

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4048115号
(P4048115)

(45) 発行日 平成20年2月13日(2008.2.13)

(24) 登録日 平成19年11月30日(2007.11.30)

(51) Int. Cl. F I
GO 1 N 27/62 (2006.01) GO 1 N 27/62 Z NAV

請求項の数 13 (全 58 頁)

| | | | |
|---------------|-------------------------------|-----------|--|
| (21) 出願番号 | 特願2002-553108 (P2002-553108) | (73) 特許権者 | 503230117 |
| (86) (22) 出願日 | 平成13年12月21日 (2001.12.21) | | ザ インスティテュート フォー システムズ バイオロジー |
| (65) 公表番号 | 特表2004-528533 (P2004-528533A) | | アメリカ合衆国 ワシントン 98103 |
| (43) 公表日 | 平成16年9月16日 (2004.9.16) | | -8904, シアトル, ノース 34 |
| (86) 国際出願番号 | PCT/US2001/050403 | | ティーエイチ ストリート 1441 |
| (87) 国際公開番号 | W02002/052259 | (73) 特許権者 | 502457803 |
| (87) 国際公開日 | 平成14年7月4日 (2002.7.4) | | ユニヴァーシティ オブ ワシントン |
| 審査請求日 | 平成15年11月18日 (2003.11.18) | | アメリカ合衆国ワシントン州98105, |
| (31) 優先権主張番号 | 09/748, 783 | | シアトル, イレヴンス・アヴェニュー・ノースイースト4311番スイート500 |
| (32) 優先日 | 平成12年12月26日 (2000.12.26) | | 4311 11th Avenue N. |
| (33) 優先権主張国 | 米国 (US) | | E., Suite 500, Seattle, |
| (31) 優先権主張番号 | 09/748, 793 | | WA 98105, U. S. A |
| (32) 優先日 | 平成12年12月26日 (2000.12.26) | | |
| (33) 優先権主張国 | 米国 (US) | | |

最終頁に続く

(54) 【発明の名称】 迅速かつ定量的なプロテオーム解析および関連した方法

(57) 【特許請求の範囲】

【請求項1】

質量分析方法であって、該方法が、

(i) 複数の親ポリペプチドイオンが質量分析される、作動上の親イオン走査様式と

(ii) 複数の親ポリペプチドイオンが同時にフラグメント化され、得られたフラグメントイオンが質量分析される、作動上の衝突誘導性解離 (CID) 走査様式との間で質量分析装置を連続的に変更する工程；

該親ポリペプチドイオンの質量を決定する工程；

該フラグメントイオンの質量を決定して、1以上の親ポリペプチドと関連付けられた1以上の特徴を決定する工程；該親ポリペプチドイオンの質量および該フラグメントイオンの質量を注釈されたポリペプチドインデックスと比較して、1以上の親ポリペプチドに関連する1以上のさらなる特徴を、該注釈されたポリペプチドインデックスと比較する工程；および

該親ポリペプチドイオンおよび該フラグメントイオンの決定された質量および1以上のさらなるポリペプチドの特徴に基づいて該注釈されたポリペプチドインデックスの1つ以上のポリペプチドを同定する工程

を包含することを特徴とする、質量分析方法。

【請求項2】

請求項1に記載の方法であって、同定されたポリペプチドを含むサンプル中の該ポリペプチドの量を定量する工程をさらに包含する、方法。

10

20

【請求項 3】

請求項 1 または 2 に記載の方法であって、2 以上の異なる集団のポリペプチドの相対量を測定する工程をさらに包含する、方法。

【請求項 4】

請求項 1 に記載の方法であって、前記 1 以上のさらなる特徴がアミノ酸組成を含む、方法。

【請求項 5】

請求項 1 ~ 4 のいずれか 1 項に記載の方法であって、前記 1 以上のさらなる特徴が、pI を包含する、方法。

【請求項 6】

請求項 1 ~ 5 のいずれか 1 項に記載の方法であって、前記 1 以上のさらなる特徴が、クロマトグラフィーの媒体における溶出順序を包含する、方法。

【請求項 7】

請求項 1 ~ 6 のいずれか 1 項に記載の方法であって、

前記作動上の衝突誘導性解離 (CID) 走査様式において、複数の親イオンが、イオン選択をおこなわずに同時にフラグメント化される、方法。

【請求項 8】

請求項 1 ~ 7 のいずれか 1 項に記載の方法であって、前記フラグメントイオンが、1 ppm より大きい ppm 精度で質量分析される、方法。

【請求項 9】

請求項 1 ~ 8 のいずれか 1 項に記載の方法であって、前記フラグメントイオンが、2.5 ppm 以上の ppm 精度で質量分析される、方法。

【請求項 10】

請求項 1 ~ 9 のいずれか 1 項に記載の方法であって、前記フラグメントイオンが、5 ppm 以上の ppm 精度で質量分析される、方法。

【請求項 11】

請求項 1 ~ 10 のいずれか 1 項に記載の方法であって、前記フラグメントイオンが、10 ppm 以上の ppm 精度で質量分析される、方法。

【請求項 12】

請求項 1 ~ 11 のいずれか 1 項に記載の方法であって、前記フラグメントイオンが、100 ppm 以上の ppm 精度で質量分析される、方法。

【請求項 13】

請求項 1 に記載の方法であって、前記フラグメント質量が、フラグメントイオンを生成するためのイオン選択をおこなわずに質量分析によって決定される、方法。

【発明の詳細な説明】

【背景技術】

【0001】

本発明は、国立衛生研究所での国立癌研究所によって授与された助成金番号 R33CA84698-01 のもとで、政府支援によってなされた。米国政府は、本発明において特定の権利を有する。

【0002】

(発明の背景)

本発明は、一般に、プロテオーム解析、ならびにより詳細には、タンパクの質混合物中に含まれるタンパク質を同定および/または定量する方法に関する。

【0003】

生物学的プロセスを研究するための古典的な生化学的アプローチは、プロセスを構成する特定の活性の経時的分画およびアッセイサイクルによる均質性までの精製、各単離された成分の詳細な構造分析、機能分析および調節分析、ならびに単離された成分からのプロセスの再構成に基づいている。ヒトゲノムプロジェクトおよび他のゲノム配列決定プログラムは、特定の種の完全なゲノム配列、従って、基本的には、これらの種によって潜在的

10

20

30

40

50

にコードされる全てのタンパク質のアミノ酸配列を次々に明らかにしている。生物学の歴史の中で前例のないこの情報源は、従来の研究法を増強し、そして、基本的に異なる研究概念における発展を触媒することが予想され、この情報源の1つがプロテオミクスである。

【0004】

多くの他の種のゲノムとともに、ヒトゲノム全体を配列決定する努力は、特別に成功している。多数の微生物種のゲノム(TIGR微生物データベース; www.tigr.org)は完全になり、そして、120を超える他の微生物種のゲノムが、配列決定の過程にある。さらに、真核生物のより複雑なゲノム、特に、遺伝子的によく特徴付けられている単細胞生物の *Saccharomyces cerevisiae*、ならびに多細胞種の *Caenorhabditis elegans* および *Drosophila melanogaster* のゲノムは、完全に配列決定されている。さらに、イネゲノムの「ドラフト配列」が公表されており、そして、ヒトゲノムおよび *Arabidopsis* ゲノムの完了は目前に迫っている。完全なゲノム配列がないにしても、豊富なDNA配列データベースは、公然に利用可能であり、このデータベースは、210万を超えるヒト発現配列タグ(EST)および120万を超えるマウス発現配列タグ(EST)を含む。

【0005】

ESTは、cDNAライブラリー中のクローンの系統的な単一通過(single pass)の配列決定により生成されている部分的遺伝子配列を表す、約300~約500の連続ヌクレオチド配列の伸長である。大半の生物学的プロセスの時間尺度において、顕著な進化の例外を除いては、ゲノムDNA配列は、静的なものに見なされ得、ゆえに、ゲノム配列データベースは、ライブラリーに類似の情報供給源を表す。配列データベース中の個々の配列に「機能」を割り当てるために、集中的な努力が進行中である。これは、既知の機能を有する配列のファミリーに対して、ある配列の統計学的に有意な類似性を示す、直鎖状配列モチーフまたは高次構造モチーフのコンピューター分析、または他の方法(例えば、種にわたる相同タンパク質機能の比較)によって試行される。他の方法をまた使用して、個々の配列の機能を決定し、これら方法は、アンチセンスヌクレオチド技術を使用する遺伝子ノックアウトおよび遺伝子発現の抑制のような実験的方法を含み、これらの方法は時間がかかり、ある場合においては、配列によりコードされるポリペプチドへの生物学的機能の割当を可能とするのになお十分ではない。

【0006】

プロテオームは、ゲノムにより発現されるタンパク質補体として定義されている。このいくらか制限的な定義は、プロテオームの静的な性質を示唆する。実際には、プロテオームは、発現されたタンパク質の型、それらの存在比、改変の状態および亜細胞の位置が、細胞または組織の生理学的状態に依存するので、非常に動的である。ゆえに、プロテオームは、細胞状態または細胞が遭遇する外部的条件を反映し得、そして、プロテオーム解析は、細胞状態を区別および研究するため、そして、これらを制御する分子機構を決定するためのゲノム全体にわたるアッセイとして見なされ得る。分化した細胞のプロテオームが、少なくとも5桁の大きさの発現の推定動的範囲で、数千~数万のタンパク質の異なる型からなると推定されることを考えると、プロテオーム解析の見通しは険しいようである。しかし、実際に発現されるタンパク質を同定し得る技術における急速な発展と組合せると、潜在的に発現されるあらゆるタンパク質の配列を列挙するDNAデータベースの利用性は、プロテオミクスを実際の提案にする。質量分析器は、現在のプロテオミクス技術が基づく不可欠な段階の1つである。

【0007】

定量的プロテオミクスは、細胞または組織によって発現された全てのタンパク質の定量および同定に関して、これらの系統的分析である。所定の時間での細胞、組織、生物学的液体またはタンパク質複合体で発現されたタンパク質は、その時点での細胞、組織の状態を正確に定義する。異なる状態における同一細胞型のタンパク質プロファイル間の定量的および定性的差異を使用して、各状態間での移行を理解し得る。従来は、プロテオーム解

10

20

30

40

50

析は、タンパク質を分離するための高分離能のあるゲル電気泳動（特に、2次元ゲル電気泳動）と、タンパク質を同定するための質量分析器との組合せを使用して、実施された。このアプローチは、経時的かつ単調であるが、より重要には、タンパク質の生物学的に重要なクラスが基本的に検出不可能という点で、基本的に制限される。

【0008】

従って、迅速で、効率的でかつ費用効率的な方法がプロテオーム解析に必要である。本発明は、この必要性を満足させ、そして、同様に関連した利点を提供する。

【発明の開示】

【課題を解決するための手段】

【0009】

（発明の要旨）

本発明は、ポリペプチドを同定する方法を提供する。本方法は、ポリペプチド集団からの親ポリペプチドのサブセットの質量と、親ポリペプチドのサブセットのフラグメントの質量を同時に決定する工程；決定された質量と注釈されたポリペプチドインデックスを比較する工程；および決定された質量を有する注釈されたポリペプチドインデックスの1つ以上のポリペプチドを同定する工程を含み得る。本発明はまた、ポリペプチドもしくはそのフラグメントに関連する2つ以上の特徴を決定する工程に（この特徴の1つはポリペプチドのフラグメントの質量であり、このフラグメント質量は、質量分析器によって決定される）；ポリペプチドに関連する特徴と注釈されたポリペプチドインデックスを比較する方法；および、この特徴を有する注釈されたポリペプチドインデックス中の1つ以上のポリペプチドを同定することによってポリペプチドを同定する方法を提供する。本方法はさらに、このポリペプチドを含むサンプル中の同定されたポリペプチドの量を定量する工程を包含する。

【0010】

（発明の詳細な説明）

本発明は、ポリペプチドもしくはそのペプチドフラグメントに関連する特徴を決定すること、決定された特徴とポリペプチド識別インデックス（*identification index*）とを比較すること、そして同一の特徴を有するポリペプチド識別インデックス中の1つ以上のポリペプチドを同定することによりポリペプチド集団からポリペプチドを同定する方法を提供する。本発明の方法は、プロテオーム解析に適用可能であり、そして、複雑なサンプル中の1つ以上のポリヌクレオチドの迅速かつ効率的な同定を可能にする。本方法は、ポリペプチド識別インデックスを生成する工程に基づき、このインデックスは、ポリペプチドに関連する特徴のデータベースである。このポリペプチド識別インデックスは、ポリペプチドの同定のために、サンプルからのこのポリペプチドに関連すると決定された特徴を比較するために使用され得る。さらに、本方法は、ポリペプチドを同定するためだけでなく、サンプル中の特定のタンパク質の量を定量するために適用され得る。

【0011】

ポリペプチドを同定するための本発明の方法は、サンプルポリペプチドの同定および定量の両方を含む、定量的プロテオーム解析、またはポリペプチド集団間の比較の実施に適用可能である。このような定量分析は、所望される場合、都合のよいことには、2つの別々のステージで実施され得る。第1の工程としては、試験されるサンプルに代表的な参照ポリペプチドインデックスが、例えば、本明細書中に記載されるように、研究中の種、細胞型または組織型から生成され得る。第2の工程は、未知のポリペプチドに関連する特徴と、参照ポリペプチドインデックス、または以前に生成されたインデックスとを比較することである。参照ポリペプチドインデックスは、特定のサンプル（例えば、細胞、亜細胞（*subcellular*）画分、組織、器官または生物）のポリペプチドを表す、ポリペプチド同定コードのデータベースである。サンプル中のかなり多数のポリペプチドの代表であるポリペプチド識別インデックスが、生成され得、これは本質的に、サンプル中で潜在的に発現される全てのポリペプチドを含む。従って、本発明の方法は、サンプルの特

10

20

30

40

50

定の生理学的状態（例えば、疾患状態）に関連するか、またはそれを定義するサンプル中のポリペプチドの決定を、有利に可能にする。さらに、いったんポリペプチド識別インデックスが生成されると、このインデックスを繰り返し使用して、サンプル（例えば、潜在的に疾患を有する個体由来のサンプル）中の1つ以上のポリペプチドを同定し得る。

【0012】

サンプル中のポリペプチドの定量のために、ポリペプチドは、同位体標識（例えば、 ^{12}C に対して ^{13}C 、水素に対して重水素、または ^{16}O に対して ^{18}O ）される、化学的に同一の分子と比較される。かなり多数の差次的な同位体は、本明細書中に開示されるように、MSによって区別される質量に十分な差異がある限り、組み込まれ得る。分子は、同位体的な差異以外は化学的に同一なので、これらの分子は、物理化学的には同じ挙動である。さらに、所望される場合、2つより多いサンプルは、十分な数の同位体標識（例えば、d0、d4、d8、d12）が利用可能である場合に比較され、その結果、複数のサンプルが比較され得、そして、MSによって区別され得る。定量は、安定な同位体希釈に基づく。サンプルを定量するための1つの方法は、化学的に同一であるが同位体的に異なる内部標準を有するサンプルをスパイクすることである。検量線は、サンプル中の分子の量を推定するために、同位体の希釈を用いて生成され得る。このような場合、スパイクされた分子は同一でなければならず、ゆえに、サンプル中の分子は、既知でなければならない。

【0013】

サンプル中のポリペプチドを定量するための別の簡便な方法は、ICATTMのような試薬を使用することである（Gygiら、Nature Biotechnol. 17: 994~999 (1999); WO 00/11208）。ICATTM型試薬（以下により詳細に記載する）は、アフィニティタグ、1つ以上の安定な同位体が組み込まれ得るリンカー部分、および、ポリペプチド中のアミノ側鎖（例えば、システイン）に共有結合し得る反応基を含む。ICATTM型試薬を使用する定量のために、類似のサンプルは、ICATTM型試薬の異なる同位体バージョンで処理される。サンプルは標識され得、そして、例えば、定量のための参照サンプルまたはコントロールサンプルを正規化するために、類似の標識化サンプルと比較され得る。サンプル中のポリペプチドを同定および定量するためのICATTM型試薬の使用は、図3中に図示される。ICATTM型試薬の異なる同位体バージョンで標識されたペプチドは、物理化学的に同一の挙動をするので、2つのサンプル中の同一のポリペプチドは、同時に精製するが、ICATTM型標識中の同位体差異に起因して、MSでなお区別可能である。従って、同一ポリペプチドの相対量は、容易に比較され得、そして定量され得る（Gygiら、前述、1999）。全ての他の走査は、フラグメント化、次いで、溶出ペプチドについての配列情報（MS/MSスペクトル）の記録に充てられる。このペプチドの由来である親ポリペプチドは、記録されたMS/MSスペクトルを用いて配列データベースを検索することによって、同定され得る。従って、この手順は、単一の解析で、タンパク質混合物の成分の相対量および同定を提供する。このような比較は、例えば、疾患を有するかまたは疾患を有すると疑われる個体由来のサンプル中の発現レベルと、健康な個体由来のサンプル中の発現レベルを比較することによる、参照サンプルに関連するポリペプチドの発現レベルの定量に、または法医学

【0014】

ポリペプチドの定量に有用であることに加えて、ICATTM型試薬はまた、この系の複雑性に対する制約として機能し、すなわち、ICATTM型試薬に反応性であるアミノ酸を含むポリペプチドまたはこのフラグメントのみが、このポリペプチドがアフィニティ単離されるか、または異なる同位体標識されたサンプルと一緒に比較される場合、標識され、そして特徴付けられる（Gygiら、前述、1999）。従って、ICATTM型試薬の使用は、このサンプルの複雑性の軽減を提供し得る。さらに、ICATTM型試薬で標識されるポリペプチドまたはこのフラグメントの能力（すなわち、このペプチドが反応性アミノ酸を含むかどうか）は、さらなる特徴と組合せて、このポリペプチドを同定す

10

20

30

40

50

るのに有用なポリペプチドに関連する特徴である。

【0015】

ICATTM型試薬の使用のさらなる利点は、サンプル中のポリペプチドの正体が、分析前に既知でなくてもよいことである。上記したように、同位体希釈（ここで、内部標準がサンプルにスパイクされる）は、差次的な同位体標識される化学的に同一の分子がサンプル中にスパイクされることを必要とし、ゆえに、定量されるポリペプチドまたはこのフラグメントが既知であることを必要とし、その結果、化学的に同一の同位体標識された分子が添加される。ICATTM型試薬を用いると、正確なポリペプチドまたはフラグメントの前もった知識は必要ない。さらに、サンプル中の様々なポリペプチドを特徴付けるために、様々な同位体標識された分子を合成する必要はない。

10

【0016】

親和性標識を組み込む標識試薬（例えば、ICATTM型試薬）を使用することに加えて、他の標識試薬が、ポリペプチドを含む2つの異なるサンプルを差次的に同位体標識するために使用され得る。例えば、異なる同位体を含む化学的に同一である2つの試薬が、2つのポリペプチドサンプルを共有結合により改変するために使用され得、ここで、これらの試薬は、親和性タグを含まない。従って、ICATTM型タグに関連する親和性単離工程を使用する代わりに、所望される場合は、他の単離工程が使用され得る。それにも関わらず、その差次的に同位体標識されたポリペプチドサンプルは、定量的分析のために比較され得る。例えば、d3を含む（3つの重水素を含む）メタノールでのエステル化を介するポリペプチドのメチル化に対するd0を含む（重水素を含まない）メタノールでのエステル化を介するポリペプチドのメチル化が、2つのポリペプチドサンプルを差次的に同位体標識するために使用され得る。同様に、ポリペプチド中の側鎖アミノ酸を改変するための周知の方法のいずれかが、差次的に標識される同位体（例えば、水素について重水素、C¹²についてC¹³、O¹⁶についてO¹⁸）（例えば、Glazerら、Laboratory Techniques in Biochemistry and Molecular Biology: Chemical Modification of Proteins, Chapter 3, pp. 68-120, Elsevier Biomedical Press, New York (1975); Pierce Catalog (1994), Pierce, Rockford, ILを参照のこと）とともに、同様に使用され得る。並行標識されるポリペプチドが、MSにより検出されるに十分な質量の差異を含む限り、多数のこの差次的な同位体が、組み込まれ得る。上記のようなポリペプチドの化学的改変に加えて、2つのポリペプチドサンプルが、O¹⁸標識H₂Oに対するO¹⁶標識H₂Oの存在下で、プロテアーゼ（例えば、トリプシンなど）で消化され得る。このプロテアーゼ消化反応は、切断されたペプチドに水を付加するので、同位体により差次的に標識されたH₂Oの存在下での切断が、差次的な標識を別個のポリペプチドサンプル中に組み込むために使用され得る。比較されるサンプルが、生じる標識されたポリペプチドが、差次的な同位体標識によってのみ本質的に異なるように、化学的に類似した様式で処理される限り、2つのポリペプチドサンプルを差次的に標識するために同位体標識を組み込むために有用な任意の方法が、本発明の方法において、特に定量的方法のために使用され得ることが、理解される。

20

30

40

【0017】

サンプルを定量するためのなお別の方法は、同位体の存在下でサンプルをインキュベートすることまたは天然に存在する同位体の除去を生じる培地中でインキュベートすること（例えば、Odaら、Proc. Natl. Acad. Sci. USA 96: 6591-6596 (1999)を参照のこと）によって、比較のために2つのサンプル中に同位体を代謝組み込み可能にする条件下でサンプルをインキュベートすることである。このような方法は、従来のように培養されたサンプル（例えば、微生物サンプル、または個体から得られた細胞の初代培養物）に特に有用である。従って、インビトロ方法およびインビボ方法の両方が、比較および/または定量のために、2つのサンプルを差次的に同位体標識するために使用され得る。

50

【0018】

本発明の方法は、ポリペプチドの特徴の決定に基づき、この決定は、所定の物理化学的特徴に基づいてそのポリペプチドの同定を可能にする。ポリペプチドを同定するように機能し得る物理化学的特徴の集合は、本質的には、そのポリペプチドについての「バーコード」であり、すなわち、ポリペプチド識別インデックスとして機能する参照データベースとその特徴を相関付けることに基づいてポリペプチドを独自に同定するに十分な、特徴の集合である。これらの方法は、多くの異なるポリペプチドを含む複合サンプルの迅速かつ効率的な分析のために特に有利である。このような複合サンプルは、他の方法を使用すると、時間がかかり、かつ非効率的である。従って、本発明の方法は、多数の異なるポリペプチドを含む複合サンプルを分析するために適用され得、そしてプロテオミクス適用において特に有用である。従って、本発明の方法は、そのプロテオームのポリペプチドを同定するために有利に使用され得る。このプロテオームは、細胞の代謝状態と関連付けられる、ポリペプチド発現および翻訳後修飾を反映するので、この方法はまた、正常なポリペプチド発現または疾患と関連する異常なポリペプチド発現を決定するため、診断適用において使用され得る。従って、本発明の方法は、疾患または状態を診断するために、臨床適用において使用され得る。

10

【0019】

本発明の方法は、異なるポリペプチドの複合混合物からの1つのポリペプチドの同定を可能にする、拘束 (constraining) パラメーターを有利に使用する。この拘束は、ポリペプチドの同定を単純にするために使用され得る。拘束は、例えば、ポリペプチドと関連する1つ以上のさらなる特徴を含むこと、複合混合物からのポリペプチドサブセットの同定、またはポリペプチドの複合混合物の分析を単純にするために使用され得る任意の型の拘束であり得る。従って、本発明の方法は、複合混合物 (多数のポリペプチドを含む) 中のポリペプチドのより効率的同定を提供し、この方法は、プロテオーム分析に特に有用である。

20

【0020】

ポリペプチド識別インデックスの作製および使用は、いくつかの利点を提供する。第1に、この方法は、特定の構造特徴を含むポリペプチドフラグメントの選択的単離を伴って使用され得、このことは、特定の化学試薬でのタグ化により利用され得る。「タグ化」フラグメントの親和性選択は、そのポリペプチド混合物を単純にし、その混合物を、タンパク質の単離および取扱いに使用され得る高度な変性条件 / 可溶化条件と適合性にする。フラグメントの選択的単離はまた、データベース検索を拘束する。例えば、本明細書中に開示される選択的システインタグ化は、そのペプチド混合物の複雑性を、約10分の1に減少する。

30

【0021】

本発明の方法の第2の利点は、これらの方法が、種々の実験室状況において容易に使用され得ることである。例えば、質量測定は、絶対的であり、そしてクロマトグラフパラメーターは、容易に標準化され得る。従って、本発明の方法により決定されるポリペプチド識別インデックスは、実験室間で容易に移転可能であり、そして異なる実験室により作成されたデータが、類似の条件下で作成されたポリペプチド識別インデックスと容易に比較され得る。この方法を、ネットワークを介して (例えば、Webベースの検索ツールの構築を介して)、アクセス可能にすることによって、この利点はさらに利用され得る。第3の利点は、これらの方法が、単一工程質量分析とともに実施され得る。この単一工程質量分析は、迅速で、単純で、かつ感度が良い。第4の利点は、これらの方法が、複合ポリペプチドサンプル中に存在する各ポリペプチドの比を正確に測定するために使用され得、但し、そのサンプルは、安定な同位体標識で改変されていることである。最後に、これらの方法は、本質的に無制限のサンプルキャパシティを有し、これにより、非常に量が少ないポリペプチドを分析する可能性を確実にする。そしてこれらの方法は、高いピークキャパシティを有し、これにより、非常に複雑なサンプルの分析を可能にする。

40

【0022】

50

本明細書中に開示されるように、フラグメント化の前に個々の親イオンを単離することに加えて、複数のイオンが、単一イオン選択を伴うことなく、並行してフラグメント化され得る(図2を参照のこと)。従って、このような方法の重要な利点は、MS/MSによるタンパク質同定の場合と同様に、各ペプチドについて別個にではなく、複数のポリペプチドについて並行して、パラメーターが容易に決定され得ることである。

【0023】

本発明の1つの実施形態において、ポリペプチド識別インデックスは、1つのポリペプチドと関連する特徴(特に、親イオン選択を用いてかまたは用いずに作成されたMS/MSによるフラグメントイオン質量測定(図2))を決定すること、そして必要に応じて、クロマトグラフ工程を含めることによって、作成される。これらの質量決定は、高い精度である必要はない。正確な質量は、所望される場合は、計算され得、そして特定のポリペプチドと関連する他の特徴を伴うインデックスへと編集され得る。十分な数の特徴が、このインデックスにおいてポリペプチドの同定を可能にするために決定される。これらの方法は、サンプルの生理学的状態に関するさらなる情報を提供するための定量とともに、必要に応じて有利に使用され得る。しかし、より単純な系の場合、例えば、微生物ゲノムまたはウイルスゲノム、または個体由来のより少ない数のポリペプチドを含む検体(例えば、脊髄液)の場合、サンプル中のポリペプチドの複雑性は、サンプル中のポリペプチドの定量分析が特定の適用のために十分であるに十分なほど、小さくあり得る。このように、サンプル中のポリペプチドの発現の定量決定が、特定の状態(例えば、疾患状態)と相関付けるに十分である場合、本発明の方法が、サンプル中のポリペプチドの定量的同定のために適用され得る。

【0024】

例示的ポリペプチド識別インデックスまたはデータベースが、実施例IVに記載される。質量分析法ベースのプロテオミクスの成長する領域は、細胞状態の調査からタンパク質を迅速に同定する新規な方法を組み込むことを必要とする。この目的のために、サンプル単純化方法(システイン含有ペプチド精製(Gygiら、Nat. Biotechnol. 17:994-999(1999))を含む)が、首尾良く使用されている。このことから天然に存在する拘束を使用して、混合物中のペプチド候補の可能な数が、高い質量正確性(約0.1~1ppm)がペプチドおよびその親タンパク質をまとめて同定し得る点(Goodlettら、Anal. Chem. 72:1112-1118(2000))にまで、減少され得る。システイン拘束を使用することなく、ペプチドおよびそのペプチド由来のほんのわずかなフラグメントイオンの高い質量正確性が、共溶出するペプチドの存在下でさえ、同じ結果を達成し得ることが、示されている(Masselonら、Anal. Chem. 73:1918-1924(2000))。本明細書中に開示される場合、FT-ICR質量分析計でなくTOFを使用するこれらの概念の組み合わせが、Saccharomyces cerevisiaeの抽出物について試験された(実施例IVを参照のこと)。これは、データベースの複雑性およびインソース(in-source)CIDを減少させてフラグメントイオンを生成するための、システイン拘束の組み合わせによって行われた。この方法は、ペプチドが共溶出する場合でさえ、有効であることが示された。このプロセスは、共溶出するペプチドの配列決定を反映するためのショットガン配列決定を可能にする。

【0025】

例えば、図2に示されるように、本発明の方法は、単一イオン選択の非存在下または供給源領域におけるイオン選択の非存在下で、実施され得る。例えば、Q₁、Q₂、Q₃などと示される領域は、四重極を指す。これらは、m/zに基づいて選択されたイオンを分離するための物理的手段である。しかし、選択されたイオンを分離するために適切な任意の適切な方法が、四重極の使用に加えて、本発明の方法において使用され得ることが、理解される。

【0026】

本明細書中で使用される場合、用語「特徴」とは、ポリペプチドに関して使用される場

10

20

30

40

50

合は、ポリペプチドの物理化学的特性を指す。物理化学的特性は、親ポリペプチドの物理化学的特性（例えば、分子量、アミノ酸組成、pIなど）、ならびにポリペプチドフラグメント（フラグメントイオンを含む）の物理化学的特性（これは、ポリペプチドと関連付けられ得、従って、親ポリペプチドと関連する特徴であると考えられる）を包含する。ポリペプチドの物理化学的特性はまた、その特定の物理化学的特性から生じるポリペプチドの測定可能な挙動を包含する。例えば、物理化学的特性としては、規定された条件下で特定のクロマトグラフ媒体に対して溶出する順序、および規定された条件下でポリアクリルアミドゲルにおいてポリペプチドが移動する位置が挙げられる。これらの特徴は、経験的に決定され得るか、またはそのポリペプチドに関する既知の情報（例えば、配列情報）に基づいて、推定され得る。

10

【0027】

本明細書中で使用される場合、用語「ポリペプチドに関連する特徴」とは、ポリペプチドおよび/またはそのポリペプチドの任意のフラグメントの、物理化学的特性を指す。このように、ポリペプチドに関連する特徴は、親ポリペプチドの特定の特徵、ならびにその親ポリペプチドのフラグメントの特徵を包含する。このフラグメントは、そのポリペプチドに関連し得るので、このフラグメントの特徵は、親ポリペプチドと関連した特徴であると考えられる。このような特徴は、例えば、ポリペプチド識別インデックスとの比較によって、ポリペプチドを同定するために使用され得る。

【0028】

本明細書中で使用される場合、用語「ポリペプチド」とは、2つ以上のアミノ酸からなる、ペプチドまたはポリペプチドを指す。ポリペプチドはまた、天然に存在する改変（例えば、翻訳後修飾（リン酸化、脂質化、プレニル化、硫酸化、ヒドロキシル化、アセチル化、糖質付加、補欠分子族付加もしくは補因子付加、ジスルフィド結合の形成、タンパク質分解、高分子複合体へのアセンブリなどを含む））によって、改変され得る。

20

【0029】

ポリペプチド（特に、リガンドポリペプチド）の改変はまた、天然に存在しない誘導体、化学合成により生成されたそのアナログおよび模倣物も包含し得、但し、そのようなポリペプチド改変は、親ポリペプチドと比較して類似する機能的活性を提示する。例えば、誘導体は、そのポリペプチドの化学的改変（例えば、アルキル化、アシル化、カルバミル化、ヨウ素化、またはそのポリペプチドを誘導体化する任意の改変）を包含し得る。そのような誘導体化分子としては、例えば、遊離アミノ酸が、塩酸アミン、p-トルエンスルホニル基、カルボベンズオキシ基、t-ブチルオキシカルボニル基、クロロアセチル基、またはホルミル基を形成するように誘導体化された、分子が挙げられる。遊離カルボキシル基は、塩、メチルエステルおよびエチルエステル、もしくは他の型のエステル、またはヒドラジドを形成するように、誘導体化され得る。遊離ヒドロキシル基は、O-アシル誘導体またはO-アルキル誘導体を形成するように、誘導体化され得る。ヒスチジンのイミダゾール窒素が、N-im-ベンジルヒスチジンを形成するように誘導体化され得る。標準的な20個のアミノ酸の1つ以上の天然に存在するアミノ酸誘導体（例えば、4-ヒドロキシプロリン、5-ヒドロキシリジン、3-メチルヒスチジン、ホモセリン、オルニチン、またはカルボキシグルタメート）を含むポリペプチド、およびペプチド結合により連結されていないアミノ酸を含み得るポリペプチドもまた、誘導体またはアナログとして含まれ得る。

30

40

【0030】

特に有用なポリペプチド誘導体としては、スルフヒドリル基の改変（例えば、親和性試薬（例えば、ICATTM型試薬）を結合するためのスルフヒドリル基の改変）が挙げられる。ポリペプチドの特に有用な改変としては、安定な同位体を有する部分を用いる、サンプル中のポリペプチドの改変が挙げられる。例えば、2つの異なるポリペプチドサンプルが、同位体的に異なる部分で別個に標識され得、そのような差次的に標識されたサンプルが、比較され得る。安定な同位体によるポリペプチドの改変は、サンプル中の個々のポリペプチドの相対量を定量するために特に有用である。

50

【 0 0 3 1 】

本明細書中で使用される場合、「フラグメント」とは、親ポリペプチドの、任意の短縮形態（カルボキシ末端、アミノ末端、またはその両方）を指す。従って、カルボキシ末端またはアミノ末端からの単一アミノ酸の欠失は、親ポリペプチドのフラグメントと考えられる。フラグメントとは、一般的に、N末端および/またはC末端でのアミノ酸の欠失を指すが、側鎖が除去されているがペプチド結合が残っている改変も包含する。フラグメントは、例えば、化学試薬、酵素、またはエネルギー投入を使用するポリペプチド切断によって生成された、短縮型ポリペプチドを包含する。フラグメントは、配列特異的切断事象または配列非依存性切断事象から生じ得る。ポリペプチドを切断するために一般的に使用される試薬の例としては、本明細書中に開示されるような、酵素（例えば、プロテアーゼ（例えば、トロンピン、トリプシン、キモトリプシンなど））、および化学物質（例えば、臭化シアン、酸、塩基、およびo-ヨード安息香酸）が、挙げられる。フラグメントはまた、質量分析法により生成され得る。さらに、フラグメントはまた、1つの切断事象から生じる短縮型ポリペプチドが、さらなる切断事象によりさらに短縮され得るように、複数の切断事象から生じ得る。

10

【 0 0 3 2 】

本明細書中で使用される場合、用語「ポリペプチド識別インデックス」とは、そのインデックスにおいて、他のポリペプチドを同定および識別するに十分な、ポリペプチドに関連する特徴の集合を指す。従って、ポリペプチド識別インデックスは、そのポリペプチドまたはそのフラグメントの特徴に基づいてポリペプチドを同定するための、ポリペプチド識別コードの集合である。ポリペプチド識別インデックスは、ポリペプチドに関連する推定された特徴（例えば、配列情報（例えば、ゲノム配列、cDNA配列、またはESTデータベース）に基づいて推定された特徴）に基づき得る。ポリペプチド識別インデックスはまた、経験的に決定された特徴、または推定された特徴と経験的に決定された特徴との組み合わせに基づき得る。「注釈されたポリペプチド（AP）インデックス」とは、そのインデックスにおいて、そのポリペプチドの各々について経験的に決定された少なくとも1つの特徴を含むポリペプチド識別インデックスを指し、これは、例えば、本明細書中に開示される方法によって、決定され得る。所望される場合、APインデックスは、経験的に決定された特徴、または推定された特徴と経験的に決定された特徴との組み合わせに基づき得る。注釈されたポリペプチドインデックスの使用は、翻訳後修飾により改変されたポリペプチドを同定するために特に有用であり、このようなポリペプチドは、配列データベースのみからの推定に基づいて推定不可能である特徴を有し得る。

20

30

【 0 0 3 3 】

「ポリペプチド識別サブインデックス（subindex）」とは、ポリペプチド識別インデックスのポリペプチド識別コードのうち全部よりは少なくしか含まない、ポリペプチド識別インデックスのサブセットを指す。サブインデックスは、例えば、10個のポリペプチド識別コードのポリペプチド識別インデックス（これは、インデックス全体のサブセットである）からの5個のポリペプチド識別コードを含み得る。サブインデックスの同定は、例えば、ポリペプチド識別インデックスの検索の複雑性を減少するために有用であり得、このことは、本明細書中に開示される分画方法によりポリペプチドサンプルに適用され得る複雑性の減少と類似する。従って、サブインデックスの検索は、インデックス全体を検索するために必要とされるよりも少ないコンピュータの時間しか必要としないという点で、有利であり得る。

40

【 0 0 3 4 】

本明細書中で使用される場合、用語「識別コード」は、ポリペプチドの同一性を決定し、そしてポリペプチドをポリペプチド識別インデックス中の他のポリペプチドと区別するのに十分な、ポリペプチドに関する特徴のセットをいう。識別コードは、基本的にポリペプチドを識別するために使用され得る注釈されたペプチドタグ、または「バーコード」である。

【 0 0 3 5 】

50

本発明は、ポリペプチドを同定するための方法を提供する。この方法は、ポリペプチドまたはそのフラグメントに関する2つ以上の特徴を決定する工程（特徴の1つは、ポチペプチドのフラグメントの質量であり、ここで、フラグメントの質量は、質量分析計によって決定される）；ポリペプチド識別インデックス（例えば、注釈されたポリペプチド指標）とポリペプチドに関する特徴とを比較する工程；および決定された特徴を有するポリペプチド識別インデックス中の1つ以上のポリペプチドを同定する工程、を包含する。このフラグメントは、1 ppmよりも大きいppmでの精度か、またはより低い精度（より高いppm）で決定され得る。この方法は、ポリペプチドに関する1つ以上のさらなる特徴を決定する工程、およびポリペプチド識別インデックスと各工程において決定された特徴とを比較する工程、をさらに包含し得る。必要に応じて、ポリペプチドに関する1つ以上のさらなる特徴を決定する工程、およびポリペプチド識別インデックスと各工程において決定された特徴とを比較する工程は、1回以上繰り返され得、ここで、ポリペプチド識別インデックス中の単一のポチペプチドを同定する特徴のセットが決定される。上記の方法、および本発明の他の方法は、サンプル中のポリペプチドの量を定量する工程をさらに包含し得る。さらに、この方法は、この方法は、ポリペプチドの2つ以上の異なる集団における相対的な存在量（すなわちポリペプチド混合物（例えば、異なるサンプル中のポリペプチドの集団））を測定するために使用され得る。

【0036】

ポリペプチドを同定するための本発明の方法は、ポリペプチド、またはポリペプチドのフラグメントに関する特徴を決定する工程を包含する。ポリペプチドを同定するために有用なポリペプチドに関する特徴は、再現可能に決定され得る特徴である。ポリペプチドまたはフラグメントの物理化学的特性としては、例えば、原子量、アミノ酸組成、部分アミノ酸配列、みかけの分子量、PI、および規定の条件下での特定のクロマトグラフィーの媒体における溶出の順序が挙げられる。ポリペプチドに関して決定されるこのような特徴は、ポリペプチドの同定のために使用される。ポリペプチドに関する特徴を決定するための方法は、以下により詳細に記載される。

【0037】

本発明の方法において特に有用な特徴の1つは、ポリペプチドまたはフラグメント、またはこれらのフラグメントの質量である。ポリペプチドのフラグメントは、質量分析計による質量決定プロセスの前または質量決定プロセスの間に生成され得る。それ故、ポリペプチドフラグメントの質量は、ポリペプチドサンプルの調製の間に生成されたポリペプチドのフラグメントの質量であり得るか、または質量分析計の間に生じるポリペプチド切断によって生成されるフラグメントの質量であり得る。

【0038】

本発明の方法において、ポリペプチドフラグメントの質量は、質量分析計によって決定され、そしてフラグメントイオンを生成するためのイオン選択の非存在下で有利に決定され得る。本発明の方法は、ポリペプチドまたはそのフラグメントの配列決定の必要なしでポリペプチドの同定を可能にする。ポリペプチドフラグメントの質量は、本明細書中に記載されるような、当該分野において公知の種々の質量分析法を用いて決定され得る。

【0039】

種々の質量分析計システムが、ポリペプチドを同定するために本発明の方法において用いられ得る。高い質量精度、高い感度および高い分解能を有する質量分析器としては以下が挙げられるが、それらに限定されない：マトリクスアシスティドレーザー脱離飛行時間型（MALDI-TOF）質量分析計、ESI-TOF質量分析計、およびフーリエ変換イオンサイクロトロン質量分析器（FT-ICR-MS）。MSの他の方法としては、MSおよびイオントラップを有するエレクトロスプレープロセスが挙げられる。イオントラップMSにおいて、フラグメントは、エレクトロスプレーまたはMALDIによってイオン化され、次いで、イオントラップに置かれる。次いで、トラップされたイオンは、イオントラップからの選択的放出においてMSによって別々に分析され得る。フラグメントはまた、イオントラップにおいて生成され得、そして分析され得る。本発明の方法において

10

20

30

40

50

使用され得る I C A T^{T M} 型試薬標識ポリペプチドは、例えば、M L D I - T O F システムまたは E S I - T O F システムを用いる一段階質量分析によって分析され得る。プロテオミクス適用のための質量分析計法が、記載されている (A e b e r s o l d および G o o d l e t t , C h e m . R e v . 1 0 1 : 2 6 9 ~ 2 9 5 (2 0 0 1) を参照のこと)。所望される場合、M S 法は、例えば、本明細書中に記載されるような I C A T^{T M} 型試薬または I D E N T 型試薬を用いて、親和性タグ化ペプチドの検出を可能にするように修飾され得る。

【 0 0 4 0 】

所望される場合、異なる M S 分析が、未知のポリペプチドの特徴を決定するためよりもポリペプチド識別インデックスを生成するために適用され得る。例えば、L C - M S / M S は、識別子についてのデータを収集するために使用され得、そして L C - E S I - T O F は、未知のポリペプチドの特徴の測定のために使用され得る。任意の M S 法および任意の M S 法の組み合わせは、サンプルが実質的に同様の様式で処置される限り、かつ M S 法が、異なる方法によって測定された質量の比較に適合する限り、使用され得る。

【 0 0 4 1 】

本発明の方法は、ポリペプチド分離工程に続く質量分析工程を含み得る。ポリペプチド分離工程および質量分析工程は、独立して実行され得るか、または「オンライン」分析方法と組み合わせられ得る。ポリペプチド分離技術の種々の様式は、質量分析器と組み合わせられ得る。例えば、ポリペプチドは、マイクロキャピラリー H P L C を用いたクロマトグラフィー、質量分析器と組み合わせられ得る固相抽出キャピラリー電気泳動システム、またはゲル電気泳動法によって分離され得る。組み合わせられたポリペプチド分離法および質量分析法の具体的な例としては、イオントラップ M S における動的排除と共に適用される E S I - M S / M S システムと組み合わせられたマイクロキャピラリー H P L C がある。

【 0 0 4 2 】

異なる型の質量分析計は、本発明の異なる適用方法について使用され得る。特定の適用 (例えば、ポリペプチド識別インデックスを生成するためのポリペプチドフラグメントの質量決定) について、ある方法は高い精度 (例えば、1 p p m 未満の精度) を提供する。しかし、本発明の方法は、低い精度 (すなわち、より高い p p m の分解能) の M S において有利であり、より高い精度の決定を必要とするより高価な装置を必要とせず、都合よく使用され得る。ポリペプチドの集団の高処理能力の分析を含む適用について、より低い精度の質量決定で十分であり得る。より低い精度の質量決定は、一般に質量決定のためにわずかな時間しか必要とされないため、より高いサンプル処理量を提供する。

【 0 0 4 3 】

質量決定を含む、本発明の方法は、より低い精度で都合よく実行され得る。例えば、高い精度の装置 (例えば、F T M S または F T I C R M S) が、0 . 2 p p m での精度で決定するために使用され得る (G o o d l e t t ら、A n a l . C h e m . 7 2 : 1 1 1 2 ~ 1 1 1 8 (2 0 0 0) ; M a s s e l o n ら、A n a l . C h e m . 7 2 : 1 9 1 8 ~ 1 9 2 4 (2 0 0 0))。非常に高い質量精度 (例えば、0 . 1 p p m) の使用は、制限として作用する。しかし、本発明の方法は、ポリペプチドに関するいくつかの特徴が、決定され得る点において有利である。さらなる特徴と組み合わせられる場合、質量は、より低い精度 (すなわち、より高い p p m) で決定され得る。より低い精度での質量の決定は、F T M S よりもより広範に利用可能であるより安価な M S 装置の使用を可能にする。質量決定は、1 p p m または 1 p p m よりも大きい p p m の精度で決定され得、そして 2 . 5 p p m 以上、約 5 p p m 以上、約 1 0 p p m 以上、約 5 0 p p m 以上、約 1 0 0 p p m 以上、約 2 0 0 p p m 以上、約 5 0 0 p p m 以上、またはさらに約 1 0 0 0 p p m 以上の p p m での精度で決定され得、これらの各々は、低い精度の M S 装置を必要とする。本発明の方法は、サンプル中のポリペプチドを同定するために、本明細書中に記載されるように、他の物理化学的特徴と組み合わせたより低い精度の M S 分析の使用を有利に可能にする。特定の適用のための M S 測定の精度は、例えば、使用されるサンプルおよび / または指標の複雑さに依存して、当業者によって容易に決定され得る。

【 0 0 4 4 】

ポリペプチドを同定するための本発明の方法としては、1 ppmより大きい精度でポリペプチドフラグメントの質量を決定する工程が挙げられ得る。それ故、本方法は、高い精度を有するMS法を必要としない。従って、低費用のMSシステムが、ポリペプチドを同定する方法において用いられ得る。高い処理能力形式（例えば、96ウェルプレートもしくは384スポットプレート形式）、またはアンアテンデッドオペレーションを可能にする自動注入システムに対する任意の質量分析計の適応は、サンプル処理能力を増大させるための利点である。

【 0 0 4 5 】

本発明の方法において、ポリペプチドまたはそのフラグメントの質量は、フラグメントイオンを生成するためのイオン選択の非存在下で決定され得る。タンパク質同定法のストラテジーの概観は、図1に示される。ポリペプチドは、例えば、ポリアクリルアミドゲル電気泳動を用いて必要に応じて分画化され、そしてポリペプチドは、ペプチドにさらにフラグメント化され得る。このペプチドは、必要に応じてクロマトグラフィーによってさらに分画され得る。クロマトグラフィーの画分、またはピン（図1において「*」によって示される）は、MSに供される。伝統的に、イオンまたは優先イオンは、衝突誘導性解離（CID）のための衝突セルにおいて選択される。単一イオンの選択は、図1のQ1に描写される。イオンは、図1のQ3に示されるように、選択され、次いでフラグメント化される。イオン選択の非存在下で、単一のイオンが選択される代わりに、何のイオン選択も適用されず、むしろ全てのイオンがフラグメント化され、多くのペプチドフラグメントを導く。このペプチドフラグメントは、特定の親ポリペプチドに対応するペプチドを決定するためにデコンヴォリューションされ、そしてポリペプチドのフラグメントの質量におけるこのような情報は、ポリペプチドに関する特徴となる（図4を参照のこと）。図1の下方に示されるように、フラグメントの質量は、任意の数のさらなる特徴と組み合わせられ得、そしてタンパク質識別子（例えば、配列データベース、または注釈されたポリペプチド指標）と比較され得、そしてこのポリペプチドは、これらの決定された特徴に基づいて同定される。

【 0 0 4 6 】

ポリペプチドに関する決定された特徴の1セットは、ポリペプチド識別インデックスにおいてポリペプチドに関する特徴と比較される。ポリペプチド識別インデックスは、指標で注釈された他のポリペプチドからポリペプチドを独自に同定および区別する、個々のポリペプチドに関する特徴の収集物である。ポリペプチドに関する決定された特徴のセットを、ポリペプチド識別インデックスと比較することによって、同じ特徴を共有するポリペプチド識別インデックスにおける1つ以上の識別子が、同定され得る。1つより多いポリペプチドが、同じ特徴を有すると決定される場合、さらなる制限（例えば、1つ以上のさらなる特徴の決定）が含まれ得る。ポリペプチド識別インデックスは、ポリペプチドの推定の特徴（例えば、遺伝子配列データベースから推定される1つ以上の特徴）に基づき得るか、または本明細書中に記載される注釈されたペプチドタグ指標を用いる場合、経験的に決定され得る。

【 0 0 4 7 】

注釈されたペプチド指標を生成する1つの例示的な方法としては、以下がある：タンパク質を収集する；同位体コード親和性タグ（ICATTM）型試薬をコードする同位体を用いてタンパク質を標識する；分子量によりタンパク質を分画する；タンパク質をペプチドへと（例えば、トリプシンを用いて）消化する；イオン交換によりペプチドを分離する；アフィニティークロマトグラフィーにより各イオン交換画分を精製する；LC/MS/MS（またはCE/MS/MS）により各アフィニティークロマトグラフィー画分を分析する；個々のMS/MSペプチドスペクトルのデータベース検索を介して全ての発現タンパク質を同定する；測定された物理化学的特性に基づく個々のペプチドおよび従ってそのペプチドの親タンパク質についての固有のバーコードを構築する注釈されたペプチドタグのデータベースを作成する。上記の方法、これらの工程の組合せ、これらの修飾体、また

10

20

30

40

50

はポリペプチドに関する特徴の決定を可能にするのに適切な任意の方法が、本明細書中に記載されるように、少なくとも1つの経験的に測定された特徴を含むポリペプチド識別インデックスを作製するために使用され得ることが、理解される。

【0048】

本発明の方法は、ポリペプチド識別インデックスとの比較のためにポリペプチドに関する1つ以上のさらなる特徴を決定する工程をさらに包含し得る。ポリペプチドに関する1つ以上のさらなる特徴を決定するプロセスに続いて、ポリペプチド識別インデックスを比較するプロセスは、単一のポリペプチドが、ポリペプチド識別インデックスから独自に同定されるまで、繰り返され得る。従って、さらなる制限が適用可能である場合、これらの制限は、ポリペプチド識別インデックスに対する比較によってポリペプチドを同定するために含まれ得る（図4Cを参照のこと）。

10

【0049】

ポリペプチドを同定するために十分な特徴の数は、決定された特徴のセットを、ポリペプチド識別インデックスと比較することにより当業者によって容易に決定され得る。ポリペプチド識別インデックス中の単一のポリペプチドの同定は、ポリペプチドをポリペプチド識別インデックス中の別のポリペプチドから区別するのに十分な特徴のセットを決定する工程いう。例えば、2つの決定された特徴が、ポリペプチド識別インデックス中の単一のポリペプチドと一致する場合、2つの特徴は、単一のポリペプチドを同定するために十分である、同様に異なるポリペプチドについて、3つの決定された特徴が、指標中のポリペプチドを独自に同定するために必要とされ得る。従って、ポリペプチドについて決定された特徴に基づいて、比較は、ポリペプチド識別インデックスを作成する。単一のポリペプチドが、同定される場合に、十分な数の特徴が決定されている。1つより多いポリペプチドが同定される場合、1つ以上のさらなる特徴が、単一のポリペプチドが決定された特徴と独自に一致するまで決定され得、これによってポリペプチドの同定を可能にする。それ故、特定のポリペプチド識別インデックスとの比較に基づいて、十分な数の特徴が、ポリペプチド識別インデックス中の特有のポリペプチドの同定を可能にするポリペプチドについて決定されている場合、当業者は、容易に決定し得る。

20

【0050】

本発明の方法は、より効率的なポリペプチドの同定、特に多数の異なるポリペプチドを含む複雑なサンプルにおける同定を可能にする、選択された制限を含むことに有利に基づく。この方法はまた、複雑なサンプルから複数のポリペプチドを同時に同定するために有利に使用され得る。従って、異なるポリペプチドに関する多数の特徴を決定するよりもむしろ、この方法は、繰り返しの様式で実行され得、所望される場合、ポリペプチド識別インデックス中の単一のポリペプチドを同定するために必要とされるさらなる制限を含むことを伴う。

30

【0051】

例えば、相同なポリペプチドは、高い配列同一性のセグメントを一般に有する。このようなポリペプチドは、例えば、類似の機能を有するポリペプチド、同じ核酸のスプライス改変体などから生じ得る。高い配列同一性のセグメントを有するポリペプチドは、通常いくつかの物理化学的特徴、特にポリペプチドの相同フラグメントに関する特徴を有し得る。それ故、高い程度の類似性を共有するポリペプチドは、類似または同一の関連する特徴のセットを有し得る。このような類似のポリペプチドについて、2つの異なるポリペプチドを区別するのに十分な所定の特徴のセットは、ポリペプチドが、類似の領域を有する場合に、ポリペプチド識別インデックス中の単一のポリペプチドを同定するために不十分であり得る。このような場合において、ポリペプチドに関する1つ以上のさらなる特徴が、決定され得、そしてさらなる特徴の決定が、対象のポリペプチドが、ポリペプチド識別インデックス中のポリペプチドと互いに区別され得るまで、繰り返され得る。ポリペプチドに関する特徴のセットを決定する方法、ポリペプチド識別インデックスと比較する方法、およびポリペプチド識別インデックス中の単一のポリペプチドが同定されるまで、さらなる特徴を決定する方法は、1つ以上のポリペプチドに適用され得る。

40

50

【0052】

従って、ポリペプチドを同定するために必要な場合、さらなる容器が考慮され得る。例えば、ポリペプチド識別インデックス中の1より多いポリペプチドが、所定のセットの特徴を有する場合、このポリペプチド識別インデックスの選択されたポリペプチドの同定、すなわち、インデックスまたはインデックスのサブインデックス中のポリペプチドのサブセットは、実質的に制約として機能する。従って、ポリペプチド識別インデックスに対する後の比較は、このサブインデックスに対してなされ得、これは、所望の場合、計算時間を減少させ得、そしてより効率的な比較を提供し得る。次いで、さらなる制約（例えば、さらなる特徴）が考慮され得、そしてサブインデックスと比較され得、これは、決定された特徴の全てを有するポリペプチドの数を減少させ得る。このような工程は、ポリペプチド識別インデックス中の単一のポリペプチドが同定されるまで、必要に応じて繰り返され得る。このようなアプローチは、複数のポリペプチドの同定を同時に決定する場合、有利である。なぜなら、ポリペプチドを同定するに十分な特徴だけが、決定される必要があるからである。従って、これらの方法は、不要なデータ取得に対して資源を無駄にすることなく、種々のポリペプチドの同定およびプロテオミクス分析に伴う複雑性の決定を容易に提供し得る。

10

【0053】

ポリペプチド識別インデックスを作製するための本発明の方法は、2つ以上のポリペプチドについて、ポリペプチドを同定するために使用され得る特徴のセットを決定する工程を包含する。ポリペプチド識別インデックス中のポリペプチドを独自に同定する特徴のセットは、ポリペプチド同定コードまたはポリペプチドについての「バーコード」を規定する。ポリペプチド識別インデックスは、インデックス付けされたポリペプチドに関連する種々の特徴を含み得る。ポリペプチド識別インデックスに含まれるポリペプチド特徴は、ポリペプチド質量、アミノ酸組成、部分的アミノ酸組成（例えば、特定のアミノ酸の存在）、pI、特定のクロマトグラフィー媒体での溶出の順番、および1以上のポリペプチドフラグメントの質量を含み得る。ポリペプチド識別インデックスは、さらに、ポリペプチドの同定のために適切な、アミノ酸配列、関連ポリペプチドに対する参照、データベース登録または文献、ならびに他の情報を含み得る。使用者は、どの型の情報がポリペプチドインデックスに有用であり、そして任意の生理学的特性またはポリペプチドに関連する情報を含み得るかを知っている。種々のポリペプチドについて多数の同定コードを含むポリペプチド識別インデックスは、複合サンプル中のポリペプチドを同定するために特に有用である。

20

30

【0054】

ポリペプチドを同定することに関する本発明の方法は、ポリペプチド識別インデックスとポリペプチドについて決定した特徴とを比較することに基づく。ポリペプチド識別インデックスは、市販または公に利用可能なデータベース（例えば、GenBank (www.ncbi.nlm.nih.gov/GenBank)）であり得、ここで、ポリペプチドの1以上の特徴（例えば、アミノ酸組成、ポリペプチドもしくはそのフラグメントの質量など）が予測される。さらに、ポリペプチド識別インデックスは、本明細書中に記載される方法によって決定される、実験的に決定された特徴に基づき得る。例えば、ポリペプチド識別インデックスは、実験的に決定された特徴、または予測された特徴および実験的に決定された特徴の組み合わせであり得る。例えば、本明細書に開示される注釈付ポリペプチド（AP）インデックスはまた、注釈付ペプチドタグ（APT）インデックスとして言及される。

40

【0055】

実験的に決定されたポリペプチドに関連する特徴のセットは、種々の方法を使用して実験的に決定され得る。ポリペプチドの同定のための例示的方法および/またはポリペプチド識別インデックスを作製するための特徴を決定するための例示的方法は、図1に示され、そして以下に記載される。この方法は、ポリペプチド同定コードを規定するために有用である。なぜなら、この方法は、ポリペプチドに関連する特徴の決定を可能にする一連の

50

工程を包含し、最後の工程は、ポリペプチドまたはフラグメントの質量を決定することである。この方法は、以下を包含し得る：(i)ポリペプチドサンプル調製；(ii)ポリペプチドタグ化；(iii)任意のポリペプチド分画；(iv)ポリペプチドフラグメント化；(v)ポリペプチドフラグメント分離；(vi)タグ化ポリペプチドフラグメントの親和性単離；(vii)高分解ポリペプチドフラグメント分離；(viii)データベース検索；および(ix)ポリペプチド識別インデックス構築（実施例Iを参照のこと）。

【0056】

ポリペプチドサンプル調製について、定量的プロテオーム分析が実行されるポリペプチドサンプルは、ポリペプチドの可溶性を維持するための標準的なプロトコールを使用して、それぞれの供給源から単離される。ポリペプチドサンプルおよびポリペプチドサンプルの調製は、以下により詳細に議論される。

10

【0057】

ポリペプチドタグ化について、サンプル中のポリペプチドは、変性され得、必要に応じて還元され得、そしてポリペプチドの化学的に反応性の基は、化学的改変試薬を用いて共有結合的に誘導体化され得る。例示的な反応性の基は、還元されたシステイン残基の側鎖を示す、スルフヒドリル基であり、これは、試薬（例えば、ICATTM（Gygiら、Nature Biotechnol. 17:994-999（1999））またはIDEnT試薬（Goodlettら、Anal. Chem. 72:1112-1118（2000））によって誘導体化され得る。他の有用な反応性の基としては、ポリペプチドのアミノ基もしくはヒドロキシル基または特定の翻訳後修飾（ホスフェート、炭水化物または脂質）が挙げられる。ポリペプチド中の化学基についての特異性を有する任意の化学反応が、この工程において適用され得る。ICATTM型試薬およびIDEnT試薬ならびに使用方法は、以下により詳細に記載される。

20

【0058】

任意のポリペプチド分画について、タグ化ポリペプチドの混合物は、任意のポリペプチド分離手順を使用して分画され得る。本発明の方法において有用な分画手順は、再現性があり、ポリペプチドを可溶性のままにし、そして高いサンプルおよびピーク容量を有する。任意の分画手順のいずれかは、低量のタンパク質について富化するため、および/または混合物の複雑性を減少させるために実行され得、そして相対的な量は維持され得る。例示的な分画方法としては、例えば、本明細書中に開示されるような、ドデシル硫酸ナトリウム-ポリアクリルアミドゲル電気泳動（SDS-PAGE）、クロマトグラフィー法（例えば、サイズ排除、イオン交換、疎水性など）が挙げられる。ポリペプチド分画方法は、以下により詳細に記載される。

30

【0059】

ポリペプチド分画について、サンプル混合物中のポリペプチド、または任意のサンプル画分が使用される場合には各画分中に含まれるポリペプチドは、配列特異的な切断（例えば、トリプシンによる切断）に供され得る。配列特異的な切断の使用は、特に有用であり得る。なぜなら、配列特異的な方法によって切断されたペプチドの末端は、制約として作用し得るからである。しかし、フラグメントを作製するために使用される切断方法は、所望の場合、配列特異的である必要はないことが理解される。配列特異的な様式でポリペプチドを切断するために有用な方法は、以下により詳細に記載される。

40

【0060】

ポリペプチドフラグメント分離について、得られたポリペプチドフラグメント混合物は、必要に応じて、1次元ペプチド分離に供され得る。高いサンプル容量を有する分離方法、少なくとも中程度の分解能および高度に再現性のある分離パターンが、この工程において有用である。1次元分離方法の例としては、アニオンおよびカチオンのイオン交換クロマトグラフィーが挙げられる。これらおよび他のクロマトグラフィー方法は、以下により詳細に記載される。ポリペプチドフラグメント分離は、必要に応じて実行され得るが、この方法は有利に使用され得、その結果、ペプチドフラグメントの特徴は「バルク」で測定

50

される、すなわち、この方法は、ペプチドフラグメントを均質にまで精製することを必要としない。

【0061】

タグ化ポリペプチドフラグメントの親和性単離について、ポリペプチドフラグメントは、このポリペプチドタグに結合する親和性試薬を使用して、各クロマトグラフィー画分から単離され得る。例えば、本明細書中に例示される I C A TTM 試薬を用いてタグ化されたポリペプチドフラグメントは、アビジンまたはストレプトアビジン親和性クロマトグラフィーを使用して単離され得る。I C A TTM 標識されたポリペプチドフラグメントの単離のために有用な親和性媒体の例は、ポリマービーズ上に固定化されたモノマーアビジンである。ビオチンとは異なる親和性タグと共に I C A TTM 型試薬を使用する場合、親和性タグに結合する、対応する親和性媒体が使用される。本明細書中に開示されるように、親和性単離分離は、例えば、ビーズが通過するのを防ぐ膜、フリットまたは他のデバイスを含む上側のチャンバのウェル中に親和性ビーズを有するマイクロタイタープレートを使用して、並行して実施される得る。サンプルは、上側のチャンバに添加され、そして親和性ビーズと共にインキュベートされ、洗浄されて非特異的結合を除去し、次いで特異的結合剤が、固体の底を有する第二のマイクロタイタープレート中に溶出する。

10

【0062】

高分解ポリペプチドフラグメント分離について、液体クロマトグラフィー E S I - M S / M S が使用され得る。親和性クロマトグラフィーカラムから溶出されたポリペプチドフラグメント混合物は、分離方法としてキャピラリー逆相クロマトグラフィー (Y a t e s 20
ら、Methods Mol. Biol. 112: 553 - 569 (1999)) を使用し、そして質量分析法として動的排除を用いるデータ依存的 C I D (G o o d l e t t r a
、前出、2000) を使用する自動化 L C - M S / M S によって、個々に分析され得る。

【0063】

全ての親和性画分の同時の高スループット分析について、高分解分離が並行して実行され得る。このためのデバイスは、並行分離キャピラリー (各々、圧電ポンプで終わる C E
、S P E - C E、H P L C が挙げられるが、これらに限定されない)、または M A L D I
M S サンプリングプレート上で迅速にサンプリングし得る他のデバイスからなる。各圧電ポンプまたは類似のデバイスは、分離からの分解を維持し、そして非常に小さいスポットでサンプルを堆積させることによって、M A L D I - T O F の感度の増大を提供する、
迅速なサンプリングを行い得る。各分離からの溶出物は、M A L D I 標的上に同時に堆積し得るか、またはマイクロタイタープレート中にも堆積され得る。

30

【0064】

データベース検索について、適切な C I D スペクトルが得られたポリペプチドフラグメントの配列は、調査中の種由来の配列データベースを検索することによって決定される。配列データベース検索プログラム (例えば、S E Q U E S T (E n g , J . ら、J . A m . S o c . M a s s . S p e c t r o m . 5 : 9 7 6 - 9 8 9 (1 9 9 4)) または類似の能力を有するプログラムは、データベースを検索するために有利に使用され得る。

【0065】

ポリペプチド識別インデックスの構築について、上記の手順によって同定された全てのペプチドの配列は、データベースに登録され得、そして上記工程の間に生じた特徴と共に注釈付けされ得る。これらの属性としては、例えば、部分的アミノ酸組成、親ポリペプチドの近似分子量 (これは、例えば、任意の分画工程によって決定され得る)、第一のクロマトグラフィー工程からの溶出の順序、第二のクロマトグラフィー工程からの溶出の順序などが挙げられ得る。

40

【0066】

集合的に、ポリペプチド識別インデックス中の各ポリペプチドフラグメントを識別する、十分な数の特徴が決定され得る。ポリペプチドを独自に同定する特徴の収集は、「バーコード」またはポリペプチド同定コードを示す。未知のポリペプチドに関連する特徴は、引き続いて決定され得、そして以前に作製されたポリペプチド識別インデックスと比較さ

50

れ得る。あるいは、ポリペプチド識別インデックスは、未知のポリペプチドと共に決定され得る。しかし、ポリペプチドに関連する特徴およびインデックス中の収集に関する情報の蓄積は、サンプルを分析する時点での必要な実験工程を最小にするために便利である。従って、引き続き実験において作製されたポリペプチドのフラグメントについて決定されたポリペプチド同定コードは、未知のポリペプチドについて新たに作製されたポリペプチド同定コードをポリペプチド識別インデックスと関連させて、未知のポリペプチドを同定することによって、サンプル中のポリペプチドを同定するために使用され得る。

【0067】

ポリペプチド識別インデックスとの比較によるポリペプチドの同定について、ポリペプチドに関連する特徴のセットは、ポリペプチド識別インデックスを作製するために本明細書に記載されるように一般に決定され得るか、あるいは等価な方法、改変方法もしくは省略方法、またはポリペプチドに関連する特徴の決定を可能にする任意の方法を使用して、決定され得る。ポリペプチドを独自に同定するために十分な特徴の数は、本明細書に開示されるように、当業者によって容易に決定され得る。これらの方法は、一般に、2以上の特徴の同定を含み、そしてインデックス中の各ポリペプチドを識別する十分な数の特徴が決定される限り、3以上、4以上、5以上、6以上、7以上、8以上、9以上、10以上、15以上、20以上、30以上またはなお50以上の特徴、あるいは任意の数の特徴を含み得る。ポリペプチド識別インデックス中に含める特徴の数は、インデックスの特定の使用および分析されるサンプルの複雑性に依存する。

【0068】

ポリペプチド識別インデックスを作製する際に、配列データベースを検索することによってポリペプチド配列を得るために、ポリペプチドに関連する特徴を使用し得る。例えば、必要に応じて質量分析によって決定されるポリペプチドまたはフラグメントの部分的アミノ酸配列を、ポリペプチドまたは翻訳された核酸配列データベースを検索するために容易に使用して、ポリペプチドを独自に記載する名前または配列同定数（例えば、登録番号）を同定し得る。従って、ポリペプチド識別インデックスは、ポリペプチド特徴（例えば、一般名称、公に利用可能なデータベースからの数字の同定コードもしくは英数字の同定コード、またはポリペプチド識別インデックス中のポリペプチド同定コードを同定するために選択された任意の他の同定コード）を含み得る。

【0069】

データベース中に親ポリペプチド配列も核酸配列も有さないか、または同定を困難にする予期しない翻訳後修飾を含むポリペプチドから配列情報を獲得するために、新規配列決定が実行され得る。同定されたアミノ酸配列を使用して、上記のようにポリペプチド配列または核酸配列のデータベースを検索し得る。新規配列決定は、種々の方法を使用して実行され得る。新規配列決定の特に有用な方法は、ポリペプチド同定のために作製されたMSデータセットを使用することを含む。質量分析を使用してポリペプチドを配列決定するための方法は、当業者に周知である（例えば、KinterおよびSherman、Protein Sequencing and Identification Using Tandem Mass Spectrometry、John Wiley & Sons、New York（2000）を参照のこと）。

【0070】

例えば、CIDのような方法によって決定されるポリペプチドまたはその部分に関する配列情報は、ポリペプチド識別インデックス中の特徴として含まれ得るが、本発明の方法は、ポリペプチドを同定するために未知のポリペプチドを配列決定する必要性を除外し、一方、配列情報は、所望の場合、ポリペプチド識別インデックスを作製する際に含まれ得る。従って、ポリペプチド識別インデックスは、ポリペプチドを同定するために十分な特徴に加えて、ポリペプチドに関連する特徴についての情報（例えば、配列情報）を含み得る。インデックス中のポリペプチドに関連する特徴に関する情報を蓄積することによって、ポリペプチドの同定は、未知のポリペプチドについての配列情報を得ることなく容易に決定され得る。

10

20

30

40

50

【0071】

クロマトグラフィー分離を使用して、ポリペプチドの特徴を決定し得る。なぜなら、ポリペプチドの物理化学的特性は、クロマトグラフィー媒体上でのポリペプチドの挙動に反映されるからである。例えば、高度に荷電したポリペプチドは、非荷電であるかまたは反対に荷電したポリペプチドが溶出するpHおよび/または塩条件とは異なる特定のpHおよび/または塩条件下で、アニオン交換カラムまたはカチオン交換カラムから溶出される。従って、ポリペプチドに関連する特徴は、ポリペプチドがクロマトグラフィーカラムから溶出される、特定のpHおよび/または塩条件であり得る。同様に、ポリペプチドが任意の型のクロマトグラフィーカラムから溶出する条件が決定され得る。ポリペプチドがカラムから溶出される溶出の順番および緩衝液条件は、ポリペプチドインデックス中で注釈を付けられるか、またはポリペプチドインデックス中の対応する値と比較するために使用される値を割り当てられ得る。例えば、値は、規定された条件下の溶出プロフィール中の相対位置、所定のセットの条件下および流速での溶出時間、外部標準画分数もしくは内部標準に対する溶出の相対的な時間または順番、塩濃度、pH、あるいは再現性を伴って決定され得る特定のクロマトグラフィーカラム上でのポリペプチドの挙動を説明する任意のパラメータであり得る。代替的方法としては、ゲル電気泳動（例えば、等電点電気泳動法（IEF））または他の分析的電気泳動方法が挙げられる。例えば、IEFは、ポリペプチドを同定するためのバーコードの作製に有用な特徴であることが示された（実施例VIを参照のこと）。ポリペプチドを分画するための方法は、当業者に周知である（Scopes、Protein Purification: Principles and Practice、第3版、Springer Verlag、New York（1993））。クロマトグラフィー方法は、従来のクロマトグラフィー形式で使用され得るか、または（例えば、バルクまたはマルチウェル形式で）バッチ結合および溶出法として使用され得る。

10

20

【0072】

タンパク質分画工程は、ポリペプチドまたはそのフラグメントの質量分析の前のポリペプチドサンプルの複雑さの軽減およびポリペプチドに関連する特性の決定の両方のために本発明の方法において有用である。上記のクロマトグラフィーによる分画に加えて、任意の周知の分画工程が、サンプルの複雑さを軽減するために使用され得、そして/またはポリペプチドに関連する決定された特性として役立つ。例示的な分画工程として、硫酸アンモニウムのような塩析またはポリエチレングリコールもしくはポリエチレンイミンのような化学物質を用いる析出、細胞成分分画、組織分画、免疫沈降、などが挙げられる（Scopes、前出、1993を参照のこと）。分画工程は、ポリペプチド群の複雑さを軽減するために使用され得る。例えば、複雑さの軽減は、特定のアミノ酸で標識されたポリペプチドを含むポリペプチドの亜群の単離において使用され得る。組織サンプルの場合、分画は、1つ以上の特定の細胞型の単離（例えば、遠心分離技術または免疫選択によって）を包含し得る。さらに、細胞成分分画のような、他の分画工程はまた、サンプルの複雑さを軽減するためにおよび/またはポリペプチドの識別のために有用な特性を提供するために適用され得る。この分画工程は、ポリペプチドについての生物学的に重要な情報を、強力に提供し得る（例えば、ポリペプチドが、細胞小器官に位置付けられるかあるいは核タンパク質、膜タンパク質、および/またはシグナル伝達複合体の一部など）。ポリペプチドの群の複雑さを有利に軽減する任意の分画工程が、本発明の方法において適用され得る。

30

40

【0073】

ポリペプチド分画工程は、ポリペプチドに関連する特性を決定するための本発明の方法において有用である。例えば、分子量に基づくタンパク質分画方法は、ポリペプチドの分子量を決定するために使用され得る。SDS-PAGE（市販のゲル溶出または分取細胞システム（BIO-RAD））およびサイズ排除クロマトグラフィーのような方法は、ポリペプチドまたはフラグメントの見かけの分子量を決定するために使用され得る。ポリペプチドおよび/またはフラグメントの分子量は、ポリペプチド識別インデックスの中に包

50

含され得る特徴である。

【0074】

ポリペプチド識別インデックスの作成においてかまたはポリペプチドの識別のために、ポリペプチドについて決定された特定の一連の特性は、使用者により選択され得、そして、ポリペプチドサンプル、ポリペプチドサンプルを調製するために使用される方法、使用される質量分析の方法および使用者の好みに依存する。ポリペプチドの特性は、任意の一時的な指示 (order) において得られ得る。例えば、ポリペプチドの特性は、時間効率または利便性を提供する指示により集結され得るか、またはサンプル処理のために選択された特定の方法による命令として集結され得る。

【0075】

ポリペプチド指標の作成において、配列情報 (例えば、CIDによって決定される) ならびにポリペプチドの他の特性が、使用され得、そしてこの配列情報は、ポリペプチドを識別するための特定の配列を有するポリペプチドの他の特性と関連するために、特に有用である。しかし、一旦、ポリペプチド識別インデックスが、作成されると、ポリペプチドについて獲得される配列情報は、必要とされないという点でこの方法は、好都合である。代わりに、ポリペプチドを識別するために十分な他の特性 (例えば、質量および/またはペプチドの間の比率ならびに他の特性の間の比率) が、決定され得、そしてそれ自体が配列情報を含むポリペプチド識別インデックスと比較され、それによってポリペプチドを識別するために、ポリペプチドを配列決定する必要性を排除する。

【0076】

ポリペプチド識別インデックスを作成するための本発明の方法は、第1および第2のポリペプチドと関連する一連の特性を決定する工程を包含し、ここで決定された特性は、第1および第2のポリペプチドを識別するために十分である。第1および第2のポリペプチドを識別するために十分な特性とは、一意的にポリペプチドに起因され得る一連の特性のことをいい、これによってポリペプチドの同一性は、ポリペプチド識別インデックスを用いて明瞭に決定され得る。一連の特性が、1つ以上のポリペプチドによって共有される場合、ポリペプチドを別のポリペプチドから識別するためのさらなる特性が、決定される。従って、ポリペプチド識別インデックスにおいて示されるポリペプチドは、各ポリペプチドを識別する一連の特性によって、互いに区別され得る。

【0077】

ポリペプチドを識別するための本発明の方法は、ポリペプチドの群に適用され得、ここで、2つ以上のポリペプチドが、識別され、そして所望の場合、サンプルにおいて同時に、多数のポリペプチドを識別するために都合よく使用され得る。従って、この方法は、単純なまたは複雑なポリペプチドサンプルに対して使用され得る。単純なポリペプチドサンプルは、例えば、1から数種のポリペプチドを含む精製されたポリペプチドサンプルであり得る。複雑なサンプルは、例えば、数個から数百のポリペプチドまたは数千または数万個のポリペプチドを含む細胞または組織溶解産物あるいは分画であり得る。本明細書中に記載した方法を使用する、ポリペプチド特性の決定は、一連の工程 (例えば、一連のクロマトグラフィー分離) から得られる実験データの収集を必要とし得る。

【0078】

複雑なポリペプチドサンプルの分析の間に得られたデータを体系化するために有用な例示的プロセスは、情報の理論的な「容器 (bin)」への分配工程を包含する。例えば、ポリペプチドのICATTM型標識化混合物は、大きさによって特定の番号の容器に分離され得る。この容器は、クロマトグラフィーカラム (例えば、サイズ排除、イオン交換、など) またはSDSポリアクリルアミドゲルのセグメントから溶出する分画であり得る。各容器の中のポリペプチドは、配列特異的切断法によって断片化され得る。あるいは、サンプル中におけるポリペプチドの分析は、分画なしにポリペプチドの識別が可能となるようなサンプルの複雑さが十分に軽減されている限り、ポリペプチドを分画することなく実施され得る。容器の中に分画されるペプチド混合物は、さらに種々の方法によって分画され得、このような方法として、例えば、イオン交換クロマトグラフィー、ICATTM型

10

20

30

40

50

標識化ペプチドの単離に伴って使用されるような親和性クロマトグラフィー、逆相液体クロマトグラフィー、または他のクロマトグラフィー技術が挙げられる。次いで、各容器のペプチドは、イオン交換クロマトグラフィーのようなさらなる分画工程によってさらに容器に入れられ、そしてもう一度、特定の番号の容器にさらに分けられる。これらの容器の各々は、逆相クロマトグラフィーのような別の分画工程によってさらに分離され得、そしてさらに特定の番号の容器に分けられ、これらの容器の各々は、質量分析機によって分析され得る。そのため、このような方法によって分析される各々のペプチドは、5つの関連する特性を有し、これらは、例えば、システイン含量、大きさ、電荷、疎水性、および質量に基づく、5桁のポリペプチド識別コードまたは「バーコード」として、示され得る。

【0079】

多数のポリペプチドに関連する特性を指標化するための本発明の方法は、コンピュータのメモリを多量に（すなわち配列の長さの2乗）、使用する。高度なデータ構造（例えば、接尾辞木（*suffix tree*）のような）は、コンピュータメモリの必要量を減らすために使用され得る（*Gusfield, Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology, Cambridge University Press (1997)*）。接尾辞木は、配列のデータベースにおける全ての接尾辞のための小型のデータ表示法である。その純粋な形態において、これらは、2乗のメモリに代わって、比例的な時間をかけて構築され得そして比例的に保存される。接尾辞木アルゴリズムについての種々の改変およびトラバーサル（*traversal*）アルゴリズムは、計算時間の最適化およびポリペプチド識別インデックスの検索と関連するコンピュータメモリの使用のために使用され得る。

【0080】

ポリペプチドに関連する一連の決定された特性は、ポリペプチド識別インデックスと比較される。種々の検索アルゴリズムは、特性を決定するために割り当てられた値を指標において注釈付けられた値と整合させるために、使用され得る。データベース検索の有効性を増加するために役立つストラテジーは、データベースの狭小化または「束縛化」である。ポリペプチド識別インデックスと関連して使用される場合、用語「束縛化」は、特性（これは、識別されるポリペプチドの1つ以上の特性と一致する）を有するポリペプチドに対応するポリペプチド識別コードの一部を含むサブインデックスを得るための、ポリペプチド識別インデックスに適用される限定のことをいう。サブインデックスは、共通の特性を有するポリペプチドの群が、ポリペプチド識別インデックス以外から選択される場合、または、ポリペプチド識別コードに含まれる特定の特質が、指標から1つ以上のポリペプチドを除外するために使用される場合、作成され得る。共通の特性は、明確な物理化学的な特質（例えば、部分的アミノ酸配列または値の範囲内に設定された任意の他の決定された特質）であり得る。例えば、質量決定における誤差を説明する値の範囲として表わされるポリペプチドフラグメントの質量は、特定の質量のポリペプチドまたはフラグメントの部分集合を選択するための束縛として役立つ。

【0081】

ポリペプチドに関連する1つの特質（これは、データベースを束縛するために使用され得る）は、部分的なアミノ酸組成である。ポリペプチドの部分的アミノ酸組成物は、特定のポリペプチドまたはそのフラグメントに存在する少なくとも単一のアミノ酸の識別を包含する。部分的なアミノ酸配列は、例えば、ポリペプチドまたはそのフラグメントを試薬で処置することによって、得られ得、その結果、1つ以上の規定されたアミノ酸を含むポリペプチドまたはフラグメントを生成を導く。例えば、配列特異的ポリペプチド切断法は、1つ以上の既知のアミノ酸残基をフラグメントのカルボキシル末端またはアミノ末端に有するフラグメントを産生する。しかし、特定のアミノ酸残基が、ポリペプチドのカルボキシル末端フラグメントまたはアミノ末端フラグメントに位置するかを知ることは、必要ではない。従って、配列特異的プロテアーゼを用いるポリペプチドの切断は、ポリペプチドまたはそのペプチドフラグメント中に対応するアミノ酸および/または配列の存在を示

10

20

30

40

50

す。同様に、試薬は、ポリペプチドまたはフラグメントの1つ以上の特定のアミノ酸残基を特異的に修飾または標識するために使用され得る。そのような改変または標識を含むポリペプチドまたはフラグメントは、特定のアミノ酸を含むことが知られている。一部のアミノ酸組成は、ポリペプチドに関連する特性を示し、これは、ポリペプチド識別サブインデックスを作成するためのポリペプチド識別インデックスを束縛するために有用であり得る。

【0082】

従って、一連の決定された特性とポリペプチド識別インデックスとの比較は、ポリペプチドの決定された特性により束縛される一連の検索を包含する。例えば、親ポリペプチドの質量またはフラグメントの質量の初期の検索を、実施し、ポリペプチドおよびフラグメントの質量値を含むポリペプチド識別サブインデックスを作成し得、このポリペプチド識別サブインデックスは、機器の誤差範囲内において、同定されるポリペプチドまたはそのフラグメントの値に類似する。作成されるポリペプチド識別サブインデックスを背景として検索される第2の特性（例えば、識別されるポリペプチド中のシステイン残基の存在）は、さらなる束縛を提供し、そしてさらなるポリペプチド識別サブインデックスを作成するために使用され得る。

10

【0083】

ポリペプチドまたはフラグメントの決定された質量は、広大なデータベース検索の効率を増加させる上記のデータベース検索を、束縛するために有利に使用され得る特性である。例えば、タンデム型MSスペクトルは、SEQUESTTMのようなソフトウェアを使用して分析され得、ここで、このソフトウェアは、データベースにおける未知のペプチドの分子量と一致するペプチドのリストを生成し、これを用いてCIDが実施され、次いで、観測される未知のCIDスペクトルを、全ての可能な等質量物(isobar)と比較する(Eng, J.ら、J. Am. Soc. Mass. Spectrom. 5: 976-989, (1994))。従って、この型の検索によって生成される、分析されるポリペプチドフラグメントと類似する分子量を有する一連のペプチドは、ペプチドフラグメントによって示される、可能な親ポリペプチドの部分集合を提供する。次いで、この部分集合は、例えば、親ペプチドを同定するために、部分的アミノ酸組成を使用して、検索され得る。当業者は、SEQUESTTMのようなソフトウェアアプリケーションを使用する特定の検索についての適切な相関スコアパラメータを、知っているかまたは容易に決定し得る。

20

30

【0084】

2つ以上のポリペプチド群を比較する方法は、2つのポリペプチド群を定量的に区別する方法(本明細書中に記載されるような、ICATTM型試薬を使用する方法、これは図3に図示される)を使用する。2つまたはいくつかの化学的に同一だが、差次的に同位体により標識されたICATTM型試薬は、この工程において使用され得る。従って、図3は、2つのサンプルを描くが、複数のサンプルが、本明細書中に記載される方法を使用して比較され得る。図3の中に描かれるサンプルは、互いに増殖条件が異なる同一のサンプル型から収集されたポリペプチド群を含む。例示的な異なる増殖条件として、異なる代謝条件の下での増殖または異なる代謝条件にある細胞、正常サンプルおよび疾患サンプル(例えば、腫瘍サンプル)の比較、薬物学的試薬で処理されていない細胞と処理された細胞の比較などが挙げられる。

40

【0085】

図3に示されるように、サンプルは、ICATTM型試薬を使用して独立に標識され、組み合わせられ、そして本明細書中に記載されるように、ポリペプチドおよび対応するフラグメントの特性が決定される。このプロセスの間に生成されるポリペプチドおよびフラグメントは、所望の場合、サンプルの処理量および感度を増大するために、MS/MSではなく、単一工程の質量分析を使用して分析され得る(Goodlettら、上記、2000)。ポリペプチドおよびフラグメントについて決定された特性は、本明細書中に記載されるように、ポリペプチドの同一性を決定するために使用される。次いで、質量スペクト

50

ルにより、ペプチドイオンのペアについて調べられ得、ここで、このペプチドイオンのペアは、プロセスを通してともに分画され、そして I C A TTM 型試薬においてコードされた質量の相違に正確に対応する質量の相違を有する。2つのピークの相対的なシグナル強度は、フラグメントポリペプチドの相対存在量を示し、そして従って、サンプル中にはじめに存在する対応する親ポリペプチドの相対存在量を示す。従って、2つのポリペプチドサンプル中に含まれるポリペプチドを比較するための方法は、2つの参照ポリペプチド指標の生成に関連し得、ここで、この参照ポリペプチド指標は、ポリペプチド識別コードに加えて、識別された各ポリペプチドに対する、ポリペプチドの量の定量的な決定を含む。

【0086】

2つ以上のポリペプチド群を比較するための代替法は、1つ以上のポリペプチドサンプルと前もって決定されたポリペプチド参照インデックスとの比較である。ポリペプチドサンプルにおける1つ以上のポリペプチドの一連の特性は、識別され得るそして、1つ以上のポリペプチドの同一性および識別されたポリペプチドの相対量を決定するために、参照ポリペプチド識別インデックスと比較され得る。所望の場合、未知のサンプルは、ポリペプチドの相対発現レベルを決定するために上記の定量法を使用する参照サンプルと比較され得る。参照サンプルは、例えば、健康な個体に由来のサンプルであり得るかまたは疾患サンプルのような別のサンプルの生理学的条件と比較するために有用なコントロール条件由来のサンプルであり得る。

【0087】

ポリペプチドの量の定量測定を含む、ポリペプチド識別インデックスは、識別インデックスを作成するために使用される特定のサンプルの「ポリペプチドプロファイル」であると考えられる。本明細書中で使用されるポリペプチドプロファイルは、ポリペプチド量を含む、特定のサンプルを生成するための、一連のポリペプチド識別コードである。

【0088】

ポリペプチドプロファイルは、プロテオミクスの方法において有用である。なぜなら、このようなプロファイルは、細胞、組織、器官および生物の間の異なる条件または状態を識別するために使用され得るからである。細胞または組織によって特定の時期に発現されるポリペプチドは、測定の時点での細胞または組織の条件を規定するために使用され得る。従って、異なる状態における同一の細胞型のポリペプチドプロファイルの間の定量的および定性的な相違は、それぞれの状態を診断するために使用され得る。このような比較についての例として、正常細胞に対する腫瘍細胞、異なる代謝条件にある細胞および特定の薬理学的試薬で処置されていない細胞に対する特定の薬理学的試薬で処理された細胞が挙げられる。2つのポリペプチドプロファイルの間の相違は、「差次的ポリペプチドプロファイル」として記載され得る。差次的ポリペプチドプロファイルは、異なる細胞型（例えば、癌性細胞および正常細胞、刺激された細胞および刺激されていない細胞）、または臨床目的の異なる組織サンプル）に由来するサンプル中に含まれるポリペプチドにおける定量的な変化を分析するために有用である。

【0089】

差次的なポリペプチドプロファイルを生成するための本発明の方法は、体液のようなサンプル中のポリペプチドプロファイルにおける変化を分析するために適用可能である。差次的なポリペプチドプロファイルは、2つの標本のポリペプチドプロファイル（例えば、正常ポリペプチドプロファイルと疾患関連ポリペプチドプロファイル）を比較することによって決定される。例えば、正常な標本状態にある代表的なポリペプチドプロファイルが、生成されそして、異常な状態または疾患状態にあることが疑われる標本と比較され得る。もしくは、疾患状態を代表する参照ポリペプチドプロファイルは、特定の疾患状態を有するかまたは有することが疑われる個体由来の標本と比較され得る。正常状態または疾患状態を代表する参照ポリペプチドプロファイルは、特定の個体または個体群由来の標本を使用して決定され得る。

【0090】

所望であれば、分析は、個体よりむしろ集団、特に参照集団またはコントロール集団に

10

20

30

40

50

対して実施され得る。このような参照集団は、未知のサンプルの比較のために用いられ得る。当業者は、本発明の方法の詳細な適用を基に適切な参照集団を決定し得る。本発明の方法は、2つのサンプル、例えば、正常状態および疾患状態の間のポリペプチド発現における差異を同定する差次的ポリペプチドプロフィールを生成するために用いられ得る。参照集団のサイズは、参照個体を選択するために用いられる基準に依存する。選択基準および本発明の方法の詳細な適用に依存して、参照集団は、比較的少数から、何千もの個体を含む多数の個体までであり得る。

【0091】

詳細に診断された疾患を有するか、疾患の兆候または症状を示す患者からのサンプルの大規模分析は、個々の状態に対する臨床的マーカーまたはマーカーの一団を同定するために有用である。疾患を有する個体からのサンプルを用いて、その疾患に対する定性および/または定量的ポリペプチド識別インデックスを生成し得る。同様に、治療的処置を受けている患者からのサンプルに含まれるポリペプチドの比較分析を用いて、この処置の成功または失敗を示す診断マーカーまたはマーカーの一団を同定し得る。この方法はまた、特定の臨床状態に対する個体の素因を示すために有用なマーカーまたはマーカーの一団の発見またはスクリーニングのために、系統的な、集団 - 広範囲スケールの、このようなサンプルの分析に適用可能である。

【0092】

本発明は、ポリペプチド識別インデックスを生成するための方法をさらに提供する。この方法は、(a) 第1のポリペプチド、またはそのペプチドフラグメントにともなう2つ以上の特徴のセットを決定する工程であって、これら特徴の1つが、このポリペプチドのポリペプチドフラグメントの質量であって、このペプチドフラグメントの質量が質量分析により決定される工程；(b) 第2のポリペプチドについて工程(a)を繰り返す工程；(c) 必要に応じて、上記第1および第2のポリペプチドにともなう1つ以上のさらなる特徴を決定する工程であって、ここで、この決定された特徴が、上記第1および第2のポリペプチドを区別するに十分であり、それによって、第1および第2のポリペプチドに対するポリペプチド識別インデックスを生成する工程を包含する。この方法はさらに、異なるポリペプチドについて、上記工程(a)から(c)を1回以上繰り返す工程を包含し得、ここで、上記決定された特徴は、上記ポリペプチドの各々を区別するに十分であり、それによって上記ポリペプチドの各々に対するポリペプチド識別インデックスを生成する。ポリペプチドの特徴を決定することについて、上記ポリペプチド識別インデックスは、本明細書に開示の任意の方法、またはポリペプチドにともなう特徴を決定するための周知の方法のいずれかをを用いて決定され得る。このポリペプチド識別インデックスは、必要に応じて、イオン選択の非存在下で質量を決定することにより得られ得る。

【0093】

ポリペプチド識別インデックスを生成するための本発明の方法は、フラグメントイオンを生成するためにイオン選択の非存在下のポリペプチドまたはフラグメントの質量決定を含み得、そしてさらに、1 ppmより大きい精度(または、所望であれば、なおより低い精度(より高いppm))でのフラグメント質量の決定を含み得る。ポリペプチド識別インデックスは、ポリペプチドアミノ酸配列を含み得るが、ポリペプチド識別インデックスを生成するため、またはポリペプチド識別インデックスを用いてポリペプチドを同定するための本発明の方法を実施するために、ポリペプチドまたはフラグメントが配列決定される必要はない。

【0094】

本発明はさらに、ポリペプチドを同定する方法を提供する。この方法は、(a) ポリペプチドの集団から親ポリペプチドのサブセットの質量、および親ポリペプチドのサブセットのペプチドフラグメントの質量を同時に決定する工程；(b) 決定された質量をポリペプチド識別インデックスに対して比較する工程；および(c) 測定された質量を有するポリペプチド識別インデックスの1つ以上のポリペプチドを同定する工程を包含する。このポリペプチド識別インデックスは、注釈されたポリペプチドインデックスであり得る。こ

10

20

30

40

50

の方法はさらに、(d) 1つ以上の親ポリペプチドに関連する1つ以上のさらなる特徴を決定する工程；(e) 上記ポリペプチド識別インデックスに対して工程(a)および工程(d)で測定された特徴を比較する工程；および(f) 必要に応じて、工程(d)および(e)を1回以上繰り返す工程であって、ここで、上記ポリペプチド識別インデックス中の単一ポリペプチドとして親ポリペプチドを同定する特徴のセットが決定される工程を包含し得る。この方法はさらに、ポリペプチドを含むサンプル中の同定されたポリペプチドの量を定量する工程を包含し得る。本明細書に開示されるように、本発明の上記の方法およびその他の方法は、特別の質量精度で実施され得る。ポリペプチド識別インデックス中の単一ポリペプチドとしての親ポリペプチドの同定は、十分な数の特徴を決定することをいい、その結果、上記ポリペプチド識別インデックス中の特定のポリペプチドが、上記決定された特徴に一致、すなわち、ポリペプチドが同定される。

10

【0095】

ポリペプチドを同定するための本発明の方法は、ポリペプチドの集団から親ポリペプチドのサブセットの質量、および親ポリペプチドのサブセットのポリペプチドフラグメントの質量を同時に決定する工程を包含する(実施例IIIを参照のこと)。親ポリペプチドのサブセットの質量の同時決定は、ポリペプチド集団を含む単一サンプルからの親ポリペプチド質量値のサブセットの獲得をいう。用語「同時」は、親ポリペプチドおよびポリペプチドフラグメントの質量を、用いたMS法が、親質量およびフラグメント質量が同じサブセットのポリペプチドに関連付けられ得るに十分な時間フレームで親ポリペプチドおよび対応するフラグメントの質量を獲得し得るように同時に決定されることを意図する。例えば、MS法でサンプリングされるポリペプチドは、カラムの流速により指示されるように、異なるサブセットのポリペプチドが、クロマトグラフィーカラムから溶出するので、経時的に変動する。同時決定は、特定のサブセットのポリペプチドが、さらなるポリペプチドの導入またはオンラインサンプリング法の結果として生じるポリペプチドサブセットのポリペプチドの損失に起因して改変される前の時間の間に生じる。

20

【0096】

ポリペプチドのサブセットの質量の同時決定は、例えば、質量決定のための単一イオンの選択の非存在下で実施され得る。例えば、単一イオンよりむしろ、いくつかのポリペプチドが選択され得る(Masselonら、Anal. Chem. 72: 1918-1924 (2000))。好ましくは、本発明の方法では、5より多いイオン、例えば、6イオン、7イオン、8イオン、9イオン、10イオン、またはそれより多い数のイオンが選択される。このような場合、好ましくは、上記ポリペプチド識別インデックスは、注釈されたポリペプチドインデックスである。

30

【0097】

あるいは、ポリペプチドのサブセットの質量の同時決定は、単一イオン選択の非存在下または供給源領域中のイオン選択の非存在下で実施され得る(図2を参照のこと)。このような場合、得られるフラグメントイオンが解析されて、どのイオンが特定の親ポリペプチドにともない、そしてそれ故、親ポリペプチドと関連する特徴として有用であるかを決定する。このような方法は、標準的なMS法におけるフラグメント化について選択されないあまり豊富でないイオンの検出および同定に有用であり得る。

40

【0098】

ポリペプチド質量およびポリペプチドフラグメント質量を同時に得るために有用な質量分析法は、ESI-TOFでのイン-ソースCIDである。この方法は、親イオンとイン-ソースCID走査との間を連続的に交互することを含む。イン-ソースCID走査は、複数にフラグメント化された親イオンの存在下でさえ、所定の親イオンに追跡し得る特定のフラグメントイオンを提供する。親フラグメントイオン系統は、適切なソフトウェアを用いて質量分析法データの解析により決定され得る。より低い精度の測定を提供するMS器具(例えば、ESI-TOF)は、ポリペプチド同定のために特有の制限を提供するために有利に用いられ得る。

【0099】

50

本発明はさらに、ポリペプチドを同定するための方法を提供する。この方法は、(a) このポリペプチド、またはそのフラグメントと関連する2つ以上の特徴を決定する工程であって、これら特徴の1つはこのポリペプチドのフラグメントの質量であり、このフラグメント質量が、フラグメントイオンを生成するためのイオン選択の非存在下で質量分析法により決定される工程；(b) このポリペプチドと関連する特徴を、ポリペプチド識別インデックスに対して比較する工程；および(c) これら特徴を有するポリペプチド識別インデックスにおける1つ以上のポリペプチドを同定する工程を包含する。この方法はさらに、(d) このポリペプチドと関連する1つ以上のさらなる特徴を決定する工程；ならびに(e) 工程(a) および工程(d) で決定された特徴を、上記ポリペプチド識別インデックスに対して比較する工程を包含し得る。本発明のその他の方法のように、この方法は、サンプル中の同定されたポリペプチドの量を定量する工程を包含し得る。これら方法はまた、特定の質量精度で実施され得る。

10

【0100】

従って、本発明はまた、ポリペプチドを同定する方法を提供する。この方法は、(a) このポリペプチド、またはそのフラグメントと関連する2つ以上の特徴を決定する工程であって、これら特徴の1つはこのポリペプチドのフラグメントの質量であり、このフラグメント質量が、2.5 ppmより大きいppmの精度で質量分析法により決定される工程；(b) このポリペプチドと関連する特徴を、ポリペプチド識別インデックスに対して比較する工程；および(c) これらの特徴を有するポリペプチド識別インデックスにおいて1つ以上のポリペプチドを同定する工程を包含する。この方法はさらに、(d) このポリペプチドと関連する1つ以上のさらなる特徴を決定する工程；および(e) 工程(a) および工程(d) で決定された特徴を、上記ポリペプチド識別インデックスに対して比較する工程を包含する。この方法はまた、より低い精度(より高いppm)で実施され得る。

20

【0101】

本明細書中に開示されるように、親タンパク質は、(1) システイン含量、(2) 正確なペプチド質量および(3) 特定イオンの選択なしの正確なペプチドフラグメント質量の組合せから同定された(実施例IVを参照のこと)。この特定の例ではシステイン含量が用いられたが、システイン含量が決定される特徴である必要はない。伝統的なタンデムMSとは対照的に、MSと一緒に入るすべてのペプチドは、個々のイオンの選択なしにフラグメント化された。同時にフラグメント化された2つのペプチドのすべてのフラグメントイオンが、タンデムMS(ここでは、フラグメントイオンが、一度に単一ペプチドについて測定される)とは対照的に、一緒に測定された。

30

【0102】

本発明の方法は、ポリペプチドと関連する特徴を決定する工程を包含する。ポリペプチドを含むサンプルは、ポリペプチドを含む単離されたポリペプチド混合物のように単純、または生物中で発現されたポリペプチドの本質的にすべてを含むサンプルのような複雑であり得る。さらに、サンプルは、所望であれば、本明細書に開示の方法を用いて分画され得る。

【0103】

ポリペプチドは、種々の供給源から単離されたサンプル中に存在し得る。例えば、ポリペプチドサンプルは、任意の生物学的流体、細胞、組織、器官もしくはその部分、または任意の種の生物から調製され得る。サンプルは、個体中に存在し得るか、または個体から得られ得るかそれに由来し得る。例えば、サンプルは、生検により得られた標本の組織学的切片、または組織培養中に配置されるか、もしくはそれに適合されている個体から単離された細胞であり得る。サンプルはさらに、亜細胞画分または抽出物であり得る。サンプルは、本明細書に記載の方法のような、ポリペプチド溶解性を維持するために適切な当該分野で公知の方法により調製され得る。

40

【0104】

標本は、特に、個体から得られるサンプルをいう。標本は、流体または組織標本とし個体から得られ得る。例えば、組織標本は、皮膚生検、組織生検または腫瘍生検のような生

50

検として得られ得る。流体標本は、血液、血清、尿、唾液、脳脊髄液またはその他の体液であり得る。流体標本は、本発明の方法では特に有用である。なぜなら、流体標本は、個体から容易に得られ得るからである。標本の収集の方法は、当業者に周知である（例えば、YoungおよびBermes、Tietz Textbook of Clinical Chemistry、第3版、BurtisおよびAshwood編、W. B. Saunders、Philadelphia、第2章、42 - 72頁（1999）を参照のこと）。

【0105】

本発明の方法において用いられるポリペプチドは、細胞、組織、器官または生物のような供給源から得られ得る。細胞を溶解するために種々の方法が当該分野で公知である。例えば、細胞は、変性剤、凍結および融解の1つ以上のサイクル、音波処理などにより溶解され得る。溶解の後、ポリペプチド混合物を、画分工程に供し得、例えば、核酸または脂質を除去するか、またはインタクトな亜細胞画分またはオルガネラを除去する。細胞を溶解および分画する方法は、当業者に周知である（Scopes、前述、1993を参照のこと）。

10

【0106】

ポリペプチドの同定には、サンプルまたは標本は、例えば、ドデシル硫酸ナトリウム（SDS）のような変性剤を含む界面活性剤を含む緩衝液のような、ポリペプチドの溶解性を維持するために適切な緩衝液中に含まれ得る。ポリペプチド可溶化に有用な変性剤は、例えば、グアニジン - HCl、グアニジン - イソチオシアナート、尿素などを含む。グアニジン - イソチオシアナートの場合、ポリペプチドの共有結合を改変し得る任意の試薬を用いた処理でのように、このような試薬は、サンプルが比較されるポリペプチド同定標識が、同じポリペプチドの比較のために十分なサンプルと実質的に同じ様式で調製されている限り、用いられ得る。当該分野で周知であるその他の変性剤が、ポリペプチドを可溶化するために同様に用いられ得る。さらに、ジチオスレイトール（DTT）、ジチオエリスリトール（DTE）、またはメルカプトエタノールのような還元剤が含まれ得る。

20

【0107】

必要に応じて、本発明の方法は、タンパク質分画工程を包含し得る。タンパク質分画は、ポリペプチド集団から1つ以上のポリペプチドを除去するために有用な任意の方法をいう。分画は、例えば、不溶性成分から可溶性成分を分離する遠心分離工程、電気泳動の方法、クロマトグラフィーの方法、または開示された任意の方法を包含し得る。クロマトグラフィー分離には、当該分野で周知の広範な種類のクロマトグラフィー媒体を用いて、ポリペプチド集団を分離し得る。例えば、ポリペプチドは、クロマトグラフィー媒体に関連する、サイズ、電荷、疎水性、特定の色素および（親和性リガンドを含む）その他の成分への結合に基づいて分離され得る。サイズ排除、ゲル濾過およびゲル透過樹脂は、サイズに基づくポリペプチド分離に有用である。電荷ベースの分離のためのクロマトグラフィー媒体の例は、強および弱アニオン交換、ならびに強および弱カチオン交換樹脂である。疎水性または逆相クロマトグラフィーもまた用いられ得る。

30

【0108】

アフィニティークロマトグラフィーもまた、用いることができ、例えば、Cibacron blueのような色素結合樹脂、ATP、NADなどのようなコファクターのアナログを含む基質アナログ、リガンド、ポリクロールまたはモノクロールいずれかの特異的抗体などを含む。例示のアフィニティークロマトグラフィー樹脂は、本明細書に開示のような、ポリペプチド上のビオチンタグに結合するアビジン樹脂のようなポリペプチド中に取り込まれ得る特異的部分に結合する親和性樹脂を含む。特定のクロマトグラフィー媒体の解像度および能力は、当該分野で公知であり、そして当業者によって決定され得る。同様に、特定の適用のための特定のクロマトグラフィー分離の有用性は、当業者により評価され得る。

40

【0109】

当業者は、特定のサンプルサイズまたは組成物について適切なクロマトグラフィー条件を決定し得、そして規定された緩衝液、カラム寸法、流速、温度、およびその他の条件下

50

で、クロマトグラフィー分離のための再現性ある結果を得る方法を知っている。必要に応じて、タンパク質分画法は、例えば、特定のクロマトグラフィー適用またはその他の分画工程の再現性を評価するために、内部標準の使用を含み得る。適切な内部標準は、クロマトグラフィー媒体または分画工程に依存して変化する。当業者は、クロマトグラフィーまたは分画工程の方法に適用可能な内部標準を決定し得る。

【0110】

ポリペプチドタグ化は、ポリペプチドサンプルの複雑性を減少させ、データベース検索の拘束を提供し、そしてポリペプチドの定量的な比較を可能にするための本発明の方法に有用である。ポリペプチドサンプルの複雑性は、そのタグを含むポリペプチドの部分集団を単離するために使用され得る親和性タグでポリペプチドをタグ化することによって減少され得る。例えば、ポリペプチドおよびフラグメントの集団は、比較的まれなアミノ酸（例えば、システイン）上で標識され得るか、または翻訳後修飾に基づいて標識され得、そしてこのタグを含むポリペプチドおよびフラグメントの部分集団が、単離され得る。従って、特定のアミノ酸を標識することにより単離されたポリペプチドおよびフラグメントの部分集団は、既知のアミノ酸を有する。本明細書中に記載されるように、既知のアミノ酸は、データベース検索の拘束に有用な部分的アミノ酸組成を構成する。定量的なポリペプチドの比較は、2つのポリペプチドまたはポリペプチド集団を差次的に標識することによって、実施され得る。ポリペプチドタグ化の任意の他の方法が、ポリペプチドの比較のために同様に適用され得るが、ICATTM型親和性試薬またはIDENT型試薬（以下により詳細に記載される）が、この目的のために特に有用である。

10

20

【0111】

ポリペプチドタグ化は、当該分野において公知の種々の方法を用いて実施され得る。ポリペプチドタグ化またはポリペプチド修飾のための試薬は、リンカー領域により隔てられた種々の成分を含み得る。ポリペプチドタグ化試薬の成分は、ポリペプチド特定の化学基を改変する反応基、（例えば、質量分析によって）検出され得る部分、およびポリペプチド単離のために使用される親和性タグを含み得る。ポリペプチドタグ化試薬の2つの例（ICATTM型およびIDENT）は、以下に詳細に記載されるが、所望される場合、任意の型のポリペプチドタグが使用され得る。

【0112】

2つのポリペプチド集団を定量的に比較するための本発明の方法は、同位体コード親和性タグ（ICATTM）法（Gygiら、Nature Biotechnol. 17: 994-999（1999）；WO 00/11208（これらの各々は、本明細書中に参考として援用される））の使用を包含し得る。ICATTM型試薬は、さらに、定量的比較を含まないポリペプチドタグ化適用に有用であり得る。ICATTM型試薬法は、質量分析を用いて容易に識別される同位体（例えば、水素および重水素）を用いて差次的に標識され得る親和性タグを使用する。ICATTM型親和性試薬は、3つの要素（親和性タグ、リンカー、および反応基）からなる。

30

【0113】

ICATTM型親和性試薬の1つの要素は、この親和性タグの同種の結合パートナーに結合することによって親和性試薬にカップリングしたペプチドの単離を可能にする親和性タグである。特に有用な親和性タグは、ビオチンであり、ビオチンは、その同種の結合パートナーであるアビジン、または関連する分子（例えば、ストレプトアビジン）に高い親和性で結合し、従って、さらなる生化学的操作に対して安定である。このICATTM型親和性試薬にカップリングしたペプチドの単離を可能にするのに十分な、その同種の結合パートナーへの結合親和性を提供する限り、任意の親和性タグが、使用され得る。親和性タグはまた、磁性ビーズまたは磁性親和性タグの単離に適する他の磁性形式を用いて、タグ化されたペプチドを単離するために使用され得る。ICATTM型親和性試薬法、またはペプチドを親和性タグ化する他の任意の方法において、共有結合トラッピングの使用が、所望される場合、タグ化されたペプチドを固体支持体に結合させるために使用され得る。

40

50

【0114】

ICAT^{T M}型親和性試薬の第2の要素は、リンカーであり、このリンカーは、安定な同位体を組み込み得る。このリンカーは、検体ポリペプチドへの反応基の結合および同種の結合パートナーへの親和性タグの結合を可能とするのに十分な長さを有する。このリンカーはまた、1つ以上の元素での安定な同位体の組み込みを可能とするのに適切な組成を有する。特に有用で安定な同位体の対は、水素および重水素であり、これらは、質量分析を用いて、各々、軽形態(light form)および重形態(heavy form)として容易に識別され得る。この重形態および軽形態が、質量分析を用いて容易に識別され得る限り、多数の同位体元素のいずれもが、このリンカーへと組み込まれ得る。例示的なリンカーとしては、4, 7, 10 - トリオキサ - 1, 13 - トリデカンジアミンベースのリンカーおよびその関連する重水素化形態である2, 2', 3, 3', 11, 11', 12, 12' - オクタジュウテリオ - 4, 7, 10 - トリオキサ - 1, 13 - トリデカンジアミン(Gygiら、(前出、1999)により記載される)が挙げられる。当業者は、上記の基準を満たす、ICAT^{T M}型親和性試薬中で有用な多数の適切なリンカーのいずれかを容易に決定し得る。

10

【0115】

ICAT^{T M}型親和性試薬の第3の要素は、反応基であり、この反応基は、検体中のポリペプチドに共有結合し得る。ポリペプチド中の側鎖アミノ酸を改変する方法は、当業者に周知である(例えば、Glazerら、Laboratory Techniques in Biochemistry and Molecular Biology: Chemical Modification of Proteins, Chapter 3, pp. 68 - 120, Elsevier Biomedical Press, New York (1975); Pierce Catalog (1994), Pierce, Rockford ILを参照のこと)。この反応基が、ポリペプチドに共有結合する限り、種々の反応基のいずれもが、ICAT^{T M}型親和性試薬に組み込まれ得る。例えば、ポリペプチドは、スルフヒドリル反応基を介してICAT^{T M}型親和性試薬に結合し得る(この試薬は、ポリペプチド中のシステインの遊離のスルフヒドリルまたは還元されたシステインと反応し得る)。例示的なスルフヒドリル反応基としては、Gygiら(前出、1999)に記載されるようなヨードアセトアミド基が挙げられる。他の例示的なスルフヒドリル反応基としては、マレイミド、アルキルハライドおよびアリアルハライド、
- ハロアシルならびにピリジルジスルフィドが挙げられる。所望される場合、このポリペプチドは、ICAT^{T M}型親和性試薬と反応する前に、還元され得る。このことは、そのICAT^{T M}型親和性試薬がスルフヒドリル反応基を含む場合に、特に有用である。

20

30

【0116】

反応基はまた、Lysのようなアミン、例えば、イミドエステル、およびN - ヒドロキシスクシンイミジルエステルと反応し得る。反応基はまた、AspもしくはGlu中に見出されるカルボキシル基と反応し得るか、または、この反応基は、他のアミノ酸(例えば、His、Tyr、Arg、およびMet)と反応し得る。反応基はまた、ホスホペプチドの選択的な標識のためのリン酸基、または他の共有結合的改変されたペプチド(糖ペプチド、リポペプチド、または本明細書中に開示される任意の共有結合的ポリペプチド改変体(modification)のいずれかを含む)と反応し得る。当業者は、本発明の方法に用いるための検体分子の修飾に最適な条件を得るために、種々の試薬、インキュベーション条件およびインキュベーション時間を用いることによって、検体分子を修飾するための条件を容易に決定し得る。

40

【0117】

ICAT^{T M}型試薬法は、ICAT^{T M}型親和性試薬を用いた、ポリペプチドのような検体分子の誘導体化に基づく。2つのサンプルの比較および/または定量的のために、コントロール参照検体および試験される個体由来の検体が、ICAT^{T M}型親和性試薬の軽形態および重形態を用いて差次的に標識される。この誘導体化された検体は、合わせられ、そしてこの誘導体化分子は、切断されて、フラグメントを生じる。例えば、ポリペプチド

50

分子は、1つ以上のプロテアーゼを用いて、ペプチドフラグメントへと酵素的に切断され得る。ポリペプチドの切断に有用な例示的プロテアーゼとしては、トリプシン、キモトリプシン、ペプシン、パパイン、Staphylococcus aureus (V8) プロテアーゼなどが挙げられる。ポリペプチドはまた、例えば、CNBr、酸または他の化学試薬を用いて化学的に切断され得る。

【0118】

一旦フラグメントに切断されると、ICATTM型親和性試薬を用いて誘導体化されたタグ化フラグメントは、親和性タグを介して単離される(例えば、ビオチン化フラグメントは、固体相中またはクロマトグラフ形式中のアビジンに結合することによって単離され得る)。所望される場合、この単離されたタグ化フラグメントは、1つ以上の代替的な分離技術(イオン交換クロマトグラフィー、逆相クロマトグラフィー、サイズ排除アフィニティークロマトグラフィーなどを含む)または電気泳動技術(等電集束法を含む)を用いてさらに分画され得る。例えば、この単離されたタグ化フラグメントは、高速液体クロマトグラフィー(HPLC)(マイクロキャピラリーHPLCを含む)によって分画され得る。

10

【0119】

これらのフラグメントは、質量分析(MS)を用いて分析される。これらの検体分子は、軽親和性タグおよび重親和性タグを用いて差次的に標識されているので、これらのペプチドフラグメントは、MSで識別され得、コントロール参照検体および試験検体由来の各々のペプチドフラグメントの相対量の並行(side-by-side)比較が可能である。所望される場合、MSはまた、対応する標識されたペプチドを配列決定するために使用され得、このことは、このタグ化ペプチドフラグメントに対応する分子の同定を可能にする。

20

【0120】

ICATTM型試薬法の利点は、軽ICATTM型試薬および重ICATTM型試薬でタグ化したペプチドの対が、化学的に同一であり、したがって、正確な定量のための相互内部標準としての役割を果たす(Gygiら、前出、1999)ことである。MSを用いて、重タグ化フラグメントおよび軽タグ化フラグメントの対のうちの低い方の質量成分と高い方の質量成分との間の強度の比は、これらのペプチドフラグメントの相対量の正確な尺度を提供する。従って、このICATTM型試薬法は、所望される場合、差次的に発現されるポリペプチドを同定するために簡便に使用され得る。

30

【0121】

IDEN^T試薬は、特定のタンパク質の官能基に同位体タグを導入することによりポリペプチドを修飾するために、使用され得る。例示的なIDEN^T試薬は、Goodlettら、前出、2000に記載される。IDEN^T試薬は、質量分析計において特有の信号を生じる同位体分布を有する少なくとも1つの要素を含む。例えば、IDEN^T試薬は、塩素、重水素、または別の要素(放射性の要素を含む)を含み得る。IDEN^T試薬は、ポリペプチド中の少量のアミノ酸(例えば、システイン)に結合するように設計され得る。IDEN^Tタグを用いたポリペプチドの標識は、ポリペプチドフラグメント質量を用いてポリペプチド識別インデックスを検索するための拘束を提供することによって、本発明の方法に適用され得る。

40

【0122】

タンパク質切断またはタンパク質フラグメント化は、データベース検索のための拘束を提供するための本発明の方法に有用である。ポリペプチドフラグメント化は、配列特異的または配列非特異的であり得る。配列特異的ポリペプチド切断は、既知のアミノ酸を含むポリペプチドフラグメントを得るという利点を提供し、この配列特異的ポリペプチド切断は、データベース検索を拘束するために用いられ得る。非特異的ポリペプチド切断を実施するために有用な試薬の例は、パパイン、ペプシン、およびプロテアーゼSgである。これらのプロテアーゼは、その反応条件を変化させることにより、所望される程度のタンパク質フラグメント化(例えば、1つのポリペプチドからの約2~4つのポリペプチドフラ

50

グメントの生成)を達成するために用いられ得る。これらのプロテアーゼを用いるための条件は、当該分野において周知である。配列特異的なポリペプチド切断を実施するために有用な試薬の例は、トリプシン、V-8プロテアーゼ、o-ヨードソ(iodoso)安息香酸、臭化シアン、および酸である。

【0123】

本発明はまた、ポリペプチドの集団から1つのポリペプチドを同定するためのポリペプチド識別インデックスを提供する。このインデックスは、このインデックス中にポリペプチドに関連する注釈された特徴のセットを含み、これらの特徴のうちの1つは、このポリペプチドのフラグメントの質量であり、このフラグメントの質量は、フラグメントイオンを生成するためのイオンの選択をすることなく、質量分析により決定される。これらの特徴は、このインデックス中の他のポリペプチドからこのポリペプチドの1つを区別するのに十分である。ポリペプチド識別インデックスは、2つ以上のポリペプチド、3つ以上のポリペプチド、5つ以上のポリペプチド、10以上のポリペプチド、20以上のポリペプチド、50以上のポリペプチド、100以上のポリペプチド、150以上のポリペプチド、200以上のポリペプチド、500以上のポリペプチド、1000以上のポリペプチド、2000以上のポリペプチド、5000以上のポリペプチド、または10,000以上のポリペプチドについての特徴さえも含み得る。ポリペプチド識別インデックスはまた、サンプル中のポリペプチドの実質的に全てを含み得る。例えば、ポリペプチド識別インデックスは、ゲノム(例えば、ウイルスゲノム、細菌ゲノム、植物ゲノム、または動物ゲノム(ヒト、非ヒト霊長類、マウス、ラット、ウシ、ヤギ、ウサギ、または他の哺乳動物種のような哺乳動物のゲノムを含む))中で発現されるポリペプチドの実質的に全てを含み得る。ポリペプチド識別インデックス中のポリペプチドの数は、ユーザーの必要性に依存し、そしてポリペプチドを同定するために用いられるサンプル供給源およびサンプル中のポリペプチド発現の複雑性に依存して変化する。*Drosophila melanogaster*についてのポリペプチド識別インデックスの作成の例が、本明細書中に開示される(実施例Vを参照のこと)。

【0124】

このポリペプチド識別インデックスは、生物全体または生物中の特定の組織もしくは細胞に関し得るか、または所望される場合、特定の亜細胞画分(例えば、オルガネラ)に関し得る。従って、試験されるサンプルに適用される複雑性の減少と同様に、特定の標的(例えば、生物、組織、細胞、または亜細胞画分)に関するポリペプチド識別インデックスは、特定の適用における特定のポリペプチドの同定のための検索の単純化に有用であり得る。例えば、特定のポリペプチドもしくはポリペプチドの群の発現、またはこのポリペプチドの発現の量が、特定の状態(例えば、疾患状態)に相関している特定の診断適用において、関連する標的に関するポリペプチド識別インデックスの使用が、用いられ得る。例えば、一群の核タンパク質が、癌細胞中で過剰発現されることが公知である場合、核タンパク質に関連するポリペプチド識別インデックスの使用が、本明細書中に開示される定量方法を用いて、個体由来のサンプル中の核タンパク質の過剰発現について試験するために、用いられ得る。さらに、標的化ポリペプチド識別インデックスの作成および関連する疾患サンプルに対する比較が、異常に発現されるポリペプチドを同定するために用いられ得、これらは、次いで、本明細書中に開示されるような診断適用に用いられ得る。

【0125】

本発明は、そのインデックスのポリペプチド(またはそれらのフラグメント)に関連する2つ以上の特徴を含む、そのインデックスのポリペプチドに関連する注釈された特徴のセットを含むポリペプチド識別インデックスをさらに提供し、これらの特徴のうちの1つは、このポリペプチドのフラグメントの質量であり、このフラグメント質量は、フラグメントイオンを生成するためのイオンの選択をすることなく質量分析によって決定され、ここでこの質量は、1ppmより大きいppmで正確に決定される。

【0126】

所望される場合、ポリペプチド識別インデックスは、コンピューター読み取り可能な媒

10

20

30

40

50

体に簡便に保存され得る。従って、本発明は、本発明のポリペプチド識別インデックス（例えば、注釈されたポリペプチドインデックス）を含む、コンピューター読み取り可能な媒体を提供する。ポリペプチド識別インデックスを含むこのようなコンピューター読み取り可能な媒体は、ポリペプチドの特徴と、このポリペプチド識別インデックスとの比較に有用であり、この比較は、コンピューター装置上で簡便に実施され得る。ポリペプチド識別インデックスが簡便に保存され得、かつサンプル中のポリペプチドの特徴および/またはサンプル中のポリペプチドの定量的量に対する比較のためにアクセスされ得るので、コンピューター装置の使用は、簡便である。ポリペプチド識別インデックスは、適切なハードウェア、ソフトウェア、および/またはネットワークを用いて（例えば、ネットワーク（インターネットを含む）に接続されたハードウェアを用いて）簡便にアクセスされ得る。

10

【0127】

種々のハードウェア、ソフトウェア、およびネットワークの組み合わせを使用することによって、ポリペプチドについて決定された特徴をポリペプチド識別インデックスに対して比較する工程を含む本発明の方法は、種々のコンフィギュレーションで簡便に実施され得る。従って、本発明は、本発明の工程に対応するコンピューター実行可能な工程を実施するためのコンピューター装置をさらに提供する。例えば、単一のコンピューター装置は、ポリペプチドについて決定された特徴をポリペプチド識別インデックスに対して比較するコンピューター実行可能な工程を実施するための指示、ポリペプチド識別インデックス、およびこのポリペプチドについて決定された特徴が、このポリペプチド識別インデックスにおける1つ以上のポリペプチドに対応するか否かを決定するための指示を含み得る。

20

【0128】

あるいは、このコンピューター装置は、このポリペプチド識別インデックスが、別個の媒体に保存されている一方で、本発明の方法の工程を実行するための指示を含み得る。さらに、ポリペプチドが、このポリペプチド識別インデックスにおける1つ以上のポリペプチドに対応するか否かを決定するための指示は、別個のコンピューター装置もしくは別個の媒体に含まれ得るか、または本方法のコンピューター実行可能な工程を含むコンピューター装置および/もしくは別個の媒体のデータベースと組み合わせられ得る。このような別個のコンピューター読み取り可能な媒体は、別のコンピューター装置、保存媒体（例えば、フレキシブルディスク、Zipディスク）またはサーバー（例えば、ファイルサーバー）であり得、これらは、搬送波（例えば、電磁搬送波）によってアクセス可能であり得る。従って、ポリペプチド識別インデックスを含むコンピューター装置またはこのポリペプチド識別インデックスが保存されるファイルサーバーは、インターネットのようなネットワークを介して遠隔でアクセス可能である。当業者は、本発明のコンピューター装置の相互接続を可能にする適切なハードウェア、ソフトウェア、またはネットワークインターフェースを知っているか、またはそれらを容易に決定し得る。

30

【0129】

所望される場合、ポリペプチド識別インデックスを用いてポリペプチドを同定する効率を最適化および増加させるために、アルゴリズムが用いられ得る。例えば、上記のように、ポリペプチド識別インデックスのサブインデックスが決定され得、そしてこのサブインデックスは、比較されるデータベースのサイズを減少させることによって、ポリペプチドをさらに同定するために使用され得る。サブインデックスを用いることによるこのような減少は、データベース検索のためのコンピューターの時間を減少させ得る拘束を提供する。当業者は、データベース検索の効率を増加させるための種々の方法を認識する。

40

【0130】

アルゴリズムを用いて、質量スペクトルデータの分析効率を改善し得る。例えば、タンデム質量スペクトルにおける相補的イオンの存在を用いて、親イオンの質量をより正確に計算するアルゴリズムを開発し得る。これは、スコア付けを改善し、そしてペプチド質量許容範囲を、アイソバリックデータベースペプチドのより小さなサブセットへと狭めることによって検索時間を低減する。プロテオーム適用範囲を増大させるために、イオントラ

50

ップにおけるESI-MS/MSは、明白なペプチド電荷状態の決定が除外されて、全てのペプチドのタンデム質量スペクトルを $[M + 2H]^2+$ および $[M + 3H]^3+$ として2回検索させるように行われ得る。最良のスコアを生じる電荷状態が保持され、他のものは捨てられる。データベース検索の前に電荷状態を決定するアルゴリズムを用いて、処理量を増大させ得る。スペクトルの品質を測定するルーチンを用いて、データベース検索の前および後でのスペクトルのカタログ化を可能にし得る。これは、検索前に乏しいスペクトルを捨てることによって処理量を増大させ、そしておそらく予想外の翻訳後修飾に起因して、本明細書中に開示されるように、良好なデータベース一致を生じることができない「良好な品質」のスペクトルについての分析の追跡を可能にする。

【0131】

本発明の種々の実施形態の活性に實質的に影響を与えない改変もまた、本明細書中に提供される本発明の定義内に含まれることが理解される。従って、以下の実施例は、本発明を例示することを意図するが、本発明を限定することは意図しない。

【実施例】

【0132】

(実施例I：注釈されたポリペプチドインデックスの作成)

本実施例は、注釈されたポリペプチドインデックスの作成およびサンプル中のポリペプチドを同定するための注釈されたポリペプチドインデックスの使用を記載する。

【0133】

注釈されたポリペプチド(AP)インデックス(注釈されたペプチドタグ(APT)インデックスまたはデータベースとも呼ばれる)の構成要素は、調査している種、細胞または組織によって産生されるタンパク質の配列特異的な化学的または酵素的断片化によって作成される特異的な構造の特徴を有する、本質的に全てのペプチドまたは選択されたペプチドの配列である。各ペプチドは、実験によって容易に決定され、そして注釈されたペプチドとこのペプチドを誘導したタンパク質との間の明白な相関を可能にする、属性または特徴について注釈が付けられる。

【0134】

例示的なAPインデックスの作成は、以下の特定の工程を含み得る：タンパク質を回収する工程；タンパク質を、同位体コード化アフィニティータグ(ICATTM)型試薬で標識する工程；タンパク質を分子量によって分画する工程；タンパク質を、プロテアーゼ(例えば、トリプシン)で消化してペプチドを作製する工程；ペプチドを、クロマトグラフィー(例えば、イオン交換クロマトグラフィー)によって分離する工程；各イオン交換画分を、(例えば、ICATTM型アフィニティータグに基づく)アフィニティークロマトグラフィーによって精製する工程；各アフィニティークロマトグラフィー画分を、LC/MS/MSまたはCE/MS/MSによって分析する工程；個々のMS/MSペプチドスペクトルのデータベース検索を介する、発現された本質的に全てのタンパク質を同定する工程；および測定した物理化学的特性に基づいて、個々のペプチドについての、従って、そのペプチドの親タンパク質についての独特のバーコードを構成する注釈されたペプチドタグのデータベースを作製する工程。

【0135】

APインデックスを、以下の通りに作成し得る：(i)タンパク質サンプルの調製；(ii)タンパク質タグ化；(iii)任意のタンパク質分画；(iv)タンパク質断片化；(v)ペプチド分離；(vi)タグ化ペプチドのアフィニティー単離；(vii)高分解能ペプチド分離；(viii)データベース検索；(ix)APインデックス(APTデータベース)構築。

【0136】

(i)タンパク質サンプル調製。定量的プロテオーム分析が実施されるべきであるタンパク質サンプル(例えば、細胞、組織、細胞画分、体液、細胞分泌物など)を、タンパク質の可溶性を維持するために、標準的プロトコルを用いてそれぞれの供給源から単離する。

10

20

30

40

50

【0137】

(ii) タンパク質タグ化。サンプル中のタンパク質を、完全に変性させ、還元し、そして還元されたシステイン残基の側鎖を示す全てのスルフヒドリル基を、それぞれ、以前 (Gygiら, Nature Biotechnol. 17: 994 - 999 (1999)) に記載された条件を用いて、軽い形態または重い形態のスルフヒドリル特異的 ICATTM 型試薬で共有結合的に誘導体化する (図3を参照のこと)。システインタグ化は、この方法の特に有用な実施であるが、タンパク質中の化学基について特異性を有する任意の他の化学的反応もまた適用され得る。

【0138】

(iii) 任意のタンパク質分画。タグ化タンパク質の混合物を、既知の標準的なタンパク質分離手順のうちのいずれか1つを用いて分画する。適用される手順は、再現性があり、タンパク質を溶液中に維持し、そして高いサンプル容量を有する。特に有用な方法は、調製的ドデシル硫酸ナトリウム - ポリアクリルアミドゲル電気泳動 (SDS - PAGE) である。

10

【0139】

(iv) タンパク質断片化。サンプル混合物中のタンパク質、または任意のサンプル画分が用いられる場合、各画分に含まれるタンパク質を、配列特異的切断に供する。特に有用な方法は、(例えば、トリプシンを用いた) プロテアーゼ切断である。

【0140】

(v) ペプチド分離。得られるペプチド混合物を、一次元ペプチド分離に供する。このペプチド分離方法は、適用したサンプルの複雑さにかかわらず、少なくとも中程度の分解能で、高いサンプル容量を有し、そして非常に再現性の高い分離パターンを生じる。特に有用な一次元分離方法は、イオン交換クロマトグラフィー (例えば、陽イオン交換または陰イオン交換) である。

20

【0141】

(vi) タグ化ペプチドのアフィニティー単離。ICATTM 型試薬でタグ化したペプチド (例えば、システイン含有ペプチドまたはアフィニティー標識され得る他のアミノ酸もしくは翻訳後修飾を含有するペプチド) を、アビジンまたはストレプトアビジンアフィニティークロマトグラフィーを用いて各クロマトグラフィー画分から単離する。特に有用なアフィニティー媒体は、ポリマービーズに固定されたモノマーアビジンである。ピオチンとは異なるアフィニティータグを有する ICATTM 型試薬を用いる場合、このタグに相補的なアフィニティー媒体を用いる。

30

【0142】

(vii) 高分解能ペプチド分離。高分解能ペプチド分離について特に有用な方法は、液体クロマトグラフィーESI - MS / MSである。アフィニティークロマトグラフィーカラムから溶出したペプチド混合物を、自動化LC - MS / MSによって、分離方法としてのキャピラリー逆相クロマトグラフィー (Yatesら, Methods Mol. Biol. 112: 553 - 569 (1999)) および質量分析法としての動的排除を伴うデータ依存性CID (Goodlettら, Anal. Chem. 15: 1112 - 1118 (2000)) を用いて個々に分析する。

40

【0143】

(viii) データベース検索。適切なCIDスペクトルが得られる全てのペプチドの配列を、調査している種からの配列データベースを検索することによって決定する。特に有用な配列データベースは、試験している種によって潜在的に発現され得る本質的に全ての完全タンパク質配列を含むデータベースである。用いられ得る配列データベース検索プログラムは、SEQUESTTM プログラム (Eng, J.ら, J. Am. Soc. Mass Spectrom. 5: 976 - 989, (1994)) または類似の能力を有するプログラムである。

【0144】

(ix) APインデックス (APTデータベース) 構築。上記の手順によって同定され

50

た全てのペプチドの配列を、データベースに入力し、そして上記の工程 (i) ~ (v i i) の間に作製された特徴 (すなわち、属性) を用いて注釈を付ける。これらの特徴 (すなわち、属性) としては、以下が挙げられるがこれらに限定されない：部分的アミノ酸組成 (例えば、各選択されたペプチド中のシステイン残基の存在；G o o d l e t t ら，前出，2 0 0 0 を参照のこと)；親タンパク質のおよその分子量 (任意の S D S - P A G E 分画によって決定した場合)；イオン交換カラムからの溶出順序または溶出時間；逆相カラムからの溶出時間、および決定された任意の他の特徴。集合的に、これらの属性は、各ペプチドについてのバーコードと同様に、データベース中のあらゆるペプチドに関して独特である。それゆえ、同じバーコードが、その後の実験において生成されたペプチドから決定される場合、これらは、この実験によって生成されたバーコードを、A P 検索 (A P T データベース) 中に存在するバーコードと単に相関させることによって、この実験によって生成されたペプチドを独特に同定する。

10

【 0 1 4 5 】

ポリペプチドの、A P インデックス (A P T データベース) との相関に関して、上記の方法によって定量的プロテオーム分析のために生成されたペプチドサンプルを、以下の例外以外は、A P インデックス (A P T データベース) について生成されたペプチドと同様に、作製し、処理し、そして正確に加工する。(i) 比較されるべき 2 (以上の) サンプル中のタンパク質を、示差的に同位体標識された I C A T ^{T M} 型試薬で標識する。化学的に同一であるが、示差的に同位体標識された、2 またはいくつかの I C A T ^{T M} 型試薬が、この工程で使用され得る。(i i) 生成されたペプチドを、M S / M S によってではなく、1 段階質量分析のみによって分析する。質量分析器は一般に、高質量精度、高感度および高質量分解を有する。これらの特徴を有する機器としては、以下が挙げられるがこれらに限定されない：M A L D I - T O F 質量分析器、E S I - T O F 質量分析器およびフーリエ変換イオンサイクロトロン質量分析器 (F T - I C R - M S)。このプロセスによって各ペプチドから決定される属性 (上記の通りに決定される全ての属性またはそれらの選択物) を、各ペプチドについてのバーコードに変換し、そして実験的に決定されたバーコードを、A P インデックス (A P T データベース) からのバーコードと相関させて、ペプチドを、それゆえ、そのペプチドの起源であるタンパク質の明白な同定を得る。続いて、質量スペクトルを、このプロセスを通じて共に分画され、かつ用いられる I C A T ^{T M} 型試薬にコードされる質量差に正確に対応する質量差を有する、ペプチドイオン対について調べる。2 つのピークの相対シグナル強度は、これらのペプチドの相対量を、それゆえ、サンプル中に最初に存在した対応するタンパク質の相対量を示す。従って、実験的に決定されたデータの、A P インデックス (A P T データベース) との相関は、分析したサンプル中のタンパク質の定量および同定を可能にする。

20

30

【 0 1 4 6 】

(実施例 I I：酵母の注釈されたポリペプチドインデックスの作成)

本実施例は、酵母についての注釈されたポリペプチドインデックスの作成を記載する。

【 0 1 4 7 】

現在の質量分析器の感度では、L C / L C / M S / M S 法 (G y g i ら，P r o c . N a t l . A c a d . S c i . U S A 9 7 : 9 3 9 0 - 9 3 9 5 (2 0 0 0)) を用いて、量の少ないタンパク質を検出するために少なくとも 5 m g の総タンパク質が必要とされることが見積もられ、そしてこの量を、本質的に実験的に確認した。G y g i ら，前出，2 0 0 0 はまた、量の少ないタンパク質の検出に「開始」プロセスが適切であり、そして比較的多量の総サンプルに順応させるに十分なサンプル容量を有することを実証した。

40

【 0 1 4 8 】

このデータベースの構築に関して、以下の手順を用いる。タンパク質標識に関して、タンパク質サンプルを、0 . 5 % S D S、5 0 m M T r i s (p H 8 . 3)、5 m M エチレンジアミン四酢酸 (E D T A) 中に 5 m g / m l のタンパク質濃度で作製する。計 2 5 m g の総酵母タンパク質を用いる。一旦タンパク質が溶解したら、サンプルを水で 1 : 1 0 希釈し、そして E D T A を添加して 5 m M E D T A 濃度を維持することによって S

50

D S濃度を低下させる。最終濃度は、0.05% SDS、5 mM Tris、5 mM EDTAである。次いで、このサンプルを100 で3~5分間煮沸し、次いで冷凍する。ジスルフィド結合の還元を、十分なトリブチルホスフィン(TBP)を添加してサンプル溶液中で5 mMを達成することによって達成する。これに続いて、サンプルを37 で30分間インキュベートする。還元したサンプルに、アルキル化試薬(例えば、ICATTM型試薬)を、サンプル中に存在するSH基の見積もり5×モル濃度過剰で添加する。アルキル化反応を、暗所で90分間進行させる。

【0149】

タンパク質分離に関して、還元してアルキル化したサンプルに、0.2容量の5×SDSゲルサンプル緩衝液を添加し、そして5分間煮沸する。次いで、冷却したサンプルを、20 cm×20 cm×1.5 mmの寸法の調製用SDSゲルに適用する。電気泳動後、ゲルを、電気泳動の次元に対して垂直方向に、等しいサイズの10枚のストリップにスライスする。これらのストリップは、インタクトなタンパク質についての10個のサイズピンを表す。次いで、ゲルストリップ中のタンパク質を、標準的なプロトコルを用いたゲル内消化に供する。

10

【0150】

ペプチド分離に関して、ゲルスライスから抽出されたペプチドを、3連続クロマトグラフィー分離に供する。最初に、これらを、陽イオン交換クロマトグラフィーによって分離する。第二に、ビオチン化ペプチドを、アビジンクロマトグラフィーによって単離する。第三に、ペプチドを、キャピラリー逆相クロマトグラフィーによってさらに分画する。

20

【0151】

陽イオン交換クロマトグラフィーに関して、陽イオン交換HPLCカラムを用いる(PolyLC Inc., Columbia MD; 2.1 mm×20 cm、5 μm粒子、300 孔径、ポリスルホエチルA強陽イオン交換物質)。以下の緩衝液を用いる: 緩衝液A(10 mM KH₂PO₄、25% CH₃CN(pH 3.0)); 緩衝液B(10 mM KH₂PO₄、25% CH₃CN、350 mM KCl(pH 3.0))。以下の勾配を行う:

| 時間(分) | % B |
|-------|------|
| 0 | 0 |
| 30 | 25 |
| 50 | 100。 |

30

【0152】

流速は、1分間当たり200 μLである。画分を、1分間~2分間の間隔で収集する。約200 μg~約5 mgのどこかでの、消化したICATTM標識総タンパク質を、通常、2 mMのサンプルループを用いてこのカラムにロードする。陽イオン交換カラムにロードする前にサンプルをpH 3.0以下まで酸性にすることが重要である。なぜなら、ペプチドは、より高いpH値では完全には荷電せず、そしておそらくカラムに粘着しないかもしれないからである。示した勾配を、できるだけ多くの二重荷電ペプチドの溶出を広げるように設計し、これらのペプチドは通常、約8~9分に始まって約15~16分まで泳動中に溶出し、その後、三重荷電ペプチドが溶出し始める。30個の画分を、勾配の持続時間の間収集する。

40

【0153】

アビジンアフィニティークロマトグラフィーに関して、UltraLink Monomeric Avidin(Pierceカタログ番号53146)を用いる。ガラスウールの小片を、ガラスピペットチューブの頸部に充填する。400 μlのアビジンクロマトグラフィー材料を、このチューブ中に充填する(スラリーは、50%希釈なので、800 μlの50%スラリーを、400 μlのアビジンクロマトグラフィー材料を得るために添加する)。ビーズを沈澱させ、そしてカラムを2×PBSで洗浄し、ビーズをチューブの側面に沈降させる。充填したカラムを、フロースルーのpHが約1まで変化するまで、0.4%トリフルオロ酢酸(TFA)を含む30%アセトニトリル(ACN)で洗い流し

50

、次いでさらに1カラム容量のACN/TFAで洗い流す。この酸性工程は、ビーズに結合したポリマーを除去するためである。このカラムを、pHが約7.2になるまで2×PBS (pH 7.2)で洗浄する。その後、カラムを、さらに3カラム容積(1200 μl)の同じ緩衝液で洗い流す。

【0154】

カラムを、3~4カラム容量のビオチンブロッキング緩衝液(PBS中2 mM d-ビオチン)で洗浄する。このビオチンは、カラム中のより多くの残存アビジン部位をブロックして、後の残りの結合部位からのサンプルの回収を確実にする。

【0155】

ゆるく結合したビオチンを、フロースルーのpHが約2.8に変化するまで、約6カラム容量(2,400 μl)の再生緩衝液(100 mM グリシン、pH 2.8)で洗い流す。このグリシン溶液を、使用前にオートクレーブすることによって滅菌し、そして4℃で保存し、4週間では、これは、1週間持続する。

【0156】

カラムを、6カラム容量の2×PBSで洗浄して、カラムを適切なpH(約7.2)に戻す。フロースルーのpHをモニタリングする。次いで、個々のまたはプールしたイオン交換カラム画分からなるペプチドサンプルをカラムに適用し、そして約20分間にわたってカラム中でインキュベートする。次いで、結合していない物質を、5カラム容量の2×PBS (pH 7.2)を適用することによってカラムから洗浄し、そして画分を収集する。カラムを、画分の収集を続けながら、5×カラム容量の1×PBS(この工程は、塩濃度を低下させるためである)および20%メタノール(MeOH)を含む6カラム容量の50 mM AMBIC(重炭酸アンモニウム)(pH 8.3)でさらに洗い流す(50 mM AMBICは、塩濃度を低下させるためであり; MeOHは、疎水性ペプチドを除去するためである)。ビオチン化ペプチドを、溶出緩衝液(30% ACN/0.4% TFA)で溶出させ、そして手動でガラス管中に収集する。これらのサンプルを、キャピラリー逆相クロマトグラフィーによってさらに分離する。

【0157】

逆相クロマトグラフィーに関して、精製したビオチン化ペプチドを、標準的なプロトコルを用いた逆相キャピラリークロマトグラフィーによって分離する。ペプチドが60分かかって溶出するように溶媒勾配を選択する。1分間の画分を収集して質量分析器で分析する場合、60個のピンがこの手順によって作製される。

【0158】

質量分析および塩基配列決定のために、RP-カラムから溶出されたビオチン化ペプチドを、データベースの作成に対してはESI-MS/MSにより、そしてデータベース検索についてはESI-MSにより分析する。データベース構築に関しては、イオントラップ質量分析計を、約1質量単位の質量精度で使用する。データベース検索のための質量測定に関しては、ESI-TOF質量分析計を、50 ppmを越える質量精度および10,000を越える解像度で使用する。これは、1000質量単位のペプチドを、1000.1質量単位のペプチドと区別し得ることを意味する。代表的なペプチドの質量範囲が、500質量単位と3000質量単位との間であると仮定する場合、その性能での質量分析計は、25,000ピンを生じる。

【0159】

酵母*S. cerevisiae*のプロテオーム分析の場合、この生物のゲノムは、約6200のオープンリーディングフレーム(ORFs)を含む。従って、この酵母のプロテオームは、差次的に改変された同タンパク質の形態を無視して、約6200個の異なるタンパク質であると予想される。トリプシンに対する、経験に由来する特異性の規則が適用される場合、酵母プロテオームの代表的な消化は、約350,000ペプチドを生じる。システイン含有ペプチドのみが溶出される場合、このサンプルの複雑性は、ICAT^TM型試薬を用いた化学誘導体化に基づき、約35,000ペプチドまで減少する。上記の手順から利用可能なピンの総数は、 $10 \times 30 \times 60 \times 25,000 = 4.5 \times 10^8$ で

10

20

30

40

50

あり、従って、全ての酵母プロテオーム分析から予想されるペプチドの数を遙かに上回る。従って、複雑な酵母抽出物のサンプルに関するデータベース検索データの作成についての手順が、単純化され得ることが期待される。例として、タンパク質についてのゲル電気泳動サイズ分類 (s i z i n g) 工程は、必要に応じて排除される。

【 0 1 6 0 】

データベースに入力されるデータの作成手順も、データベースを検索するためのデータの作成手順も、固定されない。従って、最適化のために、サンプルの複雑性の程度に依存して、利用可能なピンの数は、容易に調整され得る。概して、データベースの作成に対して選択されるピンの数は、高く、一方、データベース検索データを作成するためのピンの数は、サンプルの処理能力を最大化するために、できるだけ低く選択される。利用可能なピンの数は、種々の方法において、容易に適応され得る。第 1 に、さらなる直交分離次元の包含は、タンパク質またはペプチドについて、考慮され得る。タンパク質分離については、等電点電気泳動、イオン交換クロマトグラフィー、ハイドロキシアパタイトクロマトグラフィー、または類似の電気泳動技術もしくはクロマトグラフ技術が含まれ得る。ペプチド分離については、ペプチドサイズに基づく分離またはキャピラリー電気泳動法が含まれ得る。

10

【 0 1 6 1 】

第 2 に、上記の分離方法の分離範囲は、拡大され得る。タンパク質サイズ分類は、拡大された分離範囲を有する勾配ゲルまたはより長いゲルを使用することにより拡大され得る。ペプチドのクロマトグラフ分離法について、ピンの数は、拡大された、より浅い勾配を作り出すことによって、および/またはより頻回なサンプリングにより、容易に拡大され得る。最後に、ピンの数は、使用される質量分析計の解像度および質量精度に依存する。より高度な性能を有する質量分析計が加わることは、その手順において使用される分離方法によって提供されるピンの数を減少させる。

20

【 0 1 6 2 】

(実施例 I I I)

(複合体サンプルの MS 分析のための E S I - T O F の使用)

本実施例は、E S I - T O F により決定されるポリペプチドフラグメントの質量を用いて、ポリペプチド同定のサブインデックスを決定するための方法を記載する。

【 0 1 6 3 】

1 組のポリペプチド、またはそのフラグメントの質量を、複合体ゲノムとの使用のためのポリペプチド同定サブインデックスを生じるための質量値によってポリペプチド識別インデックスを制限するために、E S I - T O F を用いて決定した。2 つのペプチド (A S H L A G A R (P 1) および R P P G F S P R (P 2)) を、E S I - T O F 中に一緒に注入した。この混合物を、H P L C 分離中に、ペプチドの共溶出をシミュレートすることを意図した。スペクトルを、低い $V_{N o z z l e - S k i m m e r}$ (図 4 A) および高い $V_{N o z z l e - S k i m m e r}$ (図 4 B) で得た。

30

【 0 1 6 4 】

図 4 B において見られ得るように、ペプチド両方について、多数のフラグメントイオンが存在する。フラグメントイオンが $P 1 - 1 +$ イオン上に現れず、それゆえこの m / z 範囲を、図 4 B から除外した。親イオンおよびフラグメントイオンの混合物を単一の質量スペクトル中に逆重畳するために、アルゴリズムを書いた (図 4 B)。60, 884 個のヒトポリペプチドのデータベースからの可能性のあるトリプシンペプチド全ての一覧表に P 1 配列および P 2 配列を配置した。60, 884 個のポリペプチドからの可能性のあるトリプシンペプチド全てとの、P 1 の観察された質量と P 2 の観察された質量との比較 (図 4 A) は、それぞれ、57 個の同重体および 124 個の同重体を生じた。次いで、10 ppm まで算出された全同重体由来の b - イオンフラグメントおよび y - イオンフラグメントの一覧を、図 4 B において 500 m / z と 825 m / z との間で観察されたフラグメントイオンと比較した。このプロセスは、図 4 C において示されるように、P 1 および P 2 について、それぞれ 12 個および 13 個の可能性のあるポリペプチド同定の一覧を生じた

40

50

。本方法は、例えば、複合体ゲノムとの使用のためのポリペプチド同定サブインデックスを生ずるために、ポリペプチド識別インデックスに対する制限を適用するのに有用である。ペプチド上のインソース (in-source) CIDは、ESI-TOFによる連続的交代様式において、容易かつ迅速に、そしてMALDI-TOFでのインソース分解により潜在的に入手され得る。さらに、本方法は、タンデムなMSを用いて行われる場合、結果としてデータの喪失を生じず、ここで、選択されたイオンのCID中に共溶出するイオンは、分析から必然的に除外される。ペプチドの存在量が少ない場合(すなわち、S/N (signal-to-noise) が低い場合)、それらのペプチドは、データ依存性 (DD) タンデムMSプロセスによって、CIDについて選択されない。これは、スタンダードDDプロセスが、第1に基準ピーク(すなわち、1m/zウィンドー当たりの最も強いイオン)を調査し、そのピーク上でCIDを実行し、さらなる考察から動的にそのピークを除外し、そして次の最大量のイオンに進むからである。

【0165】

図4Cに示されるペプチドは、60,000個を越える、可能性のあるポリペプチドから約12個または13個のポリペプチドまでの複雑性の減少を示す。親ポリペプチドまたはその1つ以上のさらなるペプチドフラグメントについて、さらなる特性、例えば、原子質量、アミノ酸組成、部分アミノ酸配列、見かけの分子量、pI、および規定された条件下での特定のクロマトグラフ媒体による溶出の順序などが決定される。ポリペプチドを分画するための方法は、以前に述べられている(例えば、Scopes, Protein Purification: Principles and Practice, 第3版、Springer Verlag, New York (1993))。ポリペプチド識別インデックス中の単一のポリペプチドが同定されるように、十分量の特性が決定される。

【0166】

(実施例IV)

(ポリペプチド識別インデックスを作成するためのショットガン(Shotgun)ペプチド塩基配列決定)

本実施例は、高い質量精度を用いるペプチドの塩基配列決定を記載する。

【0167】

ESI-TOF LC/MSシステムにおける高い質量精度(50ppmまたはより良く)の親イオンおよびフラグメントイオンを用いて、ペプチドを塩基配列決定するための方法を開発した。本方法は、交互の高いイオン供給源勾配電位または低いイオン供給源勾配電位での精製されたタンパク質消化サンプルの導入に関し、ここで、ペプチド由来の断片化データは、記録され、そしてデータベース検索を介したペプチド配列および親タンパク質の確実な同定についての非断片化データと組み合わせられ得る。システイン含有量は、この特定の実施例において使用されたが、システイン含有量が、決定された特性である必要はない。従来のタンデムなMSとは対照的に、MSと一緒に入るペプチド全てが、別個のイオンで選択されることなく断片化される。タンデムなMSとは対照的に、同時に断片化された2つのペプチドについてのフラグメントイオンの全てを、一緒に測定し、ここで、フラグメントイオンは、単一のペプチドについて同時に測定される。

【0168】

簡単に、*Saccharomyces cerevisiae* (BWG1-7A)をYPD培地中で3のOD₆₀₀(対数期の直後)まで増殖させた。スフェロプラストを、細胞溶解性(lyticase)消化により生成し、そして続いて機械的に溶解した。遠心分離(13,000rpmで30分間)によって核を単離し、そして硫酸アンモニウムの添加により溶解した。不溶性物質をスピアウト(28,000rpmで90分間)し、そして上清を抽出した。タンパク質を硫酸アンモニウムで沈殿させた。その沈殿物を再懸濁して、20mMのHepes、20%のグリセロール、10mMのMgSO₄、1mMのEGTA、および75mMの硫酸アンモニウムに対して透析した。

【0169】

タンパク質抽出物(5.6mg)を、Speedvac(Savant Instruments; Holbrook NY)中で乾燥し、そして0.5%のSDS、50mMのTris、1mMのEDTA中に再懸濁し、そして3分間煮沸することにより変性させた。ジチオスレイトール(DTT; 5mM)を加え、そしてそのサンプルを50で30分間インキュベートした。タンパク質を、冷やしたアセトン/エタノール(EtOH)(50:50 v:v)中で沈殿させた。そのペレットを洗浄し、そして0.05%のSDS、6Mの尿素、50mMのTris、pH8.3、および1mMのEDTA中に再懸濁し、そして3mgのEZ-LinkTMPEO-Iodoacetyl Biotin(Pierce カタログ番号21334; Pierce Chemical Co.; Rockford IL)を加えた。このサンプルを暗所で攪拌しながら90分間インキュベートした。等容量の20mM DTTを加えてビオチン化反応をクエンチした。次いで、このサンプルを6倍容量で50mMのTris、1mMのEDTAで希釈した。200μgのトリプシンを加え、そしてそのサンプルを37で一晩インキュベートした。

【0170】

約2mgのトリプシンで消化された溶液をPolysulfoethyl Aの強いカチオン交換カラム(Poly LC番号202SE0503、2.1mm×200mm、300A、5μm)(10mMのK₂PO₄、25%のアセトニトリル(ACN)中300mMのKCl(pH3.0)を200μl/分で、30分間にわたり勾配=0%~25%、20分間にわたり25%~100%)、分画を5分ごとに回収した。各分画を、ABIアフニティーカートリッジ、カートリッジホルダー、および付随する緩衝系(ABI番号4326740および番号4326688-90; Applied Biosystems; Foster City CA)を用いて、ビオチン化システイニルペプチドについて精製した。精製した分画容量(0.1カチオン交換吸光度単位を214nmのUVトレースから示す)をSpeedvac中で完全に乾燥させ、そして5μl容量までの0.2%の酢酸(HOAc)中に再懸濁した。

【0171】

乾燥サンプルを、オートサンプラー/μHPLCシステムを100μm×10cmの溶融シリカを手で充填したC-18(Michrom Magic C18)カラムと連結して統合したLC Packing(San Francisco CA)によって、1分につき約700nlで60分にわたり、5%~50%のACN(溶媒B)の直線勾配を用いて逆相で分離した。溶媒Aは、0.2%のHOAc/0.001%のヘプタフルオロ酪酸である。この溶融シリカのキャピラリーカラムを、4000VDC(直流電圧)を適用するステンレス鋼の0.15mmのボア微小容量のValco union(MU1XCS6)を介して、75~30μmのテーパチップ溶融シリカ針(New Objective、FS360-75-30-N)と連結させた。ABI Mariner Biospectrometry Workstation ESI-TOF質量分析計を使用するμLC/MSにより、最初に70/25VDCの電位でノズル/スキマー(V_{NZ})を用いて、次いで200/25VDCのV_{NZ}電位を用いての再作動により、そのビオチン化ペプチドを2回分析した。

【0172】

断片化されていない潜在的ペプチドの一覧を、クロマトグラフ作動全てを手動で試験して作成した。同定可能な二重に、三重におよび四重に荷電したm/z 一同位体値(monoisotopic value)をスプレッドシート中に掲載し、そして逆重畳した。逆重畳した質量を、単一のデータベース検索プログラムに提供した。このデータベース検索プログラムは、以下の規準に適合するペプチドの一覧を与えた: 30ppmの精度を有する質量、システイニル残基の存在を含むペプチド、両端にトリプシン部位を含むペプチド、可能性のある2つのトリプシン切断部位が欠落し得る、414.1931Da(PEO-ヨードアセチル化)によって改変されたシステイン。次いで、各可能性ペプチド配列をin silicoで断片化し、可能なフラグメントイオンのm/z値の一覧を作成した。次いで、断片化したクロマトグラフ作動を、実験的にセットしたm/z範囲内にあ

10

20

30

40

50

るそれらの可能性フラグメントの存在または非存在について、手動で調べた。次いで、各可能性ペプチドの候補を、40 ppmまでの誤差範囲内で質量分析データからの実際のフラグメントとの一致として、評価した。さらに、生じたフラグメントおよび前駆物質の質量を手動で S E Q U E S T ^{T M} に入力し、そして同データベースに対して検索し、イオントラップデータと比較するために、交差相関 (X C o r r) スコアおよびデルタ相関 (d C n) スコアを作成した。最後に、同サンプルを、T h e r m o F i n n i g a n L C Q イオントラップ L C / M S (M S) 設定上で、類似の逆相分離スキームを用いて作動した。その結果を、S E Q U E S T ^{T M} (E n g r a , 前出、1994) を使用して自動データベース分析のために提示し、E S I - T O F 結果を検証した。

【 0 1 7 3 】

図5に示したこのクロマトグラフ領域のスクリーンを、非断片化作動および断片化作動の両方について要約し、そして、それぞれ、図6および図7に示す。620.3436 m/zでの三重に荷電したピークおよび930.0258 m/zでの二重に荷電したピークを、[M + H] = 1859.03に逆重畳した。酵母のオープンリーディングフレームデータベースを、30 ppmの逆重畳質量内で、この[M + H]での、完全にトリプシンから生じたシステイン含有ペプチド(2つの潜在的に失った切断を含む)について検索した。図5において、矢印は、ペプチド Y R P N C P I I L V T R (配列番号26) を指し、クロマトグラフィー上での単一のシグナルペプチドとして分離された。MSの間に、イオン選択を実施しなかったが、ペプチドはなお明確に同定された。この結果を表1に示し、ここで、配列が Y R P N C P I I L V T R (配列番号26) であるペプチドは、示される33個のイオンのうち22個により、明確に同定された(表1で、それぞれ、配列番号27~33、26および34~38)。

【 0 1 7 4 】

10

20

【 表 1 】

表 1

| 30.0 ppm の質量精度およびシステイン(PEO連結)に加えられた414.1931 Da.での[M+H] ⁺ に適合する可能性ペプチド ^{(a)(b)} | | | | |
|--|-------------------------|--------------|---|--|
| 配列 | 理論上の [M+H] ⁺ | 理論値からの 誤差 | 計器の実験上の m/z 範囲内での可能性のある フラグメントの数に一致するフラグメント イオンの数(40 ppmの誤差範囲内) | |
| LDNCVINEKGIVK | 1858.976440 | 28.8 ppm | 3/32 | |
| RPMICLKNNSLR | 1858.981201 | 26.3 ppm | 3/30 | |
| NKANICLAKNDLK | 1858.991699 | 22.8 ppm | 3/32 | |
| YLIVLTQNCHLK | 1858.991699 | 20.6 ppm | 1/30 | |
| FYTKRLCTVSVK | 1858.991699 | 20.6 ppm | 0/29 | |
| MLRNLVVRNACR | 1858.992432 | 20.2 ppm | 0/35 | |
| GCEVVVSGKLLRAAR | 1858.998901 | 16.7 ppm | 3/40 | |
| YRPNCPIILVTR | 1859.003052 | 14.5 ppm | 22/31 (XCorr: 3.24, dCn: 0.6) | |
| SIQLTLKGPSGCLK | 1859.012939 | 9.2 ppm | 0/38 | |
| VITLEKICEVAAR | 1859.012939 | 9.2 ppm | 1/30 | |
| IDRLLKCDLSR | 1859.024048 | 3.2 ppm | 1/30 | |
| RGTDVLKALCLGAK | 1859.024048 | 3.2 ppm | 0/34 | |
| VKPIRGFCQRLK | 1859.050659 | 11.1 ppm | 0/28 | |

(a) 酵母オーブンリンターディングフラグメントデータベース由来

(b) EZ-Link™ PEO-Iodoacetyl Biotin, Pierce # 21334

【 0 1 7 5 】

10

20

30

40

【表 2】

表 2

| 計器の実験上の m/z 範囲内で適合するフラグメントイオン(*)および測定された m/z 値の ppm 誤差 | | | | | | | | | | | |
|--|----|-----------------------|----------------------|----------------------|---------------------|----|--|--|--|--|--|
| Seq. | # | b (+1) | b (+2) | Y (+1) | Y (+2) | # | | | | | |
| Y | 1 | ----- | ----- | ----- | 930.0054 | 12 | | | | | |
| R | 2 | 320.1722 (36.9 ppm)* | ----- | ----- | 848.4738 (8.8 ppm)* | 11 | | | | | |
| P | 3 | 417.2250 (20.9 ppm)* | ----- | ----- | 770.4232 (7.3 ppm)* | 10 | | | | | |
| N | 4 | 531.2679 (17.9 ppm)* | ----- | 1442.7858 | 721.8968 (7.2 ppm)* | 9 | | | | | |
| C | 5 | 1048.4702 (0.7 ppm)* | 524.7390 (5.0 ppm)* | 1328.7429 | 664.8754 (4.7 ppm)* | 8 | | | | | |
| P | 6 | 1145.5229 | 573.2654 (37.7 ppm)* | 811.5406 (8.9 ppm)* | 406.2742 (3.9 ppm)* | 7 | | | | | |
| I | 7 | 1258.6071 (21.0 ppm)* | 629.8074 (7.1 ppm)* | 714.4879 (13.2 ppm)* | 357.7478 | 6 | | | | | |
| I | 8 | 1371.6912 (3.4 ppm)* | 686.3496 (0.9 ppm)* | 601.4037 (5.3 ppm)* | 301.2058 | 5 | | | | | |
| L | 9 | 1484.7753 | 742.8915 (14.5 ppm)* | 488.3196 (23.5 ppm)* | ----- | 4 | | | | | |
| V | 10 | ----- | 792.4257 | 375.2356 (1.3 ppm)* | ----- | 3 | | | | | |
| T | 11 | ----- | 842.9496 | ----- | ----- | 2 | | | | | |
| R | 12 | ----- | 921.0001 | ----- | ----- | 1 | | | | | |

表 2 は、実験的な m/z 範囲 (300 ~ 1500 m/z) 内に位置する各フラグメントの理論的質量を示し、整合性ピークは「*」で示され、それらの添付の ppm エラーと共に示される。この質量を手動で SEQUESTTM に入力し、表 1 に示される相互相関スコアおよび相関スコアを得た。最後に、このペプチドをまた、同じサンプルからのイオンラップデータを検索する場合、3.75 の XCorr および 0.325 の相関スコアで、SEQUESTTM によって見出した。

【0176】

さらに複雑なサンプルにおいて、2つのペプチドがクロマトグラフィーで共溶出した(図8を参照のこと)。共溶出ペプチドを、イオン選択を用いずに起源領域において断片化した。MSをイオン選択を用いずに行ったが、2つの共溶出ペプチドを同定した。

【0177】

図9に示されるような非断片的質量を、上記で使用したデータベース検索プログラムに

10

20

30

40

50

入力した。潜在的ペプチドをインシリコで断片化し、そして得られたフラグメント質量を図10に示される断片的スペクトル m/z 値と比較した。表3は、 $[M+H]^+ = 1486.76$ の候補ペプチドおよび整合化イオンの数を示す。ペプチド配列をVCSSTGLIR(配列番号44)であることを決定した(表3、それぞれ、配列番号39~49)。

【0178】

表 3

| 可能なペプチド整合化 [M ₁ +H] = 1486.76 @ 30.0 ppm 実際の質量 & 414.1931 Da.(システインに付加)(PEO結合化) ^{(a)(b)} | | | |
|--|--------------|---------------|---|
| 配列 | 理論値 [M+H] | 理論値からの エラー | 機器の実験的m/z範囲内 (40 ppmエラー内) での可能なフラグメントイオンの数に整合する フラグメントイオンの# |
| KHCQSQVSR | 1486.725342 | 23.3 ppm | 0/25 |
| MMIAFTCKK | 1486.728760 | 21.0 ppm | 1/25 |
| EFRCRSK | 1486.729248 | 20.7 ppm | 0/21 |
| KTLMSVCYK | 1486.746582 | 9.0 ppm | 1/26 |
| KCKNGSPNK | 1486.750366 | 6.5 ppm | 0/31 |
| VCSHTGLIR | 1486.750366 | 6.5 ppm | 9/31 (XCorr:1.5 dCn:0.708) |
| CYLQYRVK | 1486.754517 | 3.7 ppm | 0/29 |
| LTVQRPQCK | 1486.786743 | 18.0 ppm | 3/27 |
| VVQLQNICR | 1486.786743 | 18.0 ppm | 2/27 |
| FIHVPTCKK | 1486.790771 | 20.7 ppm | 1/25 |
| TIGLLCDPK | 1486.800659 | 27.3 ppm | 0/27 |

(a) 酵母オーブンリーディングフレーム
データベース由来

(b) EZ-Link™ PEO-Iodoacetyl Biotin, Pierce # 21334

【表 4】

表 4

| 可能なペプチド整合化 $[M_2+H] = 1756.84 @ 30.0$ ppm 実際の質量 & 414.1931 Da.(システインに付加)(PEO結合化) ^{(a)(b)} | | 理論値 [M+H] | 理論値からの エラー | 機器の実験的m/z範囲内(40 ppmエラー内) での可能なフラグメントイオンの数に整合する フラグメントイオンの# |
|--|-------------|--------------|------------------------------|--|
| 配列 | | | | |
| FYRCYNSYAR | 1756.793457 | 26.5 ppm | 0/25 | |
| YKDCEDWKQK | 1756.803345 | 20.8 ppm | 0/25 | |
| FTDMQLELCRSR | 1756.806641 | 19.0 ppm | 0/29 | |
| MSCSESEQFLR | 1756.806641 | 19.0 ppm | 0/29 | |
| ETYELCDQTK | 1756.813232 | 15.2 ppm | 1/26 | |
| MATDERCIFR | 1756.817871 | 12.6 ppm | 0/28 | |
| SYSSCPAVSQSVK | 1756.824341 | 16.7 ppm | 0/33 | |
| TIDFKFPECDK | 1756.828369 | 6.6 ppm | 0/29 | |
| NAPGDCVSPVQQK | 1756.835571 | 2.5 ppm | 0/33 | |
| TLACYSDAIVMR | 1756.843018 | 1.7 ppm | 1/31 | |
| GHNPCTSTISGR | 1756.846924 | 4.0 ppm | 6/32 (XCorr: 1.1 dCn: 0.751) | |
| DGVRYVIFDCR | 1756.850952 | 6.3 ppm | 0/29 | |
| MFPIGCATFLSR | 1756.858276 | 10.4 ppm | 1/31 | |
| LKGKDSFSMVCK | 1756.879395 | 22.4 ppm | 0/33 | |

(a) 酵母オーブンリーディングフレーム

データベース由来

(b) EZ-Link™ PEO-Iodoacetyl Biotin, Pierce # 21334

表 4 は、 $[M_2 + H] = 1758.84$ の候補ペプチドおよび整合化イオンの数を示す。ペプチド配列を、GHNPCTSTISGR (配列番号 60) であることを決定した (表 4、それぞれ、配列番号 50 ~ 63)。

【0180】

両方のペプチド ($[M_1 + H] = 1486.76$ および $[M_2 + H] = 1758.84$) の場合において、フラグメント m/z 値を、手動で SEQUEST™ に入力し、そし

10

20

30

40

50

てX C o r r 値およびd C n 値を得た。これらを、それぞれ表 3 および表 4 に列挙する。

【 0 1 8 1 】

表 5 および表 6 は、整合化されたフラグメントイオン(「*」)、フラグメントの理論的m/z、および実際のm/z 値のppmエラーを示す。最後に、これらのペプチドをまた、同じサンプルからのイオントラップデータを検索する場合、[M 1 + H] = 1 4 8 6 . 7 6 について、2 . 8 4 のX C o r r および0 . 1 9 8 の 相関スコア、ならびに[M 2 + H] = 1 7 5 8 . 8 4 について、3 . 1 8 のX C o r r および0 . 2 9 9 の 相関スコアで、S E Q U E S T ^{T M} によって見出した。

【 0 1 8 2 】

【表 5】

表 5

| 機器の実験的m/z 範囲 および 測定されたm/z 値の ppm エラー 内の 整合化フラグメントイオン (*) | | | | | | |
|--|----|-----------|----------|----------------------|----------------------|----|
| Seq. | # | b (+1) | b (+2) | Y (+1) | Y (+2) | # |
| V | 1 | ----- | ----- | 1468.7506 | 743.8792 | 10 |
| C | 2 | 617.2786 | 309.1432 | 1387.6821 | 694.3450 (34.6 ppm)* | 9 |
| S | 3 | 704.3106 | 352.6592 | 870.4798 (8.3 ppm)* | 435.7438 (1.4 ppm)* | 8 |
| S | 4 | 791.3427 | 396.1752 | 783.4478 (12.0 ppm)* | 392.2278 (11.5 ppm)* | 7 |
| H | 5 | 928.4016 | 464.7047 | 696.4158 (16.5 ppm)* | 348.7118 (5.7 ppm)* | 6 |
| T | 6 | 1029.4492 | 515.2285 | 559.3569 (0.9 ppm)* | ----- | 5 |
| G | 7 | 1086.4707 | 543.7393 | 458.3091 (12.9 ppm)* | ----- | 4 |
| L | 8 | 1199.5548 | 600.2813 | 401.2876 | ----- | 3 |
| I | 9 | 1312.6389 | 656.8234 | ----- | ----- | 2 |
| R | 10 | 1468.7400 | 734.8739 | ----- | ----- | 1 |

【 0 1 8 3 】

10

20

30

40

【表 6】

153

表 6

| 機器の実験的m/z 範囲および測定されたm/z 値の ppm エラー内での 整合化フラグメントイオン (*) | | | | | | |
|--|----|----------------------|----------|----------------------|----------------------|----|
| Seq. | # | b (+1) | b (+2) | Y (+1) | Y (+2) | # |
| G | 1 | ----- | ----- | ----- | 878.9274 | 13 |
| H | 2 | ----- | ----- | ----- | 850.4166 | 12 |
| N | 3 | 309.1311 (30.7 ppm)* | ----- | ----- | 781.8872 | 11 |
| I | 4 | 422.2152 (5.0 ppm)* | ----- | 1448.7235 | 724.8657 | 10 |
| P | 5 | 519.2679 | ----- | 1335.6395 | 668.3237 (20.6 ppm)* | 9 |
| C | 6 | 1036.4702 | 518.7390 | 1238.5868 | 619.7973 | 8 |
| T | 7 | 1137.5179 | 569.2629 | 721.3845 (14.6 ppm)* | 361.1961 | 7 |
| S | 8 | 1224.5499 | 612.7789 | 620.3368 | 310.6723 | 6 |
| T | 9 | 1325.5977 | 663.3027 | 533.3047 | ----- | 5 |
| I | 10 | 1438.6818 | 719.8448 | 432.2570 (17.4 ppm)* | ----- | 4 |
| S | 11 | ----- | 763.3608 | 319.1730 (18.5 ppm)* | ----- | 3 |
| G | 12 | ----- | 791.8715 | ----- | ----- | 2 |
| R | 13 | ----- | 869.9221 | ----- | ----- | 1 |

10

20

30

40

これらの結果は、ペプチド配列のポジティブ同定が、純粋な形態または他のペプチドと同時に（すなわち、それぞれ単一または複数の溶出ペプチド）、ペプチドについて得られることを示す。共溶出ペプチドを、一段式ESI-TOF質量分析計で起源内CIDを用いて「ショットガン」配列決定した。タンデムMSを介して従来のペプチド配列決定に関係して、親イオンは連続して選択され、システイン制約の組み合わせ、高い質量正確度お

50

よび高い質量分解能は、ペプチドが同時にショットガン配列決定されることを可能にする。この方法はまた、1つの分析におけるICAT標識化ペプチドの分析のためのICAT (Gygiら、Nature Biotechnol. 17:994-999 (1999))で使用され得る。

【0184】

これらの結果は、親タンパク質が決定された特徴の組み合わせから同定され得ることを示す。この実施例において、決定された特徴は、(1)システイン含量、(2)正確なペプチド質量および(3)特定のイオンの選択なしでの正確なペプチドフラグメント質量、である。

【0185】

(実施例V)

(ポリペプチド識別インデックスとしてのDrosophilaデータベースの作製)
この実施例は、Drosophilaタンパク質を同定するためのデータベースの作製を記載する。

【0186】

Drosophilaを、完全なゲノム配列が決定され、そしてゲノム配列が多くのエクソンを含む種の代表的なモデル生物として選択した。実験の目的は、完全なゲノム配列が決定されている種由来のペプチドおよびタンパク質の大きなセットの作製、多次元クロマトグラフィーによってペプチドを分離すること、CIDスペクトルおよび配列データベース検索の作製を介してペプチドを同定すること、および分画/同定プロセスの間、得られた情報のバーコードでそれぞれの同定したペプチドに注釈を付けることである。

【0187】

Drosophila melanogaster由来のシュナイダー(S2)細胞(10^9)を、適切な組織培養培地での培養によって調製した。細胞を溶解し、そして得られた溶解物を、核画分、細胞質画分およびミクロソーム画分への差次的遠心分離によって分画した。それぞれの画分を、2つの等量に分け、そしてそれぞれのサンプル中に含まれるタンパク質を、同位体的に通常ICAT試薬(d0)または同位体的に重いICAT試薬(d8)のいずれかで標的化した。次いで、同じ細胞成分分画を表す標的化サンプルを合わせ、1:1の同位体的標的化タンパク質のタンパク質混合物を生成し、そしてトリプシン処理した。得られたペプチドを、実施例IIに記載されたように陽イオン交換クロマトグラフィー、次にタグ化ペプチドの選択的単離のためにアビジンクロマトグラフィー、および逆相クロマトグラフィーによって分画した。分離したペプチドを分析し、そしてLC/Qイオンラップ質量分析計中のCIDに供した。得られたペプチドCIDスペクトルを、ハエゲノム検索コンソーシアムおよびSequenceTM検索アルゴリズムからのハエ配列データベースを用いて配列データベース検索に供した。

【0188】

ペプチドCIDスペクトルを用いた配列データベース検索によってタンパク質を同定するために、以下の検索パラメータを使用した:最低1つのシステイン残基を含むダブルトリプシンペプチドのみを考えた。0.1より大きい相関因子を必要とし、そして2.0より大きいX-相関係数を、二重に荷電した前駆体イオンのために必要とした(単一に荷電したペプチドに対して、 $X\text{-corr.} > 1.5$; 三重に荷電したペプチドに対して、 $X\text{-corr.} > 2.5$)。

【0189】

核画分に対して、1470のタンパク質を同定した。細胞質画分に対して、967のタンパク質を同定した。ミクロソーム画分に対して、1500をこえるタンパク質を同定した。従って、この分析において、約4000のタンパク質を同定した。従ってこれらのデータは、多次元クロマトグラフィー、タンデム質量分析および特定の組織、細胞または種の注釈されたペプチドデータベースを確立するための配列データベース検索の使用を補助する。

【0190】

10

20

30

40

50

これらの結果は、多数のポリペプチドの同定のために使用され得るポリペプチド識別インデックスが作製され得ることを示す。

【0191】

(実施例VI)

(ポリペプチド同定の識別特徴としての等電点の使用)

この実施例は、ペプチドの同定の特徴として等電点(pI)の使用を記載する。

【0192】

ペプチドの等電点を、ペプチドバーコードに有用な識別成分として試験した。バーコードの成分としてpIを用いる利点を挙げる：i)pIは、ペプチドのアミノ酸配列から容易にそして正確に予測可能である；ii)pIは、IEFによって正確に測定され得る；そしてiii)ペプチドは、IEFゲルから効率的に回収され得る。

10

【0193】

実施例Vに記載された八エペプチドサンプルのイオン交換画分を、固定化pH勾配(IPG-IEF)を用いてポリアクリルアミドゲルでの等電集束法によってさらに分離した。約pI6.8~7.0のセグメントに集束するペプチドを切りだし、そしてこの画分に含まれるペプチドを抽出し、そしてLC-MS/MSによって分析した。得られたCIDスペクトルを、実施例Vに記載されるように配列データベース検索に供し、そしてそれぞれの同定されたペプチドのpIをまた、最良の配列データベースに整合する配列を用いて計算した。表7は、この方法の例示としての結果の部分的リスト(配列番号64から117)を示す。ほとんど同定されたペプチドのpIは、狭いpI範囲付近にかたまり、pIは配列から正確に計算され得ることを示した。

20

【0194】

データは、最良の整合がゲルから抽出された範囲の外側のpIを生じる、(少ない)数のペプチドが存在することをさらに示した。検索スコアが、辺縁性であり、ペプチドが高比率の欠損トリプシン切断部位を通常含み、そしてこのペプチドの多くは非トリプシンであるため、これらのペプチドのより密接な試験は、配列データベース整合が、抽出されたゲル領域のpI範囲の外側のこれらのペプチドに対してほとんど確実に誤っていたことを示唆した。集団的に、これらのデータは、ポリペプチド識別インデックスからポリペプチドを同定するためのバーコードの成分としてペプチドpIを用いるの能力を示す。

【0195】

30

ある。

【図面の簡単な説明】

【0197】

【図1】図1は、質量分析器(MS)測定、およびタンデム質量分析器(MS/MS)測定に基づくタンパク質同定戦略の概略図を示す。

【図2】図2は、親イオンの同定に特徴的なペプチドイオンから選択されたフラグメントイオンを生成するための2つの異なる方法を示す。図2Aは、親イオンの選択を示し(Q1において)、この親イオンは、衝突セル中でフラグメント化される(Q2)。このフラグメントの質量スペクトルが決定される(Q3において)。図2Bは、単一の親イオンの選択に代わって、多数の親イオン(「供給源領域」中に示される)は、イオン化後領域または衝突セルにおいて、同時にフラグメント化されることを示す。次いで、このフラグメントイオンは、Q1または他の質量分析器中で分析され、その結果、多数の親イオンからのフラグメントイオンからなる質量スペクトルを生じる。

10

【図3】図3は、注釈されたペプチドタグのポリペプチド識別インデックスを使用して、2つのポリペプチド集団を比較し、そして定量するための方法の工程を示す。

【図4-1】図4Aおよび図4Bは、質量分析器(MS)を使用したポリペプチドの同定を示す。図4Aおよび図4Bは、ESI-TOFを使用して得た2つのポリペプチドの質量スペクトル(P1、ASHLGLAR(配列番号1);およびP2、RPRGFSPR(配列番号2))を示す。スペクトルは、低V_{ノズル}-スキマー(10V)(図4A)および高V_{ノズル}-スキマー(240V)(図4B)で取得される。

20

【図4-2】図4Cは、質量分析器(MS)を使用したポリペプチドの同定を示す。図4Cは、P1(各々、配列番号3~12、2、13および14)およびP2(配列番号15~24、1および25)についての、各々、12の可能なポリペプチド同定物および13の可能なポリペプチド同定物を示す。

【図5】図5は、*Saccharomyces cerevisiae*抽出物のクロマトグラフィー分析を示す。矢印で示されたペプチドは、非断片的MS分析および断片的MS分析によって解析された。

【図6】図6は、図5中の矢印によって示されたペプチドの非断片的MS分析を示す(YRPNCPIILVTR;配列番号26)。

【図7】図7は、図5中の矢印によって示されたペプチドの断片的MS分析を示す(YRPNCPIILVTR;配列番号26)。

30

【図8A】図8は、クロマトグラフィー分析を示す。図8Aは、基準ピーククロマトグラムを示す。

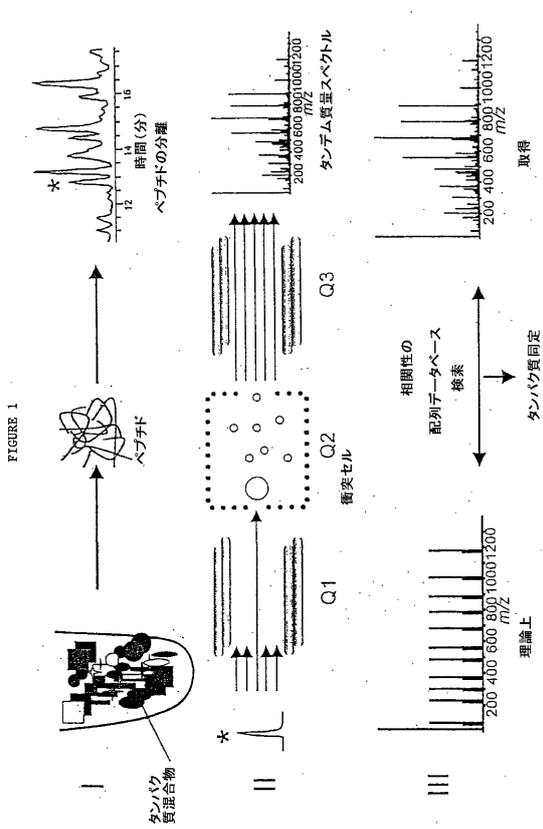
【図8B】図8は、クロマトグラフィー分析を示す。図8Bは、496.26 m/z [M₁+3H]での選択されたイオンクロマトグラムを示す。

【図8C】図8は、クロマトグラフィー分析を示す。図8Cは、586.30 m/z [M₂+3H]での選択されたイオンクロマトグラムを示す。

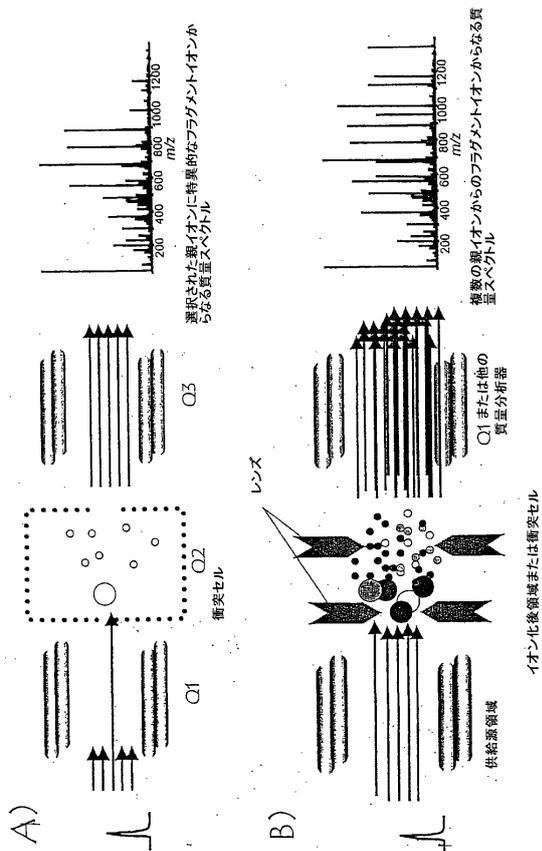
【図9】図9は、非断片化MS分析を示す。

【図10】図10は、断片化MSを示す。

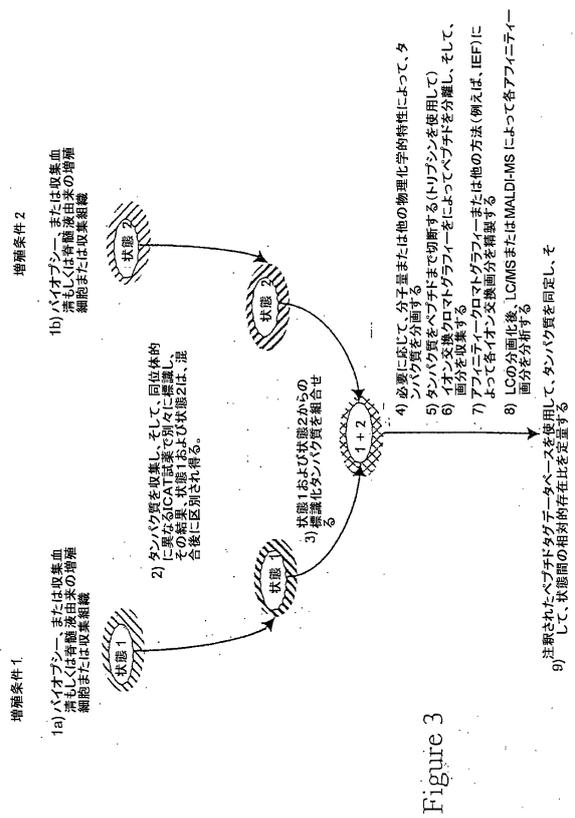
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 - 1 】

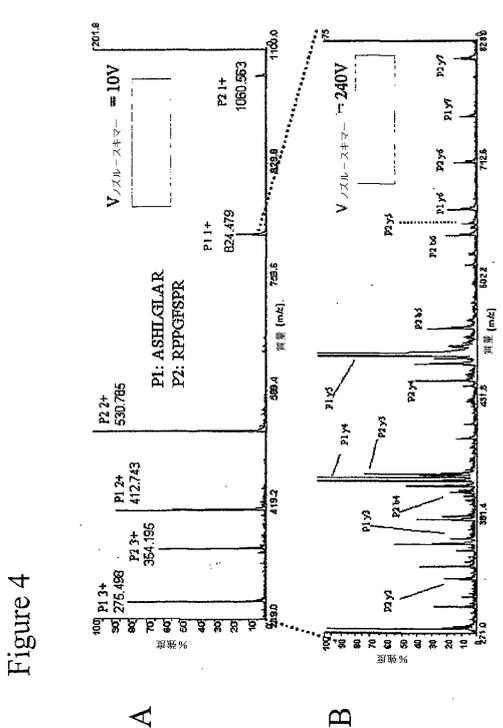


Figure 3

Figure 4

FIGURE 1

FIGURE 2

Figure 3

Figure 4

【 図 4 - 2 】

C

ヒトタンパク質データベース :
60,884 配列エントリー

分析に使用されるフラグメントイオン質量 :

710.385
753.435
807.374
886.402
904.451
1001.564

ブラジネニン RPPGFSPFR MH+ 1060.5676
(10 PPM, 144トリプシンペプチドマッチ)

フラグメントイオン制約を使用した13 マッチ :

ACISEILPSK
GVRYSEGFK
RANLISQCR
RCGLPSSGKR
RDITLEASR
RERETLEK
RLTEEBRK
RLVEVDSSR
RNLLDHR
RPHAAOPGAR
RPPGFSPFR **
RPQTATASTK
RRPSAYQAL

ASHLGLAR MH+ 824.4739
(10PPM, 57トリプシンペプチドマッチ)

フラグメントイオン制約を使用した12 マッチ:

ACYIKVK
ADPLPRR
APYAFAAK
AFVFGRK
AHAEIRK
AHEAKR
ALEAHR
ALOFFAK
AMAIYKK
APDPLRL
ASHLGLAR **
AVAGHLTR

FIGURE 4

【 図 5 】

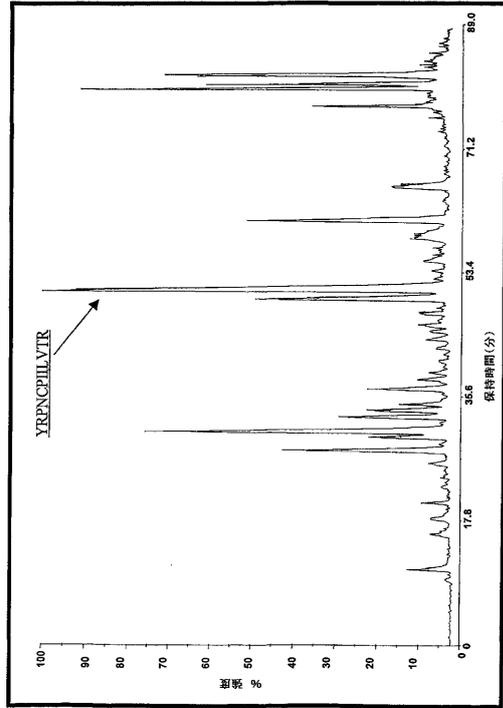


FIGURE 5

【 図 6 】

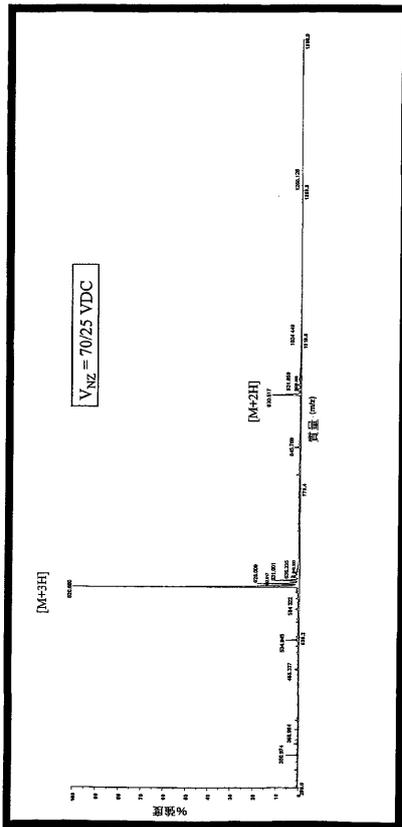


FIGURE 6

【 図 7 】

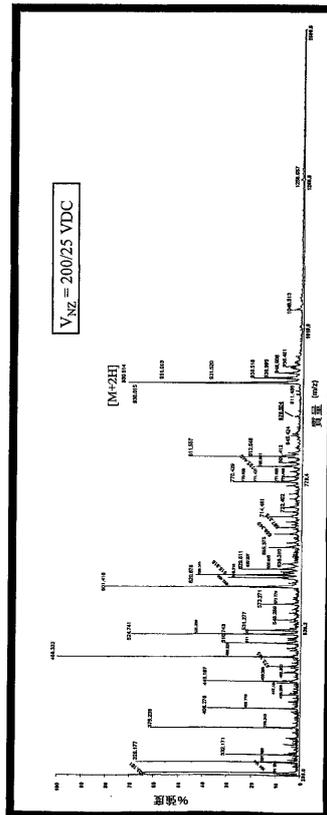


FIGURE 7

【 8 A 】

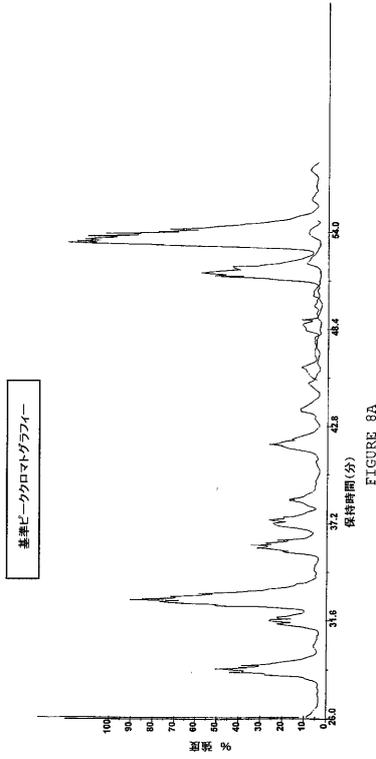


FIGURE 8A

【 8 B 】

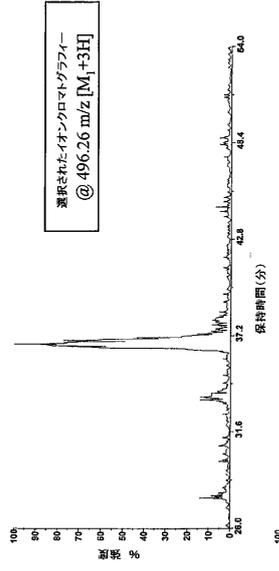


FIGURE 8B

【 8 C 】

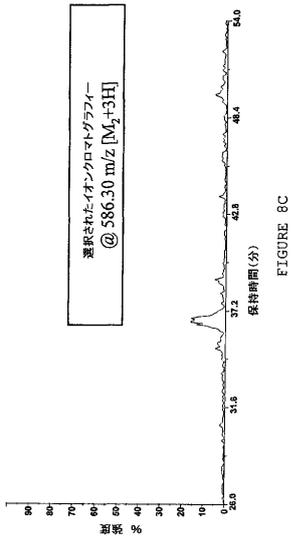


FIGURE 8C

【 9 】

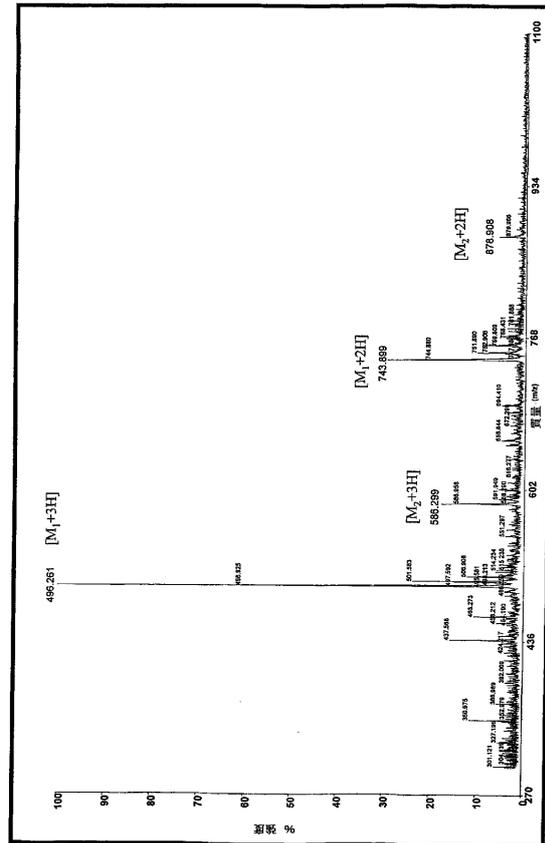
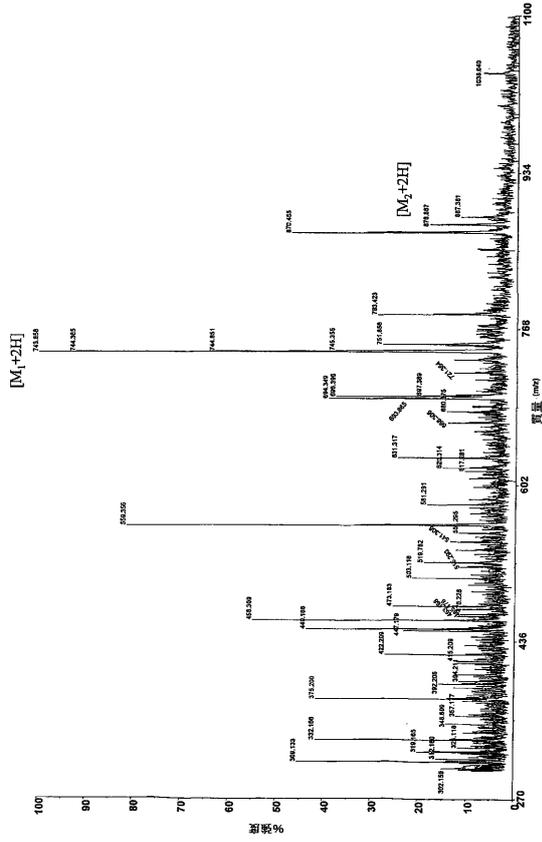


FIGURE 9

【 10 】



フロントページの続き

(74)代理人 100078282

弁理士 山本 秀策

(74)代理人 100062409

弁理士 安村 高明

(74)代理人 100113413

弁理士 森下 夏樹

(72)発明者 グッドレット, デイビッド アール.

アメリカ合衆国 ワシントン 98199, シアトル, トゥウェンティー - ナインス アベニ
ュー ウエスト 4411

(72)発明者 エーバーソルド, ルドルフ エイチ.

アメリカ合衆国 ワシントン 98040, マーサー アイランド, エスイー 78ティーエ
イチ ストリート 8609

審査官 島田 英昭

(56)参考文献 John R.Yates, "Mass Spectrometry and the Age of the Proteome", Journal of Mass Spectrometry, 1998年 2月19日, Vol.33, pp.1-19

Steven P.Gygi, et.al., "Quantitative analysis of complex protein mixtures using isotop
e-coded affinity tags", Nature Biotechnology, 1999年 7月13日, Vol.17, OCTOBER
, pp.994-999

(58)調査した分野(Int.Cl., D B名)

G01N27/62-27/70

H01J49/00-49/48

JSTPlus(JDream2)