
30– 35	NVHEVK	725.39
36– 38	DHK	399.20
39– 45	FIARFFK	928.54
46– 62	QPTFCSHCTDFIWGFGK	2087.91
63– 76	QGFQCQVCCFVVK	1796.80
77– 91	RCHEFVTFSCPGADK	1810.80
92–101	GPDTDDPRSK	1087.50
102–103	HK	284.17
104–105	FK	294.18
106–131	IHTYGSPTFCDHCGSLLYGLIHQGMK	2992.39
132–141	CDTCMDNVHK	1279.49
142–158	QCVINVPSSLCGMDHTEK	1987.90

^{a)} All cysteine residues (C) are modified by carbamido-methyl group.

(PKC α) (accession No. P17252) を例に挙げて説明するが、ヒトの PKC α は 672 残基からなるタンパク質であり、図や表が複雑になることから、簡略化するために、フォル

Table 3. Monoisotopic Mass Values of Genetically Coded 20 Amino Acid Residues

3-Letter Code	1-Letter Code	Monoisotopic Mass
Ala	A	71.04
Cys	C	103.01
Asp	D	115.03
Glu	E	129.04
Phe	F	147.07
Gly	G	57.02
His	H	137.06
Ile	I	113.08
Lys	K	128.09
Leu	L	113.08
Met	M	131.04
Asn	N	114.04
Pro	P	97.05
Gln	Q	128.06
Arg	R	156.10
Ser	S	87.03
Thr	T	101.05
Val	V	99.07
Trp	W	186.08
Tyr	Y	163.06

ボールエステルやジアシルグリセロールの結合部位を含む N 末端側 158 残基までの領域を例にあげる。PKC α (1–158) には Lys¹⁵⁸ を含めて 14 のリジン残基が含まれているが (Fig. 3), LEP を用いて消化した場合、14 の断片ペプチドが生成される (Table 2).

3.2 ペプチドマスリスト

アミノ酸 (残基) はそれぞれ固有の質量を有している (Table 3). 断片ペプチドの質量も、アミノ酸組成が異なれば、それぞれ固有の質量を有する。

PKC α (1–158) の LEP フラグメント (24–29), GALRQK を例に挙げて説明する。この断片ペプチドを構成するアミノ酸残基の残基質量 (単一同位体質量 monoisotopic mass) は、N 末端から、グリシン (G) が 57.02、つづいてアラニン (A) が 71.04、ロイシン (L) が 113.08、アルギニン (R) が 156.10、グルタミン (Q) が 128.06、リジン (K) が

128.09 である。したがって、この断片ペプチドのプロトン化分子 MH^+ の理論質量（単一同位体質量）はこれらの残基質量の和に、水 1 分子分のモノアイソトピック質量 18.01 と、プロトンの質量 1.01 を加えた 672.41 となる。この方法で、GALRQK のアミノ酸配列（アミノ酸組成）を数値化することが可能となる。同様に、他の消化断片もアミノ酸配列（組成）に従って数値化することができる。このように、任意のタンパク質を任意のアミノ酸残基特異的なプロテアーゼを用いて消化した際に生成されるすべての断片ペプチドについてこの作業を行えば、当該タンパク質から生成される特異的な断片ペプチド群の理論質量のリスト、ペプチドマスリストを作成することができる。Table 2 の右側のカラムが、PKC α (1-158) を LEP 消化した際に得られる断片ペプチド群の理論ペプチドマスリストとなる。この理論ペプチドマスリストを理論的なマススペクトルとして表すことができる。PKC α (1-158) を LEP 消化した際に得られる断片ペプチド群の理論マススペクトルを Fig. 4 に示した。

タンパク質は 1 次構造、すなわちアミノ酸配列に独自性があり、異なるタンパク質は、当然のことながら異なるアミノ酸配列を有している。したがって、アミノ酸残基特異的なプロテアーゼを用いて消化した際に得られる断片ペプチド群のアミノ酸配列の組み合わせはタンパク質に固有のものとなる。断片ペプチド群を構成する個々のペプチドのアミノ酸配列（正確にはアミノ酸組成）が異なれば、その質量も異なることから、アミノ酸残基特異的なプロテアーゼを用いて消化した際に得られる断片ペプチド群の質量の組み合わせはタンパク質に固有のものとなる。この断片ペプチド群の質量の組み合わせがペプチドマスリストである。したがって、アミノ酸配列が異なれば、ペプチドマスリストもタンパク質に固有のものとなる。さらに、このペプチドマスリストをスペクトル表記したもので、すなわち断片ペプチド群の理論マススペクトルもタンパク質に固有のものとなる (Fig. 5)。

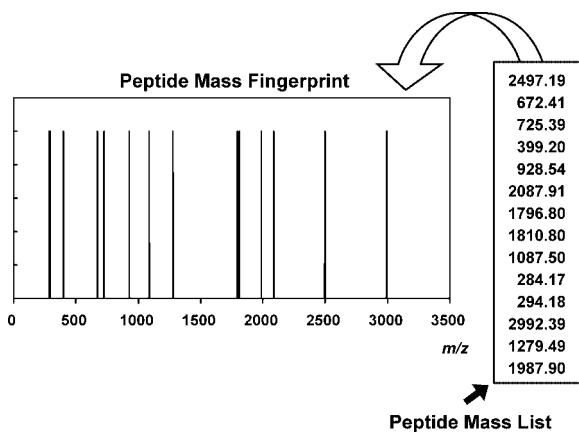


Fig. 4. Theoretical peptide mass list of PKC α (1-158), which is obtained by digestion with LEP, and the corresponding theoretical peptide mass fingerprint.

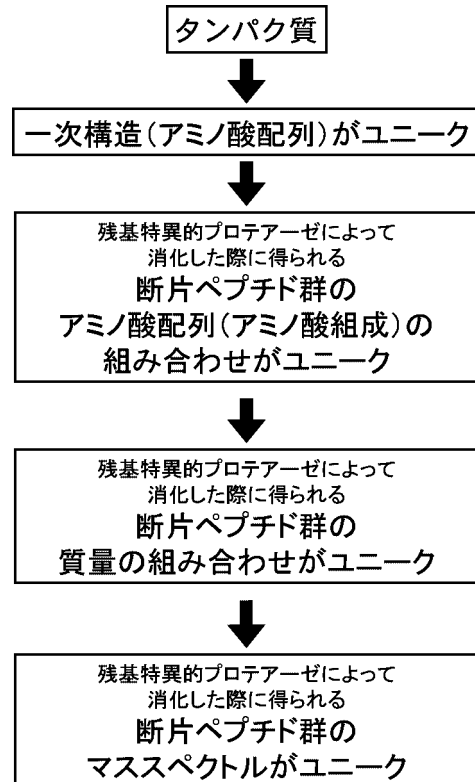


Fig. 5. Principle of protein identification by peptide mass fingerprinting.

我々人間がもつ目に見える形質の中で、個人に固有なものの一つに指紋がある。同じゲノムを有する一卵性双生児でも異なるほど極めて独自性の高い形質である。指紋と言えば犯罪捜査に用いられるイメージが強いが、最近では、放射性同位元素使用施設などの種々の管理区域への入退室や、パソコンのログオンの際に必要な個人認証システムにも使用されている。タンパク質の場合、ペプチドマスリストはタンパク質に固有のものであり、これをスペクトル表記したものが、正しくタンパク質を識別するための「指紋 (Fingerprint)」、すなわち “Peptide Mass Fingerprint” である。これまでは 1 例として、PKC α (1-158) の LEP 消化によって得られる断片ペプチドの理論ペプチドマスリスト (Table 2) に基づき、Peptide Mass Fingerprint (Fig. 4) を作成する過程を説明してきたが、配列データベース上に存在するすべてのタンパク質について同様の作業を行えば、タンパク質の指紋である Peptide Mass Fingerprint のデータベースを作成することができる。

3.3 データベース検索

ある未知のタンパク質を、あるアミノ酸残基特異的なプロテアーゼを用いて消化し、それぞれの断片を分離することなく消化物をそのまま質量分析することによって、その未知タンパク質の Peptide Mass Fingerprint を得ることができる。この実験的に得られた未知タンパク質の Peptide Mass Fingerprint が、配列データベース上のすべてのタンパク質を理論的に断片化した Peptide Mass Fingerprint の中で、どれが一番適合するかを検索すれば、統

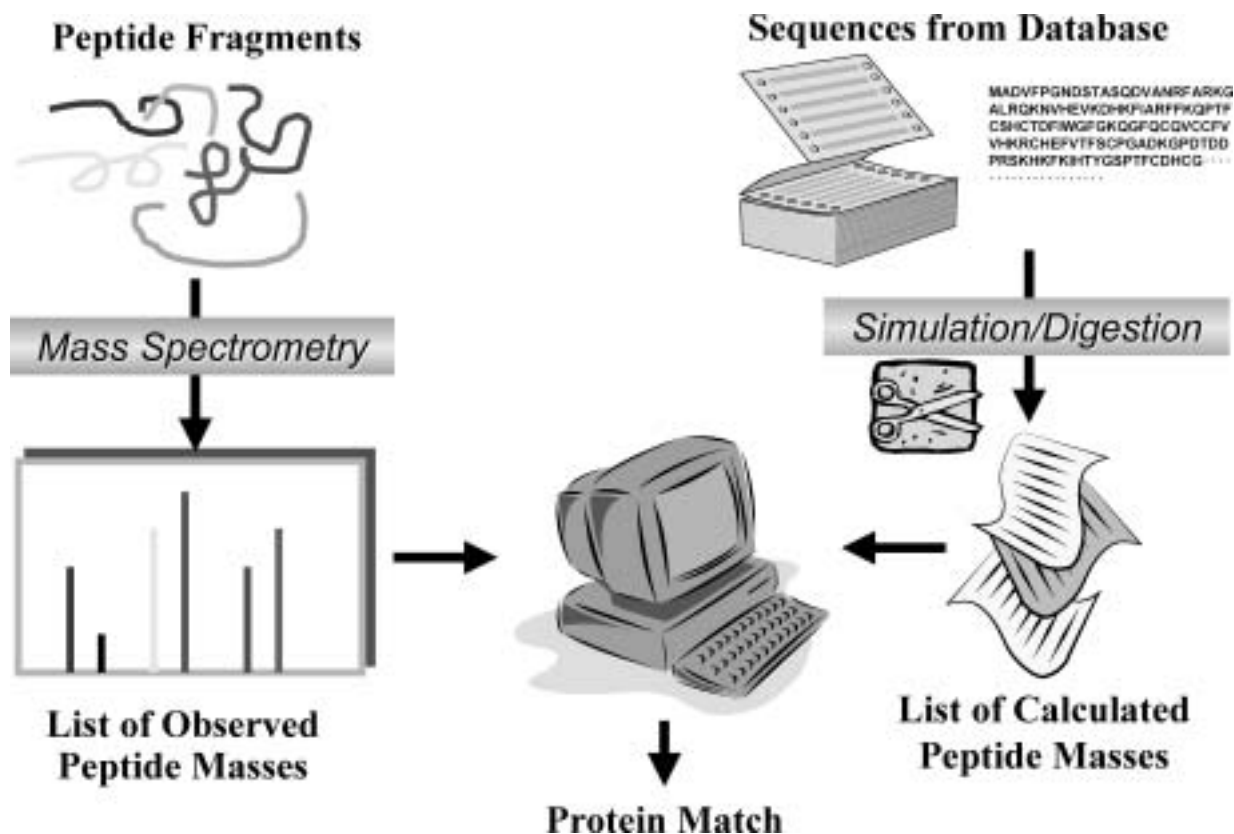


Fig. 6. A schematic illustration of the concept of protein identification by peptide mass fingerprinting.

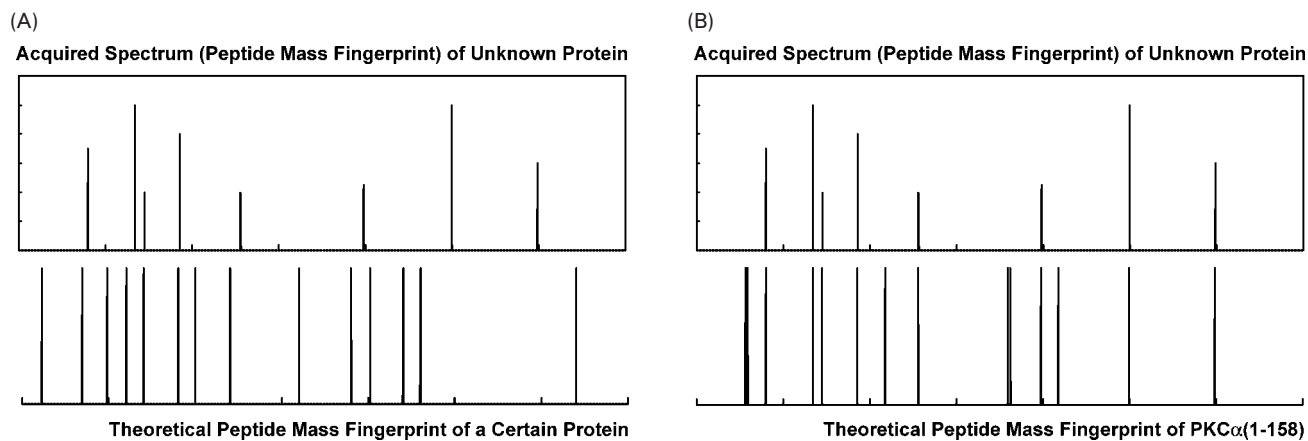


Fig. 7. Comparison of an acquired peptide mass fingerprint of unknown protein with theoretical peptide mass fingerprints of mismatch (A) and the closest match (B).

計的に最も可能性の高いタンパク質を示すことが可能となる (Fig. 6). Fig. 7A 上は, ある未知タンパク質を LEP 消化し, 消化物を質量分析して得たマスペクトル, すなわち Peptide Mass Fingerprint である. 次に, 配列データベース中のすべてのタンパク質をコンピューター上でバーチャルに LEP 消化し, 得られた仮想断片ペプチドの配列から理論的な Peptide Mass Fingerprint を作成する. データベースに 30 万種類のタンパク質がエントリーされていれば, すべてのタンパク質についてこのバーチャルな酵素消化を行う. この作業によって得られたすべての理論 Peptide Mass Fingerprint と, 測定された未知タンパク

質の Peptide Mass Fingerprint との比較を行う. 正に, 犯行現場から採取された犯人の指紋と, データベースに登録されている膨大な指紋データとを照合する作業と同じである. 測定された未知タンパク質の Peptide Mass Fingerprint は, データベースに登録されているほとんどのタンパク質の理論 Peptide Mass Fingerprint とは合致しない (Fig. 7A). 検索に用いたデータベースに“未知”タンパク質の配列情報が登録されているならば, 理想的には, “未知”タンパク質の Peptide Mass Fingerprint と合致する理論 Peptide Mass Fingerprint が存在するはずであるが, 実際にはさまざまな理由から“未知”タンパク質の Pep-

ptide Mass Fingerprint と全く同じ理論 Peptide Mass Fingerprint を探し出すことはまず不可能である。したがって、実際の同定では、測定された Peptide Mass Fingerprint に最も似ている理論的な Peptide Mass Fingerprint を探している (Fig. 7B)。ここでは、理解を助けるために、画像としてのマススペクトルを比較させているが、コンピューターによって行われる実際の検索では、当然ながら、画像どうしの比較ではなく、数値どうしの比較が行われている。

Fig. 8 に、実測された 6 種類のタンパク質の Peptide Mass Fingerprint (LEP 消化) を示した。A と B はいずれも血清アルブミンであるが、ウシ (Fig. 8A) とヒト (Fig. 8B) の違いである。同じ種類のタンパク質であっても生物種が異なることによってアミノ酸配列も異なり、その結果 Peptide Mass Fingerprint も異なっていることが明確にわかる。残りの 4 種の Peptide Mass Fingerprint はすべてヒト由来のタンパク質である。違いがわかりやすいものを故意に選んだわけではないにもかかわらず、タンパク質の種類が異なることで Peptide Mass Fingerprint も明確に異なることが、ほんの 6 種のタンパク質の実測データを比較しただけでも実感できる。

3.4 Peptide Mass Fingerprint 実測データの特徴

実際の解析において、理論 Peptide Mass Fingerprint と全く同じものを実測データとして得ることができない理由には二つの要因がある。一つ目の要因は、マススペクトル上に観測されたピークの横軸のずれである。これは、質量分析の際の測定誤差に起因するものであり、測定精度を上げる努力をすれば、この点に関しては理論値と全く同一のデータを得ることはできないまでも、限りなく近似したデータを得ることは可能である。

二つ目の要因は、酵素消化によって生成された断片ペプチドの一部がイオンとして検出されないことである。上述の測定誤差の問題は、検出された (存在している) イオンの測定値と理論値との間に差が生じているという問題であるが、この場合、パーシャルな酵素消化によって理論的には生成されるはずの断片ペプチドに由来するイオンの一部がマススペクトル (Peptide Mass Fingerprint) 上に存在していないという測定精度以前の問題である。その理由には下記のことが考えられる。

- 1) 試料の前処理の過程で、すべての断片ペプチドが同じ収率で回収されない。

タンパク質のプロテアーゼ消化物のマススペクトルを測定する前には、必ず無機塩などのイオン化を阻害する夾雑物を除くために前処理 (脱塩処理) を行う。多くの場合、逆相高速液体クロマトグラフィー用のカラム充填剤を用いて行われるが (上述、「2. 概要」参照)、充填剤に対する吸着 (保持) 効率や溶出効率はペプチド分子の疎水性に依存する。親水性の高いペプチドは吸着効率が低く、充填剤に吸着 (保持) されずに洗浄の過程で無機塩類とともに洗い流されてしまう。一方、疎水性の高いペプチドは充填剤への吸着効率は

高いが、逆に溶出効率が低いために、充填剤に保持され続け、通常の溶出条件で充填剤から溶出させることが難しい。したがって、この前処理を施した後の試料は、親水性や疎水性の高い断片ペプチドを失った状態であり、質量分析計に導入される前段階ですべての断片ペプチドが含まれていない試料となっている。また、充填剤から溶出される断片ペプチドも、ペプチドの疎水性によって溶出効率が異なるので、すべてのペプチドが均等に溶出されるわけではない。

- 2) すべての断片ペプチドが同じイオン生成効率をもっていない。

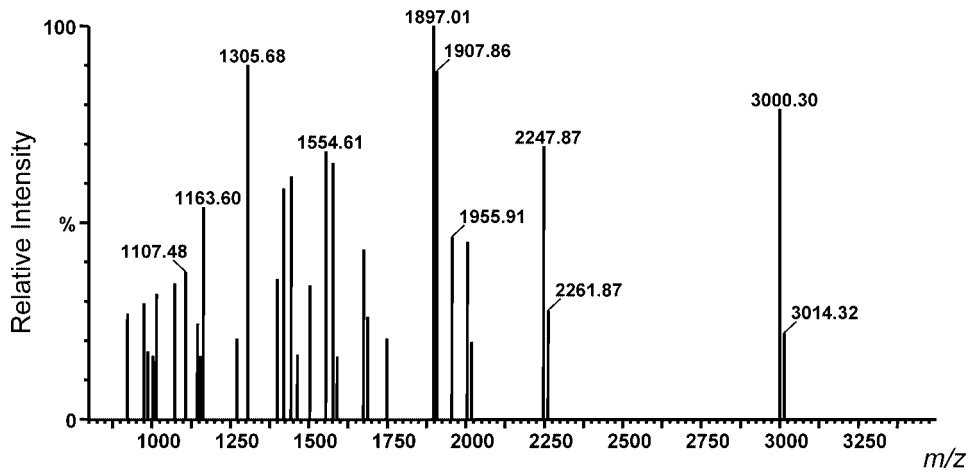
現在用いられている質量分析計は、電荷をもったイオンの質量を測定するための分析機器で、電荷をもたない電氣的に中性の分子の質量を測定することはできない。したがって、測定試料中の分子は、質量分離部に導入される前に、イオン源において、ESI 法や MALDI 法などのイオン化法によってイオン化されなければならない。ペプチド分子のイオン生成効率は、等電点や質量等のペプチドの物理的・化学的性質に依存している。アミノ酸配列が異なれば、このような物理的・化学的性質も異なることから、消化後の断片ペプチドのイオン生成効率は千差万別であり、同じモル数の断片ペプチドが存在していても個々のペプチドのイオン生成効率を反映して、観測されるイオンの強度が異なってくる。生成効率の低いペプチドはイオン強度が微弱であるか、観測されないことも多い。イオン生成効率に差がある化合物を混合して測定する場合、この差が拡大するという「サプレッション効果」も知られており、測定試料中に存在しているすべてのペプチド分子がマススペクトル上に観測されるわけではない。ペプチド分子のイオン生成効率は、用いるイオン化法によっても異なってくるので、同じ試料を測定しても ESI 法と MALDI 法で同じマススペクトルが得られることはほとんどない。

- 3) 質量分析計の質量分離部はすべてのイオンを同じ効率で通過させて、検出器に到達させてはいない。

プロテインマススペクトロメリーの分野では、ITMS や、静電界ミラー (リフレクター) を備えた TOFMS が汎用されているが、いずれの分析計も、すべての質量範囲のイオンを同じ効率で質量分離部の中を通過させてはいない。リフレクターを備えた TOFMS を用いた場合は、高質量になればなるほど観測されるイオン量が低下する。ITMS の場合は、一度イオンをトラップする関係上、設定したトラップ効率の最も高い m/z (質量電荷比) から離れれば離れるほどトラップ効率は低下するので、高質量領域に加えて、低質量領域のイオン量も減少する。リニア型 TOFMS は、広い質量領域にわたるイオンを分離することができるうえ、原理的には、ペプチド程度の質量をもつイオンであればすべてのイオンを同じ効率で通過させ、検出することが可能である。この点では理想

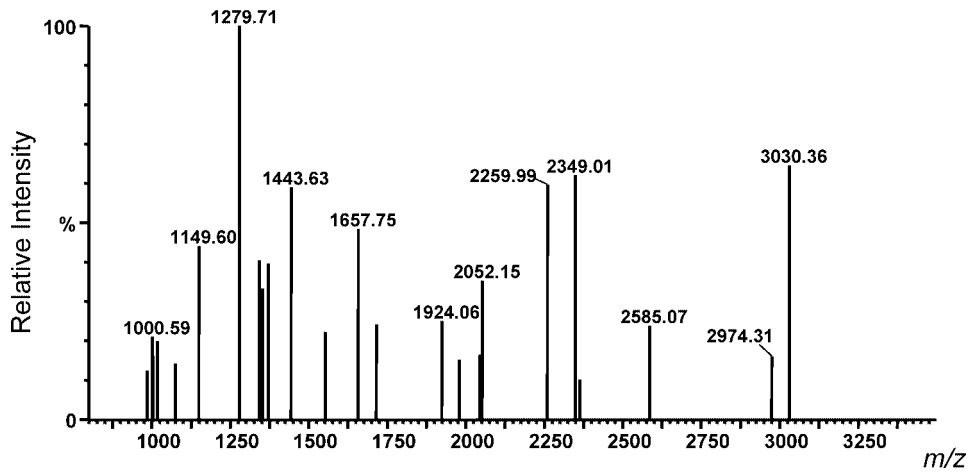
(A)

Serum albumin (Bovine)



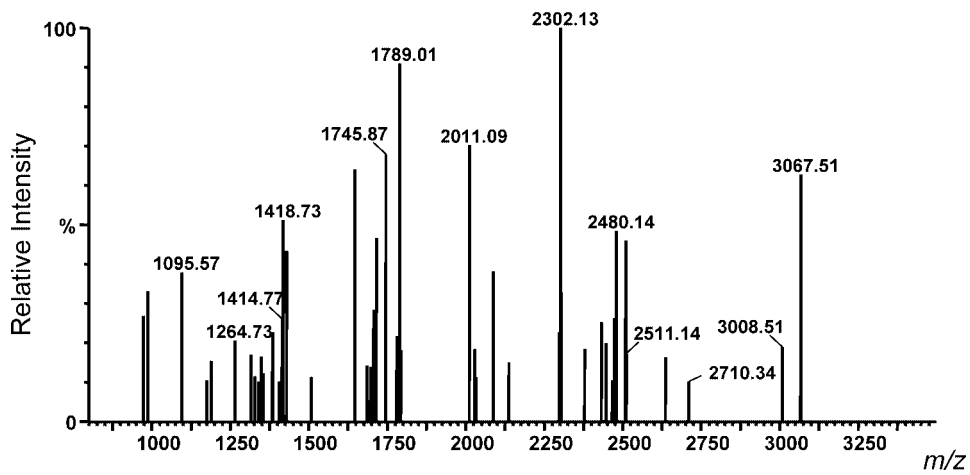
(B)

Serum albumin (Human)



(C)

Aldehyde dehydrogenase (Human)



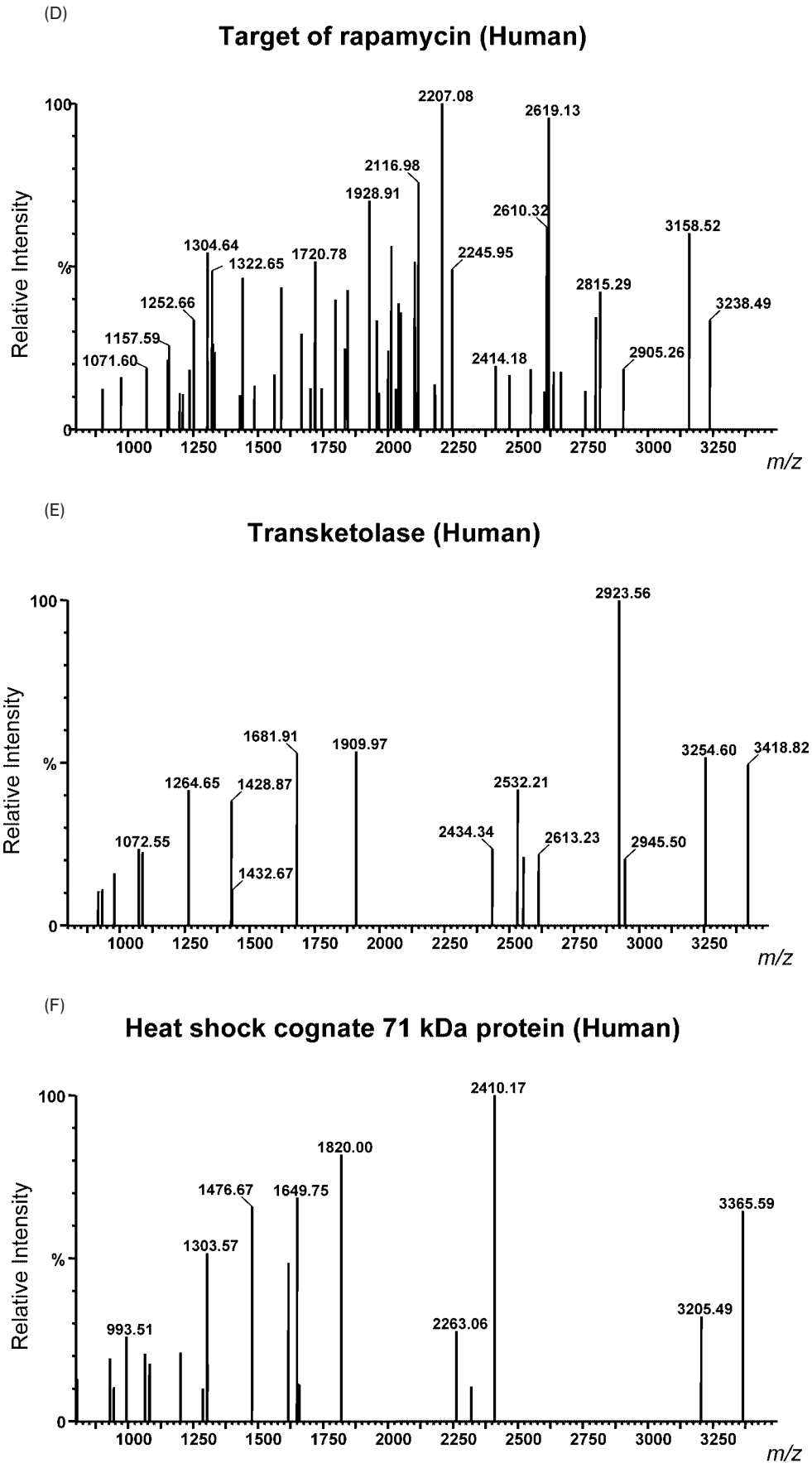


Fig. 8. Observed peptide mass fingerprints of bovine serum albumin [accession No.: P02769] (A), human serum albumin [accession No.: P02768] (B), human aldehyde dehydrogenase [accession No.: P00352] (C), mammalian target of rapamycin (mTOR) of human [accession No.: P42345] (D), human transketolase [accession No.: P29401] (E), and heat shock cognate 71 kDa protein of human [accession No.: XP_377283] (F).

的な分析計ではある。しかし、用いるイオン化法にも依存するが、分解能が低いことが難点である。

上述した三つの理由のうち、酵素消化によって生成されたすべての断片ペプチドがマススペクトル上に観測されないことに最も大きく影響している要因は、1) であり、次に2) である。3) の影響は相対的には少ない。

一つのタンパク質分子からアミノ酸配列の異なる特異的な断片ペプチドが生成する性質を原理的に利用しているPMF法であるが、断片ペプチドのアミノ酸配列が異なるがゆえに、前処理の過程で断片ペプチドの一部を損失し、またイオン生成効率が異なるというジレンマがある。実際の解析では、理論的に生成するすべての断片ペプチドが検出されることは皆無に近く、タンパク質の分子量にもよるが、十分量の試料を用いても半数以上の断片が検出されることも極めてまれなケースである。しかし、タンパク質のアミノ酸配列や測定精度にも依存するが、10種にも満たない断片ペプチドからなる部分的なペプチドマスリストであっても、PMF法によってタンパク質を同定することは可能である。

3.5 ペプチドマスリスト作成時の留意点

上述したように、理論的に生成されるすべての断片ペプチドが観測されるわけではなく、また検出されてもその相対的なイオン強度が異なることから、PMF法を利用する検索に必要なデータは、ペプチドマスリスト、すなわちイオンの観測質量（多くの場合はMH⁺）のリストであり、イオンの強度データは用いられない。したがって、検索の際には、イオン強度の強いイオンに由来するデータと、ノイズレベル付近の微弱なイオンに由来するデータとは全く同じ条件で扱われるので、実測されたマススペクトルからペプチドマスリストを作成する際には、False Positive（偽

同定）を防ぐために Threshold（閾値）の見極め、すなわち、有意な信号として考え、マスリストに加えるイオン強度と、ノイズとして考え、マスリストには加えないイオン強度との境界値の設定に留意しなければならない。また、PMF法において利用するデータはすべてペプチドマス、すなわち断片ペプチドの質量の値という1種類の情報のみであることから、検索結果の信頼度は実測値の精度に大きく依存する。測定には質量精度の優れた分析計を用いることは言うまでもなく、測定の際には常に厳密な質量較正を行う必要がある。筆者らは通常外部標準法で30~50 ppmの精度を確保した条件で測定したデータを用いてPMF法を行っている。内部標準法によってさらに精度を上げることも可能である。測定精度の高いデータを用いれば、検索、同定結果の信頼度が上昇し、検索結果そのものに悩む必要はなくなることから同定プロセスにおけるトータルのスループットも上昇する。

3.6 PMF法用のソフトウェア（検索エンジン）

PMF法によってタンパク質を同定するためには検索のためのソフトウェア、サーチエンジンが必要となる。これまでに異なる検索アルゴリズムを用いた数々のソフトウェアが開発されているが、その多くはインターネット上で公開されており、オンラインで検索することが可能である。PMF法用のサーチエンジンを利用することが可能なウェブサイトはTable 4に示した。ほとんどのソフトウェアはライセンスを取得することによってin-houseのサーバーを構築することが可能である。

4. MS/MSによるタンパク質の同定

4.1 MS/MSを利用したタンパク質同定法の発展

MS/MSを利用してタンパク質を同定する際、従来は

Table 4. Search Engines for Protein Identification by PMF on the Internet (as of March 2004)

Search engine	Resource	Uniform Resource Locators (URL)
Peptide Mass Fingerprint (MASCOT [®])	Matrix Science	http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF
MS-fit (ProteinProspector)	University of California, San Francisco (UCSF)	http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm
ProFound (PROWL)	Rockefeller University	http://prowl.rockefeller.edu/profound-bin/WebProFound.exe
ProFound [™] (Knexus [™])	ProteoMetrics (Genomic Solutions)	http://65.219.84.5/service/prowl/profound.html
PeptideSearch (Protein identification by peptide mass data)	European Molecular Biology Laboratory (EMBL)	http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html
PeptIdent (ExPASy)	Swiss Institute of Bioinformatics (SIB)	http://us.expasy.org/tools/peptident.html
PepMAPPER	University of Manchester Institute of Science and Technology	http://wolf.bms.umist.ac.uk/mapper/
MassSearch	Eidgenössische Technische Hochschule (ETH) Zürich	http://cbrg.inf.ethz.ch/Server/ServerBooklet/MassSearchEx.html#exMassSearch

CID (もしくは PSD) によって得たプロダクトイオンマススペクトルを「解読」する必要があった。BLAST や FASTA などの相同性検索プログラムを用い、解読したアミノ酸配列と相同な配列を有するタンパク質を検索することによってタンパク質の同定が行われてきた。現在では、配列データベースが充実したことに加え、同定のための種々のアルゴリズムが開発されたことから、対象となるタンパク質が配列データベースに登録されている場合は、PMF 法と同様に、観測されたプロダクトイオンマススペクトルから得た情報を用いるデータベース検索によって、迅速にタンパク質の同定を行うことが可能となっている。

4.2 MS/MS を利用する 2 種類の検索アルゴリズム— Product Ion Mass Fingerprinting 法と Peptide Sequence Tag (PST) 法—

上述したように (「2. 概要」参照)、配列データベースと MS/MS 解析によって得たデータを利用してタンパク質の同定を行うためのアルゴリズムは、大きく 2 種類のカテゴリーに分類することができる。一つはデータベース検索を行う前に、得られたプロダクトイオンマススペクトルの解読を全く行わずに、プロダクトイオンマススペクトルのデータ (プロダクトイオンマスリスト) を数値データとして直接データベース検索に用いるアルゴリズムで、1994 年、当時、米国・シアトルの University of Washington に所属していた Yates ら³⁰⁾の研究グループによって提唱された原理、アルゴリズムから発展したものである。Yates らが開発したアルゴリズムは、現在 SEQUEST[®] と名づけられた検索エンジンのベースになっている。筆者らはこのカテゴリーに属するタンパク質同定法を Product Ion Mass Fingerprinting 法と称している。

もう一つのアプローチは、部分的ではあるがプロダクトイオンマススペクトルの解読を必要とする方法で、同じく 1994 年、当時、ドイツ・ハイデルベルクの European Molecular Biology Laboratory (EMBL) に所属していた Mann らの研究グループによって提唱された PST (Peptide Sequence Tag) 法がその代表的なアルゴリズムである³⁴⁾。

5. Product Ion Mass Fingerprinting 法—プロダクトイオンマスリストを用いたデータベース検索—

5.1 プロダクトイオンマスリスト

PMF 法では、データベース検索に供する質量分析のデータは、タンパク質をアミノ酸残基特異的なプロテアーゼを用いて消化した際に得られる断片ペプチドの質量のリスト、ペプチドマスリストである。これに対し、Product Ion Mass Fingerprinting 法の場合は、断片ペプチドの MS/MS 解析によって得たプロダクトイオン (フラグメントイオン) の質量のリストである。タンパク質をアミノ酸残基特異的なプロテアーゼを用いて断片化した場合、切断される部位が特異的に規定されていることから断片ペプチドはランダムには生じない。同様に、MS/MS の場合、CID によるフラグメンテーションによって生成されるプロダク

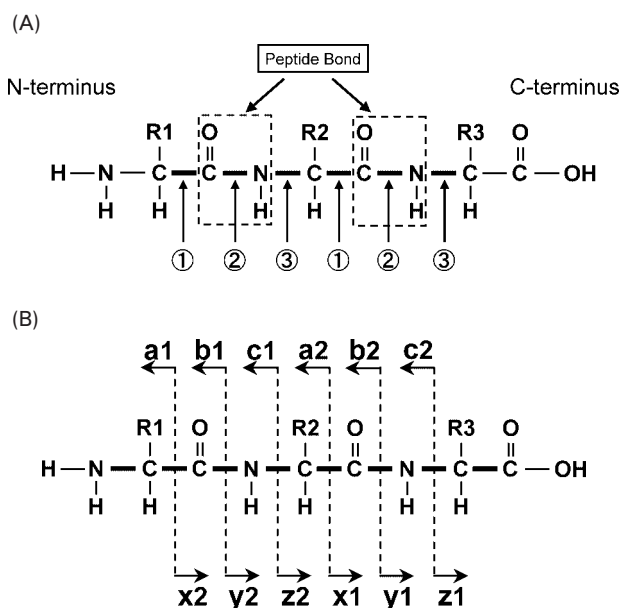


Fig. 9. Peptide fragmentation. (A) Peptide backbone; (B) Three types of fragmentation of peptide backbone.

トイオンも、ペプチド分子を構成する種々の原子間の結合がランダムに切断されて生成されるわけではない。ペプチドのフラグメンテーションは、主としてペプチドの主鎖骨格を形成する結合が開裂する。ペプチドの主鎖骨格を構成する結合にはプロリンが関与する結合を除いては下記の 3 種類の結合が存在する (Fig. 9A)。

- ① α 炭素原子とカルボニル基 (CO) 間の結合
- ② カルボニル基とイミノ基 (NH) 間、いわゆるペプチド結合
- ③ イミノ基と α 炭素原子間の結合

したがって、ペプチド主鎖の結合が開裂したプロダクトイオンは N 末端側が 3 種類、C 末端側が 3 種類、計 6 種類のイオンに分類することができる (Fig. 9B)³⁹⁾。①の結合が開裂した場合の N 末端側のイオンは a タイプ、C 末端側のイオンは x タイプと呼ばれている (Fig. 10A)。②の結合、すなわちペプチド結合が開裂した結果生成されるプロダクトイオンは、N 末端側が b type、C 末端側が y タイプと呼ばれている (Fig. 10B)。③の結合が開裂した N 末端側のプロダクトイオンは c タイプ、C 末端側のプロダクトイオンは z タイプと呼ばれている (Fig. 10C)。異なるペプチド主鎖の結合が開裂した同じタイプの一連のプロダクトイオンの質量差からペプチドのアミノ酸配列 (シーケンス) を読み取ることができることから、a, b, c, x, y, z の 6 タイプのプロダクトイオンは特にシーケンスイオンと呼ばれている。観測されるプロダクトイオン (シーケンスイオン) の種類はフラグメンテーションを誘起する条件や用いる質量分析計によって異なるが、例えば LC/MS/MS に汎用されている ESI-ITMS や ESI-Q-TOFMS の場合 (Q は Quadrupole: 四重極の略)、ペプチド主鎖を構成する 3 種の結合の中では、専らペプチド結合が開裂して生じたプロダク

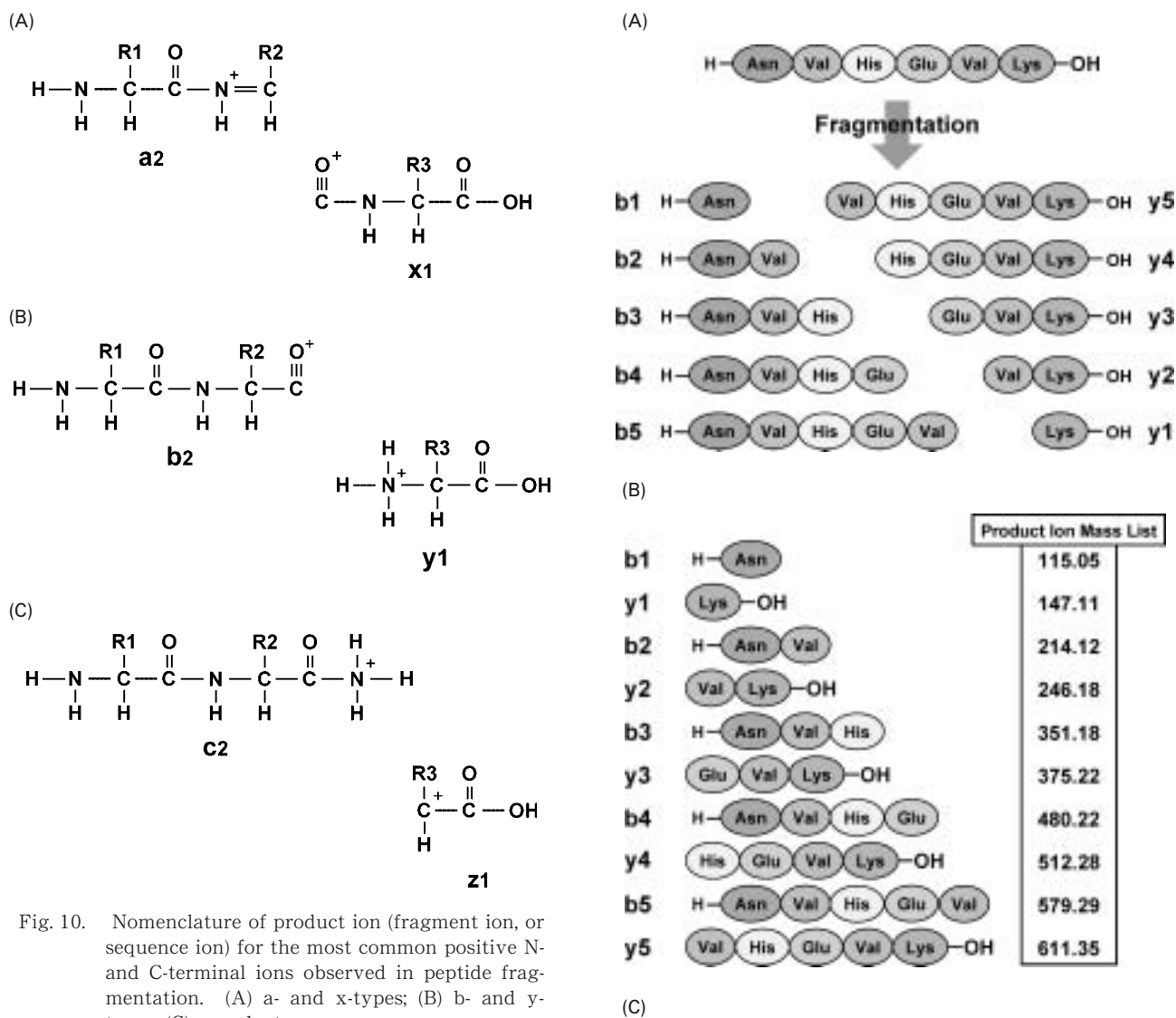


Fig. 10. Nomenclature of product ion (fragment ion, or sequence ion) for the most common positive N- and C-terminal ions observed in peptide fragmentation. (A) a- and x-types; (B) b- and y-types; (C) c- and z-types.

トイオン, すなわち b タイプと y タイプが観測される. 生成されるプロダクトイオンの質量はアミノ酸配列に依存するので, アミノ酸配列が既知であれば, 観測されるプロダクトイオンマスペクトルを予測することが可能である. 例えば, PKC α (1-158) を LEP 消化した際に得られるペプチドのうち, 30→35 残基目に位置するヘキサペプチド NVHEVK からは, 理想的には CID によって 5 種類の b type プロダクトイオンと 5 種類の y type のプロダクトイオン, 計 10 種類の断片が生じる (Fig. 11A). これらのプロダクトイオンの質量は理論的に計算することが可能であり, この理論質量のリストが, ヘキサペプチド NVHEVK の理論プロダクトイオンマスリストとなる (Fig. 11B). この理論プロダクトイオンマスリストをマスペクトルとして表すことができる (Fig. 11C).

ペプチドは 1 次構造, すなわちアミノ酸配列に独自性があり, 異なるペプチドは, 当然のことながら異なるアミノ酸配列を有している. したがって, ペプチドのアミノ酸配列が異なれば, フラグメンテーションによって生じるプロダクトイオン群のアミノ酸配列の組み合わせが固有のもの

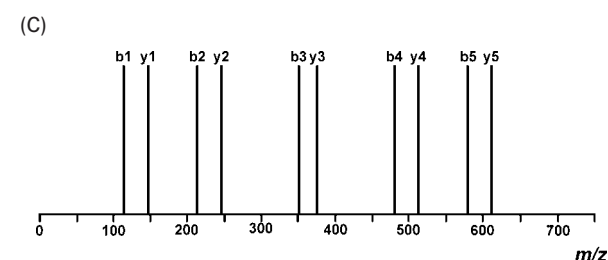


Fig. 11. Calculated product ion mass list and spectrum of a hexapeptide NVHEVK obtained by digestion of PKC α with LEP. (A) Five b-type product ions and five y-type product ions observed in theoretical fragmentation of the peptide NVHEVK; (B) The calculated product ion mass list of the peptide; (C) The calculated product ion mass spectrum.

となる. 個々のプロダクトイオンのアミノ酸配列が異なれば, イオンの質量も異なることから, フラグメンテーションによって生じるプロダクトイオン群の質量の組み合わせはペプチドに固有のものとなる. このプロダクトイオン群の質量の組み合わせが「プロダクトイオンマスリスト」である. したがって, このプロダクトイオンマスリストをマスペクトル表記したもので, すなわちプロダクトイオンマス

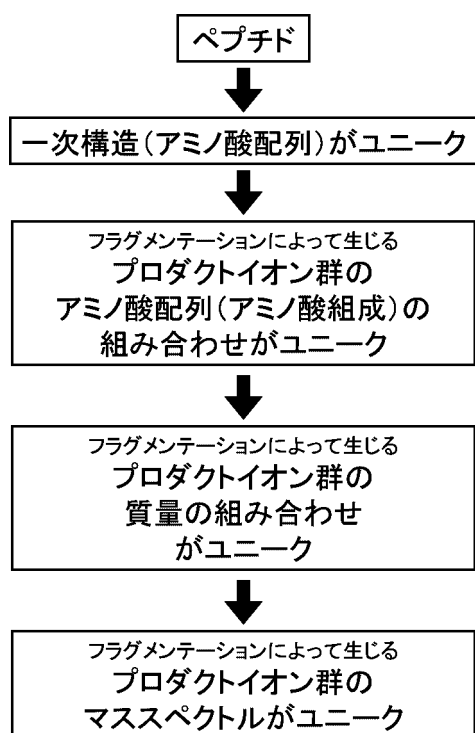


Fig. 12. Principle of identification of peptide by product ion mass fingerprinting.

ペクトルもペプチドに固有のものとなり (Fig. 12), ペプチドのプロダクトイオンマススペクトルは, 正に「ペプチドの指紋 (Fingerprint)」とも言うべき特徴である。

アミノ酸残基特異的なプロテアーゼを用いて消化した際に得られる任意の断片ペプチドのアミノ酸配列が, 当該タンパク質に固有のものであれば, その断片ペプチドを MS/MS 解析した際に得られるプロダクトイオンマススペクトルもタンパク質に固有のものとなり, 原理的には, 1 種類の断片ペプチドのプロダクトイオンマススペクトルから 1 種類のタンパク質分子を一義的に特定することが可能となる。すべての断片ペプチドが, それぞれのタンパク質に固有の配列をもっているとは限らないので, 必ずしも 1 種類の断片ペプチドのプロダクトイオンマススペクトルから 1 種類のタンパク質を一義的に特定できないこともあるが, 多くの場合, 1 種類の断片ペプチドからの情報によって, ある程度のタンパク質の種類を特定することは可能である。またタンパク質を消化した際には, 同一の試料に複数の断片ペプチドが含まれていることから, 他の断片ペプチドから得たプロダクトイオンマススペクトルの情報によってさらに詳細にタンパク質分子を限定することが可能となり, 断片ペプチドからの情報が増えれば増えるほど同定の精度は上昇する。

5.2 Product Ion Mass Fingerprinting 法の基本手順

Product Ion Mass Fingerprinting 法の基本的な手順は次のとおりである。

- 1) アミノ酸残基特異的なプロテアーゼを用いて未知タンパク質を消化し, 断片ペプチドを得る。
- 2) 一つ以上の任意の断片ペプチドを MS/MS によ

て解析し, プロダクトイオンマススペクトルを測定する。これによって断片ペプチドのプロダクトイオンマスリストを得る。

- 3) 配列データベースに登録されているすべてのタンパク質を, 未知タンパク質の消化に使用したプロテアーゼの特異性に従いバーチャルな酵素消化を行い, 生成されるすべての断片ペプチドの理論質量を計算する。
- 4) プロダクトイオンマススペクトルが測定された断片ペプチド (プリカーサーイオン) の測定値と設定した誤差範囲内で一致する理論質量をもつバーチャルな断片ペプチドを探索し, リストアップする。
- 5) リストアップされたすべての仮想断片ペプチドについて, 理論的なプロダクトイオンマスリストを作成する。
- 6) 作成された理論プロダクトイオンマスリストの中で, 実測されたプロダクトイオンマスリストに最も適合するものを検索する。
- 7) 有意に適合した仮想断片ペプチドが由来するタンパク質を提示する。

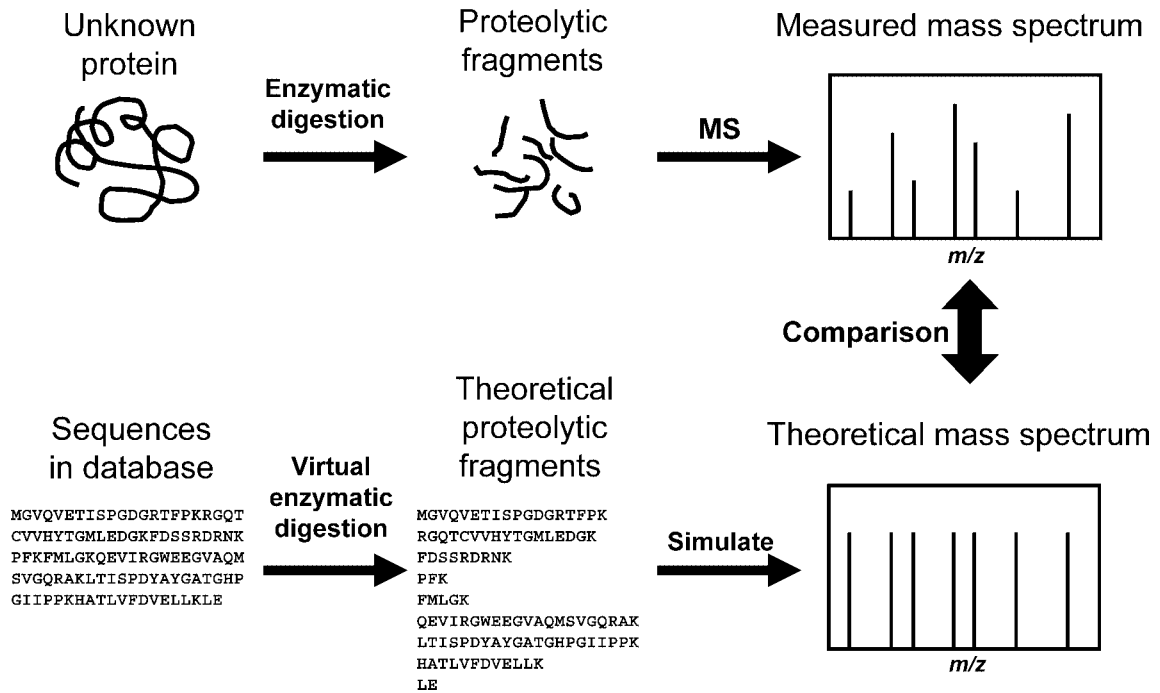
Product Ion Mass Fingerprinting 法によるタンパク質同定において, プロダクトイオンマスリストの情報から直接特定されるものはタンパク質ではなく, 断片ペプチドのアミノ酸配列である。LC/ESI-MS/MS 法や MALDI-MS/MS を用いてある未知タンパク質の同定を試みる場合, 通常は複数のプロダクトイオンマススペクトルを得ることができる。得られた複数のプロダクトイオンマスリストは, それぞれ個別に上述のアルゴリズムに従ってデータベース検索が行われ, 互いにその検索結果に依存することなく, 個別にペプチドのアミノ酸配列の同定が行われる。複数の断片ペプチドが, 同じタンパク質に由来するものであれば, データベース検索の結果マッチしたアミノ酸配列は, 当然ながらすべて同じタンパク質に由来することとなる。タンパク質の同定精度は, マッチしたプロダクトイオンマススペクトルの数に応じて相対的に上昇する。

5.3 PMF 法との比較, 同定原理の類似点

Product Ion Mass Fingerprinting 法によるタンパク質の同定原理も, その基本は PMF 法と同様である。PMF 法の場合は, 消化断片のペプチドマスリストを検索用のデータ (マスリスト) として用いているのに対し (Fig. 13A), Product Ion Mass Fingerprinting 法の場合は, CID (PSD) によって生成されたプロダクトイオンのマスリストを検索に用いている (Fig. 13B)。断片化 (Fragmentation) の手段が, 液相反応の酵素消化か, 気相反応の CID (PSD) かの違いである。ペプチド由来のイオンを開裂させて得たプロダクトイオンマススペクトル (プロダクトイオンマスリスト) が, ペプチドの指紋 (Fingerprint) であり, この指紋を照合してペプチド (タンパク質) を同定する作業は正しく Fingerprinting (指紋鑑定) そのものである。それゆえ, 筆者らは, このカテゴリーに含まれるすべてのアルゴリズムを包括する総称として “Product Ion Mass Finger-

(A)

Peptide Mass Fingerprinting



(B)

Product Ion Mass Fingerprinting

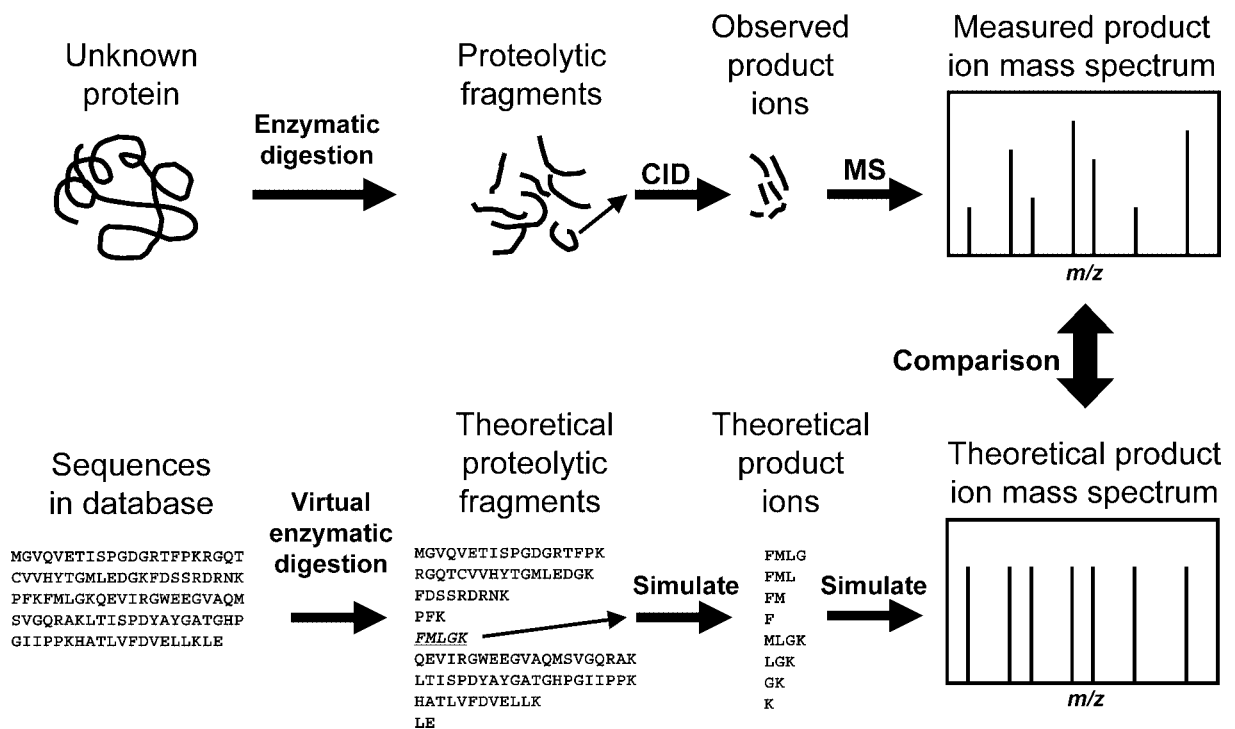


Fig. 13. Comparison between peptide mass fingerprinting and product ion mass fingerprinting in principle of protein identification. Schematic illustrations of protein identification by peptide mass fingerprinting (A) and product ion mass fingerprinting (B).

printing” 法という名称を与えた。PMF が、「ペプチドマス (Peptide Mass)」のリストを用いた Fingerprinting (指紋鑑定) であるのに対し、このカテゴリーに含まれる同定法は、「プロダクトイオンマス (Product Ion Mass)」のリストを用いた Fingerprinting である。ただし、Product Ion Mass Fingerprinting 法は、フラグメント (プロダクトイオン) のマスリストに加えて、プリカーサーイオンの質量の情報も検索に用いられる。性質の異なる 2 種類の情報による検索であることから、1 種類の情報だけで検索を行う PMF 法に比べて Product Ion Mass Fingerprinting 法の同定精度は高くなる。

5.4 Product Ion Mass Fingerprinting 法用のソフトウェア (検索エンジン)

Product Ion Mass Fingerprinting 法によってタンパク質を同定するためには、PMF 法と同様、検索エンジンが必要である。SEQUEST[®] が開発されて以来、異なる検索アルゴリズムを用いた数々のソフトウェアが開発されているが、その多くはインターネット上で公開されており、オンラインで検索することも可能である。Product Ion Mass Fingerprinting 法用の検索エンジンを利用することができるウェブサイトを Table 5 に示した。PMF 法 (Table 4) と同様、ライセンスを取得すれば in-house のサーバーを構築することが可能なソフトウェアも含まれている。残念ながら、SEQUEST[®] (TurboSEQUEST[™]) はインターネット上には公開されておらず、サーモエレクトロン社より供給されている。

5.5 検索に用いられるプロダクトイオンマスリストデータ

PMF 法に供するペプチドマスリストには、イオンの強度のデータは使用されないが、Product Ion Mass Fingerprinting 法による同定の場合、プロダクトイオンの測定質量に加え、強度データを求められることが多い (PepFrag と MS-Tag Simple では必要ない)。当然ながら、プリカーサーイオンの測定質量は、プロダクトイオンマスリストを比較させる断片ペプチドをリストアップする際に必要な情報である。SEQUEST[®] (TurboSEQUEST[™]) や

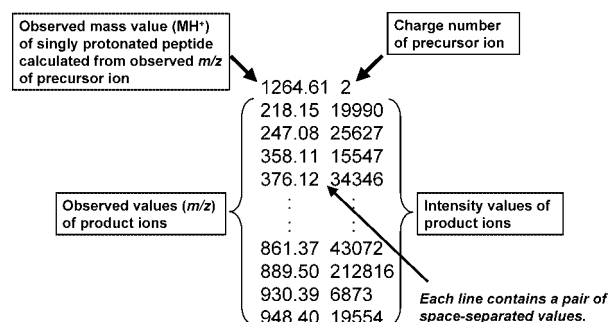


Fig. 14. An example of DTA file format.

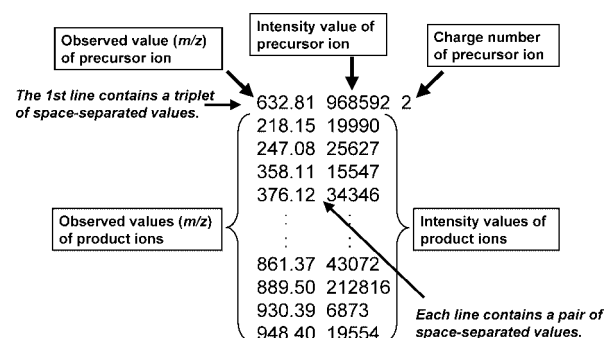


Fig. 15. An example of PKL file format.

MASCOT[®] (MS/MS Ions Search) を用いた検索の際に使用するプロダクトイオンマスリストのファイル形式 DTA file format の 1 例を Fig. 14 に示す。Micromass[®] (Waters 社) の解析用ソフトウェア MassLinx[™] (ProteinLinx[™]) によって作成される PKL file format も DTA file format とほぼ同様で、1 行目に表示されるプリカーサーイオンの情報の表記形式だけが異なる (Fig. 15)。DTA file format の場合は、プリカーサーイオンの測定値は MH⁺ の値で表記され、次にスペースが挿入されてイオンの価数が表記される。PKL file format の場合は、1 行目にマススペクトル上に観測されているプリカーサーイオンの m/z の実測値が表記され、スペースの後に続いてプリカーサーイオンの強度データ、そして同じくスペースの後

Table 5. Search Engines for Protein Identification by Product Ion Mass Fingerprinting on the Internet (as of March 2004)

Search engine	Resource	Uniform Resource Locators (URL)
MS/MS Ions Search (MASCOT [®])	Matrix Science	http://www.matrixscience.com/cgi/search-form.pl?FORMVER=2&SEARCH=MIS
MS-Tag (ProteinProspector)	UCSF	http://prospector.ucsf.edu/ucsfhtml4.0/mstag.htm
MS-Tag Simple (ProteinProspector)	UCSF	http://prospector.ucsf.edu/ucsfhtml4.0/mstagfd.htm
PepFrag (PROWL)	Rockefeller University	http://prowl.rockefeller.edu/prowl/pepfragch.html
Sonor ms/ms [™] (Knexus [™])	ProteoMetrics (Genomic Solutions)	http://65.219.84.5/service/prowl/sonar.html
Cocoozo	National Institute of Advanced Industrial Science and Technology	http://www.cbrc.jp/cocoozo/input.do

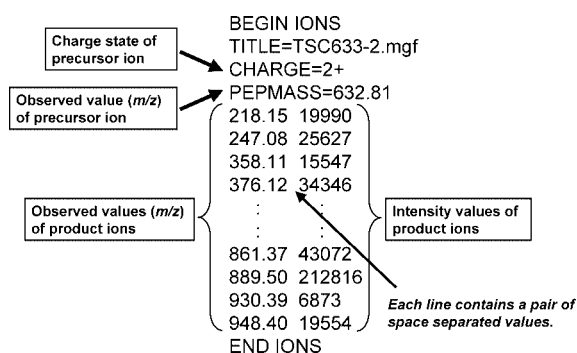


Fig. 16. An example of MGF format.

にプリカーサーイオンの価数が表記される。MASCOT[®] (MS/MS Ions Search) に用いられる Mascot generic file (MGF) format (拡張子は mgf) も基本的な入力情報は上記 2 種の形式とほぼ同じで、一つのプリカーサーイオンに由来するプロダクトイオンマスリストデータセットの始めと終わりを“BEGIN IONS”と“END IONS”とそれぞれ明記している点や、TITLE を挿入できる点、またプリカーサーイオンの m/z の実測値や価数をそれぞれ“PEPMASS=”や“CHARGE=”の後に記載する点でファイル形式がわかりやすくなっている (Fig. 16)。

ここでは、すべてプリカーサーイオンの価数を 2 価とした場合の例を示した。プリカーサーイオンの価数の値は、プリカーサーイオンの m/z 値から、 MH^+ の値を計算する際に使われるのみならず (PKL file format や MGF format の場合)、プロダクトイオンの価数の最大値を規定している値でもある。MALDI 法では問題にならないが、ESI 法を用いてペプチドをイオン化した場合、多くの場合、1 分子のペプチドに複数のプロトンが付加して生成された多価イオンからのプロダクトイオンマススペクトルが得られる。プリカーサーイオンが 2 価であった場合には、プロダクトイオンマススペクトルには 2 価と 1 価のプロダクトイオンが混在することになる。プロダクトイオンマスリストに表記される数値は m/z の値であり、この数値だけでそれぞれのプロダクトイオンの価数は判断できない。したがって、データベース検索を行う際には、1 価のプロダクトイオンの m/z と、2 価のプロダクトイオンの m/z の 2 種類のリストが混在した理論プロダクトイオンマスリストが作成され、実測されたプロダクトイオンマスリストとの比較が行われている。

分解能の高い分析計を用いてプロダクトイオンマススペクトルを取得した場合、データ解析用のソフトウェアを利用して、イオンの価数を判別し、多価のイオンから 1 価のイオンに変換し、多価イオンが混在するプロダクトイオンマススペクトルからすべてのイオンが 1 価のイオンによって構成されるプロダクトイオンマススペクトルに変換することも可能である。当然、この処理を施した方が精度の高い同定が可能となるが、この処理はコンピューターに負担を強いるプロセスであり、LC/MS/MS によって取得された大量のデータを変換する際には、コンピューターの

処理能力によっては、長い処理時間が必要となる。

5.6 Product Ion Mass Fingerprinting 法の特徴

Product Ion Mass Fingerprinting 法では、同一の試料中に混在する複数の断片ペプチドに由来する個々のプロダクトイオンマスリストは、それぞれが全く独立して非依存的にデータベース検索が行われ、個別にマッチするアミノ酸配列 (断片ペプチド) が同定される。したがって、理論的には、同一の試料中に多数のタンパク質が混在している複雑な混合物を測定した場合にも、原理的にはそれぞれのタンパク質を同定することは可能である。PMF 法を用いた場合、原理的には、2~3 種類の混合物であっても何の問題もなくタンパク質を同定することは可能ではあるが、あまり複雑な混合物を解析することには適していない。

あるタンパク質をプロテアーゼ消化した場合、ある程度の試料量があれば通常は同一タンパク質に由来する複数の異なる断片ペプチドのプロダクトイオンマスリストが 1 回の解析で得られる。その場合、それぞれのプロダクトイオンマスリストの検索は、非依行的に行われるが、結果として、いずれのプロダクトイオンマスリストの検索においても同一のタンパク質が有意な候補として提示される。したがって、それぞれのプロダクトイオンマスリストの検索結果についての確認、検証を相互に行うことが可能となる。理論的には 1 種類の断片ペプチドのプロダクトイオンマスリストからタンパク質を同定することは可能であるが、Product Ion Mass Fingerprinting 法は、上記のように任意のプロダクトイオンマスリストの解析結果を、別の解析結果によって確認、検証することが可能となる auto-confirmatory な方法である⁴⁰⁾。

これらの理由から、Product Ion Mass Fingerprinting 法は特に LC/MS/MS を利用したタンパク質の同定に適した方法である。

6. Peptide Sequence Tag (PST) 法

6.1 Peptide Sequence Tag

MS/MS 法によって取得したデータを利用して、配列データベース検索を行いタンパク質の同定を行うためのもう一つのカテゴリーは、部分的ではあるがプロダクトイオンマススペクトルの解読を必要とする方法である。この方法の代表的なアルゴリズムが Peptide Sequence Tag (PST) 法である。

ペプチドの MS/MS 解析によって得たプロダクトイオンマススペクトルから *de novo* Sequencing によってアミノ酸配列の決定を試みても、必ずしもペプチドのアミノ酸配列を完全に決定するために必要なすべてのプロダクトイオン (シーケンスイオン) が観測されるわけではない。プロダクトイオンの種類とその強度は、ペプチドのアミノ酸配列に大きく依存し、プロダクトイオンマススペクトルの質はプリカーサーイオンのイオン量にも依存している。試料の量が十分に確保できる市販品や合成ペプチドを用いた解析とは異なり、試料の量が限られた生体由来のタンパク質、ペプチドを解析する場合、測定されたプロダクトイ

オンマススペクトル上に完全なアミノ酸配列を決定するために必要な最小限のプロダクトイオンが観測されるケースはまれである。しかし、ペプチドの全配列を決定することはできなくても、数残基程度の部分配列を決定することが可能なプロダクトイオンが観測されるケースは多い。観測されたプロダクトイオンの情報だけでは、全アミノ酸配列が決定できない場合、BLAST や FASTA などの同源性検索プログラムを用い、配列データベースから解読した部分アミノ酸配列と相同な配列を検索し、その結果得られた同源性の高いアミノ酸配列の前後の配列情報を利用することによって解析したいペプチドの全アミノ酸配列を類推、決定することも行われていた⁴¹⁾。

しかしながら、FASTA や BLAST などの同源性検索を行うためには、部分アミノ酸配列とはいえ、ある程度の長さのアミノ酸配列が必要とされる。したがって、プロダクトイオンマススペクトルから 2, 3 残基分の配列しか解読できない場合には、上述の方法を利用してペプチドの全アミノ酸配列を解析し、タンパク質を同定することはできなかった。Mann と Wilm は、2, 3 残基の部分配列しか解読できない場合であっても、「解読した数残基の部分配列」、「その部分配列を解読する際に用いたプロダクトイオンの質量」、および「プリカーサーイオンの質量」の 3 種の情報を用いてデータベース検索を行うことによって、断片ペプチドのアミノ酸配列を特定し、タンパク質を同定するためのアルゴリズムを考案した³⁴⁾。彼らは 2, 3 残基の部分配列しか解読できないようなプロダクトイオンマススペクトルからでも読み取ることができる上述の 3 種の情報によって規定できる配列情報 (タグ) を “Peptide Sequence Tag” と呼んだ³⁴⁾。

Fig. 17 に PKC α をトリプシンを用いて消化した際に得られたヘキサペプチド NVHEVK を MS/MS 解析した際、b タイプのシークエンシオンがすべて観測された場合のプロダクトイオンマススペクトルを示した。同一シリーズのすべてのシークエンシオンが観測されれば、その観測質量に基づき断片ペプチドのアミノ酸配列を完全に解読することが可能である。しかしながら、前述したように、試料の量が限られている生体由来のタンパク質から得た断片ペプチドを解析した際には、Fig. 17 のように完全なアミノ酸配列を解読できる質をもったプロダクトイオンマススペクトルが得られることはまれである。シークエンシオンシリーズが部分的にしか観測されないプロダクトイオンマススペクトルが得られた場合を例に挙げる。Fig. 18 に示したプロダクトイオンマススペクトルは、Fig. 17 と同じペプチドを解析した際に得られたものである。ここでは b2, b3, b4 のシークエンシオンが観測されているが、b1 と b5 のイオンは観測されていない。このプロダクトイオンマススペクトルで観測されている m/z 480.22 のイオンと m/z 351.18 のイオンとの測定値の差 129.04 からグルタミン酸残基が、加えて m/z 351.18 のイオンと m/z 214.12 のイオンとの測定値の差 137.06 からヒスチジン残基がこのペプチドに存在していることを読み取ることがで

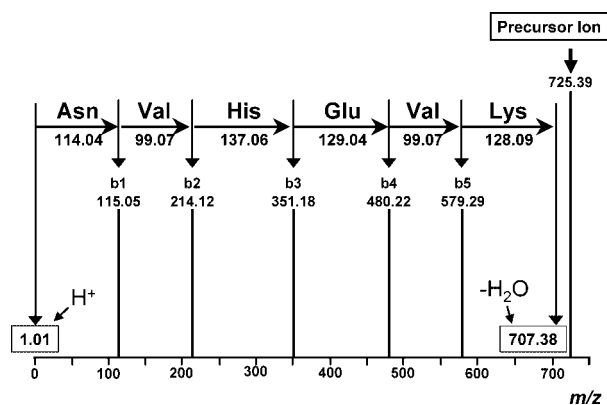


Fig. 17. *de novo* sequencing of a tryptic peptide from human PKC α . All ions observed in the product ion mass spectrum are a complete set of b-type sequence ions enough to determine the complete amino acid sequence.

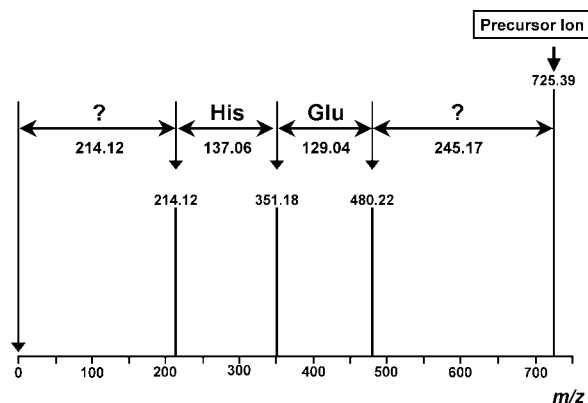


Fig. 18. An example of product ion mass spectrum in which observed product ions are not enough to determine the complete amino acid sequence.

きる。しかしながら、プロダクトイオン m/z 214.12, 351.18, 480.22 の観測質量の情報だけではこれらのイオンが b タイプであるか y タイプであるかというシークエンシオンの種類を判別することはできないので、この 2 種のアミノ酸残基の配列順序は確定できない。すなわち、従来の *de novo* Sequencing では、Fig. 18 に示したプロダクトイオンマススペクトルの情報からでは、このペプチド中に、His-Glu もしくは Glu-His の配列が存在していることだけしか解読することができない。5 種のシークエンシオンのうち、たった 2 種が観測されなかっただけで 6 残基からなるアミノ酸配列のうち、4 残基分の配列が決定できなくなり、さらに残りの 2 残基の順序すら決定できなくなってしまうのである。このような例は実際の解析において決してまれなケースではない。

Mann と Wilm は、*de novo* Sequencing によって数残基程度の部分アミノ酸配列しか解析できないプロダクトイオンマススペクトルにおいても、3 種類の「部分配列」、すなわち、[Region 1], [Region 2], [Region 3] の質量の情報を読み取ることができると考えた (Fig. 19)。どのようなペプチドであっても、そのアミノ酸配列は [Region 1]+[Region 2]+[Region 3] と表すことができ、[Region 1],

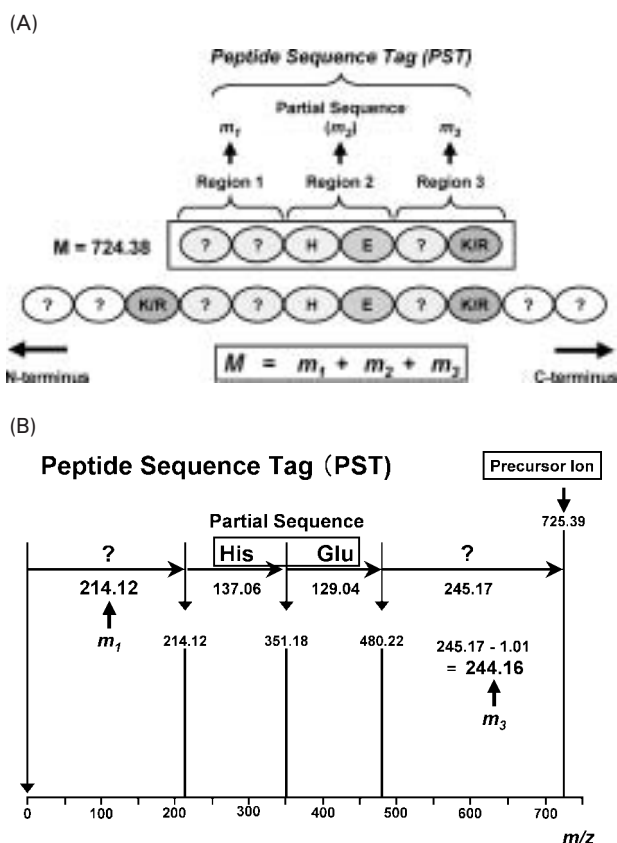


Fig. 19. (A) Schematic diagram of “peptide sequence tag” defined by Mann and Wilm. (B) Decipherment of a “peptide sequence tag” from the product ion mass spectrum shown in Fig. 18, in which observed product ions are not enough to determine the complete amino acid sequence.

[Region 2], [Region 3] の質量をそれぞれ $[m_1]$, $[m_2]$, $[m_3]$ とした場合、ペプチドの質量 M は $[m_1] + [m_2] + [m_3]$ となる (Fig. 19A)。これらの3種類の「部分配列」の情報は、それぞれ、プロダクトイオンマスペクトルを解読することによって得た $[m_1]$, [partial sequence (m_2)], $[m_3]$ の質量情報から規定することができる。

$[m_1]$ は *de novo* Sequencing によって部分配列 [partial sequence] を読み解く際に利用したシークエンシイオンのうち、シークエンシイオンが b シリーズであるならば、最も低質量のイオンの実測値と 0 との差 (= 実測値) で、この質量は読み解いた部分配列 [Partial Sequence] の N 末端側の部分アミノ酸配列 [Region 1] を規定する数値である。

[Partial Sequence] は正に、*de novo* Sequencing によって読み解いた部分配列である。

$[m_3]$ は *de novo* Sequencing によって部分配列 [Partial Sequence] を読み解く際に利用したシークエンシイオンのうち、シークエンシイオンが b シリーズであるならば、最も高質量のイオンの実測値と断片ペプチドの実測質量 (M) との差で (プロトン化分子の質量 MH^+ との差ではない)、この質量は読み解いた部分配列 [partial sequence] の C 末端側の部分アミノ酸配列 [Region 3] を規定する数

値である。

Mann と Wilm は、プロダクトイオンマスペクトルから読み取ることのできる $[m_1] + [\text{partial sequence}] + [m_3]$ からなる情報を “Peptide Sequence Tag” と名づけ、データベース検索によって断片ペプチドのアミノ酸配列を特定し、タンパク質を同定するための特異的な指標になることを示した³⁴⁾。

6.2 PST 法によるタンパク質同定の基本手順

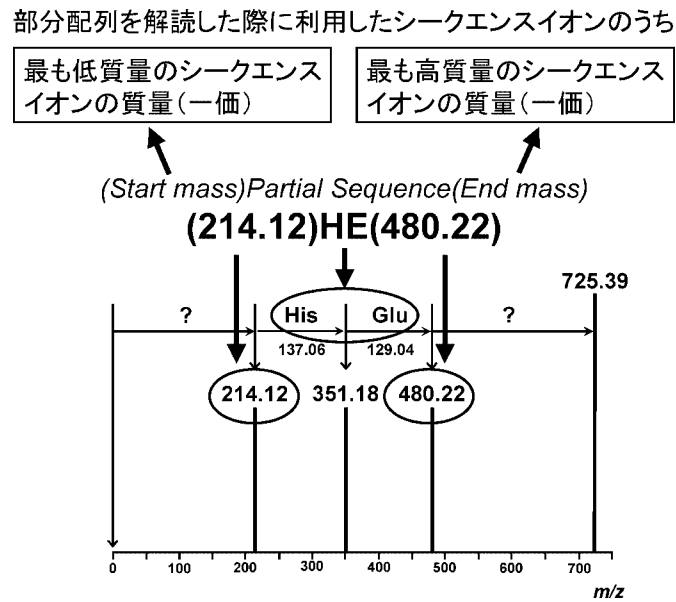
PST 法によってタンパク質を同定する際に、プロダクトイオンマスのリストの情報から読み取ったタグ情報 PST によって直接特定されるものはタンパク質ではなく、断片ペプチドのアミノ酸配列である。プロダクトイオンマスペクトルの情報からペプチドのアミノ酸配列を特定し、特定されたアミノ酸配列に基づいてタンパク質を同定するアプローチは、Product Ion Mass Fingerprinting 法によるタンパク質同定と同じである。したがって、PST 法によるタンパク質同定の基本手順は、Product Ion Mass Fingerprinting 法の手順とほぼ同じとなる。PST 法の場合は、プロダクトイオンマスペクトルを解読して、PST を読み取る必要があるが、Product Ion Mass Fingerprinting 法の場合は、スペクトルを解読する必要はないことは、これまで幾度となく述べてきた。PST 法によるタンパク質同定の基本手順は次のとおりである。

- 1) アミノ酸残基特異的なプロテアーゼを用いて未知タンパク質を消化し、断片ペプチドを得る。
- 2) 一つ以上の任意の断片ペプチドを MS/MS によって解析し、プロダクトイオンマスペクトルを測定する。
- 3) 測定されたプロダクトイオンマスペクトルから部分アミノ酸配列を解読し、その結果に基づきタグ情報 PST を読み取る。
- 4) 配列データベース上のすべてのタンパク質を、未知タンパク質の消化に使用したプロテアーゼの特異性に従いパーシャルな酵素消化を行い、生成されるすべての断片ペプチドの理論質量を計算する。
- 5) プロダクトイオンマスペクトルが測定された断片ペプチド (プリカーサーイオン) の測定値と設定した誤差範囲内で一致する理論質量をもつパーシャルな断片ペプチドを探索し、リストアップする。
- 6) 設定誤差条件を満たすすべての仮想断片ペプチドの中で、実測されたプロダクトイオンマスペクトルから読み取った PST によって規定される条件を満たすアミノ酸配列をもつ断片ペプチドを検索する。
- 7) 有意に適合した仮想断片ペプチドが由来するタンパク質を提示する。

6.3 検索に供する PST データ入力方法

PST 法によってタンパク質を同定するためには、データベース検索を行う前に *de novo* Sequencing によって部分アミノ酸配列を読み解く必要がある。この際、C 末端から、もしくは N 末端からアミノ酸配列を逐次読み取ることができない場合、(PST 法が有効なのはこうした場合で

(A)



(B)

Protein identification by peptide sequence tags

Fill out the form and press Start PeptideSearch to perform a database search.

If you want to search with sequence data produced by e.g. Edman degradation, then use the sequence only search.

Protein mass range [kDa]: 0 < Mr < 300

Cleavage agent: Lys-C

Cysteine is: Carbamidomethyl-Cys

Oxidized Methionine

Organism: Human

Peptide mass (neutral): 724.39 Monoisotopic mass

Mass accuracy: 0.005 %

Peptide sequence tag: (214.12)HE(480.22) [Syntax: (start mass)XYZ(end mass)]

Match regions: 1 and 2 and 3

Number of missed cleavage sites per peptide: 0

Pattern Match: Search by: B-type sequence ions

Fig. 20. (A) Schematic diagram of the syntax form of “peptide sequence tag” for protein identification by using “PeptideSearch” and “Sequence Query/MASCOT[®]” programs. (B) Search form of “Peptide Search” program for protein identification by peptide sequence tags.

あるが) 観測された複数のプロダクトイオンの質量の差からアミノ酸残基を特定し、部分的なアミノ酸配列を読み解くことは可能であっても、このアミノ酸配列が b シリーズのシーケンスイオンから読み取ったものか、 y シリーズのシーケンスイオンから読み取ったものかを判断できな

い場合が多い。しかし、 b シリーズか y シリーズかを判断しなくても PST 法を利用してタンパク質同定を行うことは可能である。 Mann らが開発した PST 法用のサーチエンジン PeptideSearch を用いてデータベース検索する際の PST データの入力形式は、 (Start Mass) Partial Se-

quence (End Mass) である (Fig. 20). Start Mass は, *de novo* Sequencing によって部分配列 [Partial Sequence] を読み解く際に利用したシーケンスイオンのうち, 最も低質量のイオンの実測値である. Fig. 20A の例の場合, Start Mass は 214.12 である. 一方, End Mass は, *de novo* Sequencing によって部分配列 [Partial Sequence] を読み解く際に利用したシーケンスイオンのうち, 最も高質量のイオンの実測値である. Fig. 20A の例の場合, End Mass は 480.22 である. イオンの質量の高低は, シーケンスイオンが b シリーズか y シリーズかであるかに依存することではない. 一方, Partial Sequence は, 部分配列を読み取ったシーケンスイオンが b シリーズであっても y シリーズであっても, 必ず低質量側から高質量側に向かって読み取った配列を入力する. もし, 部分配列を読み取ったイオンが y シリーズであった場合は, 結果的には通常とは逆の C 末→N 末方向の配列を入力することになる. 部分配列は, 必ずしも N 末→C 末方向に入力する必要はなく, 低質量→高質量方向で入力する. このように, PST データの入力形式には, シーケンス (プロダクト) イオンが b シリーズであるか y シリーズかであるかの情報を加える必要はない. 検索を始める前に, “Pattern Match: Search by:” のところで, Y-type sequence ions もしくは B-type sequence ions を選択する必要があるが (Fig. 20B に表示されている一番下のパラメーター), Y-type, B-type の 2 パターンの検索を試してみれば問題のないことである. 多くの場合, 結果的に異なるパラメーターを選択した方では, 有意なペプチドが同定されることはまれである.

6.4 PST 法と Product Ion Mass Fingerprinting 法の相違点

このカテゴリーに含まれる方法は, SEQUEST[®] (TurboSEQUEST[™]) や MS/MS Ions Search/MASCOT[®] に代表される Product Ion Mass Fingerprinting 法と同じ MS/MS によって得たプロダクトイオンマススペクトルを利用する同定法ではあるが, 基本的な検索アルゴリズムは下記の点で明確に異なる.

- 1) PST 法では, データベース検索の際, プロダクトイオンとブリーカーイオンの質量情報以外に, *de novo* Sequencing によって部分アミノ酸配列を読み解くなど, 得られたプロダクトイオンマススペクトルを解読した結果を検索に利用しなければなら

い. Product Ion Mass Fingerprinting 法は, プロダクトイオンマススペクトルからアミノ酸配列の解読を行う必要は一切ない.

- 2) PST 法は, 検索に供するプロダクトイオンの質量情報 (Fig. 20A の入力形式では Start Mass と End Mass に入力する質量情報) は, すべて同じシーケンスイオンシリーズに属するプロダクトイオンの情報である必要がある. 一方, Product Ion Mass Fingerprinting 法は, 検索に供するプロダクトイオンの質量情報は, 同じシーケンスイオンシリーズに属するプロダクトイオン由来である必要はない. 通常は, 複数のシリーズ (タイプ) のシーケンスイオンが混在しているマスリストを利用している.

PST 法は, プロダクトイオンマススペクトルを解読し, 部分アミノ酸配列を読み取る作業が必要であり, 迅速な検索は困難である. 近年は *de novo* Sequencing 用のソフトウェアも発達してきてはいるが, 数残基程度の部分アミノ酸配列だけしか読み取ることのできないような情報の少ないプロダクトイオンマススペクトルの場合, *de novo* Sequencing 用のソフトウェアでこの部分アミノ酸配列を正確に読み取るとは難しいように思われる. したがって, 多くの場合, マニュアルで部分アミノ酸配列を読み取る必要がある. この作業はある程度の経験が必要とされ, 解読結果に個人差が生じる部分でもある. しかしながら, 正確な解読が行われれば, 少ない情報でも, 正確に同定することが可能な有用な方法である.

PST 法は, Yates らが SEQUEST[®] の基本となったアルゴリズムを発表した年と同じ 1994 年に発表された方法であるが, 上記のような理由で, 自動化も難しく, 作業もやや難解で繁雑でもあり, スループットも低いことから, PMF 法や Product Ion Mass Fingerprinting 法と比較して利用頻度は少ない.

6.5 PST 法用のソフトウェア (検索エンジン)

PST 法は, プロダクトイオンマススペクトルから Mann と Wilm が定義した PST を読み取り, このタグ情報を入力し, 検索する方法である. この方法によってアミノ酸配列およびタンパク質を同定するための検索エンジンには, Mann らの開発した PeptideSearch がある. Sequence Query (MASCOT[®]) を用いて PST 法によるデータベースサーチを行うことも可能である. タグ情報の入力

Table 6. Search Engines for Protein Identification by PST on the Internet (as of March 2004)

Search engine	Resource	Uniform Resource Locators (URL)
PeptideSearch (Protein identification by peptide sequence tags)	EMBL	http://www.narrador.emblheidelberg.de/GroupPages/PageLink/peptidesearchpage.html
Sequence Query (MASCOT [®])	Matrix Science	http://www.matrixscience.com/cgi/search-form.pl?FORMVER=2&SEARCH=SQ
MS-Seq (ProteinProspector)	UCSF	http://prospector.ucsf.edu/ucsfhtml4.0/msseq.htm

方法は PeptideSearch と同じであるが、シーケンスイオンのタイプを指定する必要はなく、また複数の断片ペプチドのタグ情報を入力することも可能である。MS-Seq (ProteinProspector) は、タグ情報の入力方法が PeptideSearch や Sequence Query (MASCOT[®]) とは異なっている。オンラインで検索することが可能なサーチエンジンを Table 6 に示した。

7. おわりに

本稿は、質量分析法と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について、本誌「解説 (Commentary)」の投稿規定、「質量分析の分野で、比較的一般性のある特定の主題について、専門分野外の読者を対象として解説する。学生会員にも理解できるような内容と表現になるような配慮が必要である。」に沿った形で解説した。専門家にとっては常識であることに対しても、初歩的な説明を加え、またなるべく一般化した説明を心がけたため、単純に記述しすぎる箇所や、簡易すぎて冗長な部分があることをお許しいただきたい。筆者らは数年前から学部学生の実習に「PMF によるタンパク質の同定」を導入している。図もふんだんに盛り込ませていただき、本稿を実習の副読本として利用することも念頭に置いた。

本稿で紹介したソフトウェア、検索エンジンは、インターネット上に公開されていて、誰もが気軽に試すことのできるものとどめ、特定の機器メーカーの解析ソフトウェアに付属しているツールや、ベンチャー企業が販売しているソフトウェアなど、ライセンスを取得してのみ使用が許されるソフトウェアの名前は必要がない限り割愛した。このようなソフトウェアにも優れたものが多いので、別の機会に紹介させていただきたい。また本稿では、方法論を一般化して説明するために、各々のソフトウェアの特徴に関してはほとんど触れてはいないが、どれ一つ同じソフトウェアはなく、すべてのソフトウェアは開発者の意図を反映した独自の特徴をもっている。

プロテオミクスや昨今のプロテインマスマスペクトロメトリーは、多くの点で情報科学といえる要素をもっており、より高い同定精度を目指して既存のアルゴリズムも改良され、進歩し続けている。また、新しいアルゴリズムも考案され、そのアルゴリズムに基づき新しいソフトウェアも次々に開発されている。最近では、米国ハーバード大学の Gygi らが、ペプチドの MS/MS 解析において生成されたプロダクトイオンの強度比とアミノ酸配列との相関をソフトウェア自らが学習し、プロダクトイオンの質量に加えて、強度比データも利用したマスマスペクトルの照合を行うことによって断片ペプチドのアミノ酸配列を特定し、タンパク質を同定する検索アルゴリズムを考案した⁴²⁾。検索アルゴリズムの開発の歴史や、最近開発された有用な検索アルゴリズムの紹介、これまでに開発された検索エンジンの比較など各論的な解説も別の機会に譲りたい。

近年、多くの研究機関に質量分析計が設置され、身近な分析機器の一つになった。生命科学研究を推進するうえ

で、今や質量分析法は不可欠な基盤技術に成長した。質量分析計が普及すること自体はたいへん好ましいことである。質量分析計の操作法も簡便になり、配列データベースも充実し、各種のソフトウェアも発達したことから、タンパク質のプロテアーゼ消化物を質量分析し、タンパク質を同定するという作業は、すでに一部の専門家だけが行うものではなくなっている。機器の維持管理さえしっかり行われていれば、同定までの一連の作業自体は決して難しいものではない。マニュアル書も数多く出版されており、ある程度の質をもった結果であれば、比較的簡単に出すことが可能となっている。それゆえか、マスマスペクトルの測定やデータベース検索の作業をブラックボックスの中に閉じ込めてしまいがちになっている。しかし、これでは科学とは言えない。観えるべきものも観えてはこない。データベース検索用のソフトウェアは、質量分析のデータを、アミノ酸配列やタンパク質の名前に変換するだけのものであり、マスマスペクトルの質の悪さをカバーするものではない。ましてや、タンパク質の同定を行う人間の犯したミスのカバーするものでもない。ディスプレイに映し出されるタンパク質の同定結果の質は、解析に供したタンパク質試料の質、質量分析のデータの質、ソフトウェアを利用する人間の知識や経験の質を越えることは決してない。タンパク質を解析し同定する作業は、プロテオミクスや他の生命科学研究を進める過程において、多くの場合、研究の初期段階に行われる過程である。したがって、タンパク質の同定結果は、その後の研究の成否を左右する極めて重要な結果であることを認識しなければならない。不幸にも誤って真実とは異なるタンパク質を同定した場合、その誤った同定結果に基づいて行われた機能解析などの実験結果が全く無意味なものとなる。このような「誤認逮捕」や「冤罪」を未然に防止するためには、タンパク質の試料を調製する過程から、マスマスペクトルの測定、ソフトウェアを用いたデータベース検索までのすべての過程において、より質の高い結果を得るための努力を惜しまず、常に細心の注意を払い作業を行わなければならない。特に、タンパク質の指紋 (Fingerprint) を採取する過程、すなわちタンパク質のプロテアーゼ消化物を質量分析する過程において、高精度、高感度、高分解能のマスマスペクトルを得ることは最も大切である。事件の犯行現場に残された犯人の指紋の採取方法 (質量分析) や、指紋の照合方法 (データベース検索) が適切でない場合、犯人 (タンパク質) を特定できないか、もしくは無実の人 (無関係のタンパク質) を逮捕 (同定) してしまうことになる。質量分析やデータベース検索をブラックボックスの中に入れてしまうことはたいへん危険なことである。マスマスペクトルの生データやデータベース検索の結果の検証を常に心がけることを勧める。

インフルエンザ菌ゲノムの全塩基配列を決定した Venter が、この業績を成就させた舞台となった TIGR を設立したのは 1992 年であり、翌 1993 年、インフルエンザ菌のゲノムシーケンスプロジェクトを開始している⁴³⁾。このプロジェクトは、全生物種の中で初めてゲノム

の全塩基配列を決定したという意義に加え、後にセラ・ジェノミクス社がヒトゲノムのシーケンシングプロジェクトに採用した Whole Genome Shotgun Sequencing 法を利用して初めてゲノムの全塩基配列を決定したという意義も忘れてはならない。このプロジェクトの成果発表は1995年まで待たなくてはならないが、Venterらがこのプロジェクトを開始した1993年、五つの研究グループが、独自にPMF法の原理を発表した^{25)~29)}。翌1994年、Yatesらは、Product Ion Mass Fingerprinting 法の原理、SEQUESTTMの基盤となったアルゴリズムを発表し³⁰⁾、同年、MannとWilmはPST法を発表している³⁴⁾。そして、同じく1994年、イタリアのシエナで開催されたThe 1st Siena 2D Electrophoresis MeetingでオーストラリアのWilkinsが“Proteome”の概念を初めて提唱している⁴⁴⁾。プロテオミクスが注目され始めてからも久しく、2004年3月現在で、“Proteomics,” “Molecular & Cellular Proteomics,” “Journal of Proteome Research,” “Proteome Science”とプロテオミクス、プロテオームを冠した4誌の専門誌が発行されている。現在、質量分析法を利用してタンパク質の同定を行う際のイオン化法としては、専らMALDI法とESI法の2種類のイオン化法が使われているが、この2種類のイオン化法の開発に大きく貢献した2名の研究者に2002年ノーベル化学賞が授与されるに至った²⁴⁾。Venterらがインフルエンザ菌のゲノムシーケンシングプロジェクトを開始してから、そしてMann²⁵⁾、Henzel²⁶⁾、Yates²⁷⁾、Pappin²⁸⁾、James²⁹⁾らがPMF法の原理を発表してからちょうど10年目のことであった。今や、プロテオミクスは大きな樹に成長し、生命科学の主要な研究領域の一役を担っているが、若い研究者たちによってこのプロテオミクスという巨樹の「種」がまかれたのは、ほんの十数年前のことだったのである。PMF法の考案者の一人、Henzelは、この時期の一連の「発見」を“Mass Spectrometry Evolves, 1992”と表現している⁴⁵⁾。

文 献

- R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, et al., *Science*, **269**, 496 (1995).
- J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al., *Science*, **291**, 1304 (2001).
- International Human Genome Sequencing Consortium, *Nature*, **409**, 860 (2001).
- D. M. Desiderio, Jr., R. Burgus, T. F. Dunn, W. Vale, R. Guillemin, and D. N. Ward, *Org. Mass Spectrom.*, **5**, 221 (1971).
- J. Hughes, T. W. Smith, H. W. Kosterlitz, L. A. Fothergill, B. A. Morgan, and H. R. Morris, *Nature*, **258**, 577 (1975).
- K. Biemann, “Biochemical Applications of Mass Spectrometry,” First Supplementary Vol., ed. by G. R. Waller and O. C. Dermer, Wiley Interscience, NY (1980), Chap. 15, pp. 469–525.
- H. D. Beckey, *Int. J. Mass Spectrom. Ion. Phys.*, **2**, 500 (1969).
- Y. Shimonishi, Y.-M. Hong, T. Matsuo, I. Katakuse, and H. Matsuda, *Chem. Lett.*, 1369 (1979).
- Y. Shimonishi, Y.-M. Hong, T. Kitagishi, T. Matsuo, H. Matsuda, and I. Katakuse, *Eur. J. Biochem.*, **112**, 251 (1980).
- Y. Shimonishi, Y.-M. Hong, T. Takao, S. Aimoto, H. Matsuda, and Y. Izumi, *Proc. Japan Acad.*, **57B**, 304 (1981).
- A. E. Ashcroft, “Ionization Methods in Organic Mass Spectrometry,” The Royal Society of Chemistry, Cambridge (1997).
- M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. N. Tyler, *Nature*, **293**, 270 (1981).
- M. Barber, R. S. Bordoli, G. V. Garner, D. B. Gordon, R. D. Sedgwick, L. W. Tetler, and A. N. Tyler, *Biochem. J.*, **197**, 401 (1981).
- K. Biemann and H. A. Scoble, *Science*, **237**, 9928 (1987).
- 下西康嗣, “続生化学実験講座2 タンパク質の化学(上),” 日本生化学会編, 東京化学同人, 東京(1987), 7章, pp. 375–390.
- K. Biemann, *Biomed. Environ. Mass Spectrom.*, **16**, 99 (1988).
- K. Biemann, “Methods in Enzymology,” Vol. 193, ed. by J. A. McCloskey, Academic Press, San Diego, CA (1990), Capt. 25, pp. 455–479.
- 高尾敏文, 下西康嗣, “新生化学実験講座1, タンパク質II, 一次構造,” 日本生化学会編, 東京化学同人, 東京(1990), 7.1章, pp. 237–263.
- M. Yamashita and J. B. Fenn, *J. Phys. Chem.*, **88**, 4451 (1984).
- K. Tanaka, Y. Ido, S. Akita, Y. Yoshida, and T. Yoshida, *Proc. 2nd Japan-China Joint Symp. on Mass Spectrometry*, pp. 185–188 (1987).
- 吉田多見男, 田中耕一, 井戸 豊, 秋田智史, 吉田佳一, 質量分析, **36**, 59 (1988).
- K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, and T. Yoshida, *Rapid Commun. Mass Spectrom.*, **2**, 151 (1988).
- M. Karas and F. Hillenkamp, *Anal. Chem.*, **60**, 2301 (1988).
- K. Markides and A. Gräslund, Advanced Information on the Nobel Prize in Chemistry 2002, The Royal Swedish Academy of Sciences, <<http://www.nobel.se/chemistry/laureates/2002/chemadv02.pdf>> (2002).
- M. Mann, P. Højrup, and P. Roepstorff, *Biol. Mass Spectrom.*, **22**, 338 (1993).
- W. J. Henzel, T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley, and C. Watanabe, *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5011 (1993).
- J. R. Yates III, S. Speicher, P. R. Griffin, and T. Hunkapiller, *Anal. Biochem.*, **214**, 397 (1993).
- D. J. C. Pappin, P. Højrup, and A. J. Bleasby, *Curr. Biol.*, **3**, 327 (1993).
- P. James, M. Quadroni, E. Carafoli, and G. Gonnet, *Biochem. Biophys. Res. Commun.*, **195**, 58 (1993).
- J. K. Eng, A. L. McCormack, and J. R. Yates, III, *J. Am. Soc. Mass Spectrom.*, **5**, 976 (1994).
- D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell, *Electrophoresis*, **20**, 3551 (1999).
- D. M. Creasy and J. S. Cottrell, *Proteomics*, **2**, 1426 (2002).
- K. Sato, K. Yoshino, A. A. Tokmakov, T. Iwasaki, K. Yonezawa, and Y. Fukami, “Xenopus Protocol: Cell Biology and Signal Transduction, Methods in Molecular Biology,” ed. by J. Liu, Humana Press, Totowa, NJ (2004) (in press).
- M. Mann and M. Wilm, *Anal. Chem.*, **66**, 4390 (1994).
- URL: <http://www.matrixscience.com/search-form-select.html>
- URL: <http://prospector.ucsf.edu/>

- 37) T. Masaki, M. Tanabe, K. Nakamura, and M. Soejima, *Biochem. Biophys. Acta*, **660**, 44 (1981).
- 38) T. Masaki, T. Fujihashi, K. Nakamura, and M. Soejima, *Biochem. Biophys. Acta*, **660**, 51 (1981).
- 39) P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.*, **11**, 601 (1984).
- 40) A. Ducret, I. Van Oostveen, J. K. Eng, J. R. Yates, III, and R. Aebersold, *Protein Sci.*, **7**, 706 (1998).
- 41) 吉野健一, 米澤一仁, “プロテオミクスの最新技術”, 深見泰夫編, シーエムシー出版, 大阪 (2002), 10 章, pp. 81–92.
- 42) J. E. Elias, F. D. Gibbons, O. D. King, F. P. Roth, and S. P. Gygi, *Nat. Biotechnol.*, **22**, 214 (2004).
- 43) J. Shreeve, “The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World,” Alfred A. Knopf, NY (2004).
- 44) K. L. Williams and D. F. Hochstrasser, “Proteome Research: New Frontiers in Functional Genomics, Principles and Practice,” ed. by M. R. Wilkins, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, Springer Verlag, Berlin (1997), Chap. 1, pp. 1–12.
- 45) W. J. Henzel, C. Watanabe, and J. T. Stults, *J. Am. Soc. Mass Spectrom.*, **14**, 931 (2003).

Keywords: Peptide mass fingerprinting, Protein identification, Database search, MS/MS, Peptide sequence tag, Product ion mass fingerprinting