

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5181908号
(P5181908)

(45) 発行日 平成25年4月10日(2013.4.10)

(24) 登録日 平成25年1月25日(2013.1.25)

(51) Int.Cl.

G O 1 N 27/62 (2006.01)

F I

G O 1 N 27/62

D

請求項の数 3 (全 12 頁)

(21) 出願番号 特願2008-200332 (P2008-200332)
 (22) 出願日 平成20年8月4日(2008.8.4)
 (65) 公開番号 特開2010-38664 (P2010-38664A)
 (43) 公開日 平成22年2月18日(2010.2.18)
 審査請求日 平成23年7月7日(2011.7.7)

(73) 特許権者 000001993
 株式会社島津製作所
 京都府京都市中京区西ノ京桑原町1番地
 (74) 代理人 110001069
 特許業務法人京都国際特許事務所
 (74) 代理人 100095670
 弁理士 小林 良平
 (72) 発明者 梶原 茂樹
 京都市中京区西ノ京桑原町1番地 株式会
 社島津製作所内
 (72) 発明者 山田 真希
 京都市中京区西ノ京桑原町1番地 株式会
 社島津製作所内

最終頁に続く

(54) 【発明の名称】 質量分析データ解析装置

(57) 【特許請求の範囲】

【請求項1】

M Sⁿ分析(n = 2)可能な質量分析計により収集されたスペクトルデータに基づいて被検試料中のペプチド・タンパク質を同定するためにデータベース検索を行う質量分析データ解析装置において、

a) スペクトルデータに基づいて被検試料に由来するピークのピーク情報を集めたピークリストを作成するピークリスト作成手段と、

b) アミノ酸配列に起こり得る修飾・変異に関する条件を任意の数だけユーザが指定するための第1条件指定手段と、

c) 該第1条件指定手段により指定される条件以外のデータベース検索のための検索条件をユーザが指定するための第2条件設定手段と、

d) 前記第1条件指定手段により指定された条件を順次1つずつ選択しつつ選択した条件と前記第2条件指定手段により指定された検索条件とを組み合わせた条件の下で、それぞれ前記ピークリスト作成手段により作成されたピークリストをデータベースと照合することによりペプチド・タンパク質の候補を抽出する、という処理を、前記第1条件指定手段により指定された条件の全てが終了するまで又は所定の終了条件が満たされるまで繰り返すデータベース検索実行手段と、

e) 前記データベース検索実行手段により抽出された各ペプチド・タンパク質の候補の確度に基づいた情報を出力する出力手段と、

を備え、非確定的な修飾・変異が最大でも1箇所では生じないとの排他的条件の下に

10

20

、前記第1条件指定手段により指定可能な非確定的修飾・変異の種類の数
の制約をなくす又は広げるようにしたことにより、被検試料中のペプチド・タンパク質を同定する際の擬陽性の減少又はその増加の抑制と同定見逃しの減少とを両立するとともに検索時間の増大を抑制するようにしたことを特徴とする質量分析データ解析装置。

【請求項2】

請求項1に記載の質量分析データ解析装置であって、

アミノ酸配列の修飾・変異についての既知の全ての条件を格納しておく修飾・変異データベースを備え、前記第1条件指定手段は、修飾・変異データベースを利用して任意の数の条件の指定を可能としたことを特徴とする質量分析データ解析装置。

【請求項3】

請求項1又は2に記載の質量分析データ解析装置であって、

前記データベース検索実行手段は、抽出されたペプチド・タンパク質の候補の確度が所定の閾値以上であるものが探索された時点でデータベース検索の繰り返しを打ち切ることを特徴とする質量分析データ解析装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、タンパク質由来のペプチド混合物を含む被検試料を質量分析し、これにより得られた質量分析データを用いてペプチドのアミノ酸配列を推定することによってタンパク質を同定する質量分析データ解析装置に関する。

【背景技術】

【0002】

近年、ポストゲノム研究としてタンパク質の構造や機能の解析が急速に進められている。このようなタンパク質の構造・機能解析手法（プロテオーム解析）の一つとして、質量分析計を用いたタンパク質の発現解析や一次構造解析が広く行われるようになってきており、四重極型イオントラップや衝突誘起分解（CID）などによって特定のピークの捕捉と開裂を行う、いわゆるMSⁿ分析（nは2以上の整数）が威力を発揮している。一般にMS²（=MS/MS）分析では、まず、分析対象物から特定のm/zを有するイオンをプリカーサイオンとして選別し、該プリカーサイオンをCIDによって開裂させる。その後、開裂によって生成したイオン（プロダクトイオン）を質量分析することによって、目的とするイオンの質量や化学構造についての情報を得ることができる。

【0003】

上記のようなMSⁿ分析を利用してタンパク質を同定する場合、まず、タンパク質を適当な酵素で消化してペプチド断片の混合物としてから、該ペプチド混合物を質量分析してマススペクトルを取得する。続いて、ペプチド混合物のマススペクトルデータの中から、例えば単一のペプチドに由来する一組の同位体ピーク群をプリカーサイオンとして選択し、該プリカーサイオンをCIDにより開裂させて生成されたイオン（プロダクトイオン）の質量分析、つまりMS/MS分析を行う。また、1回の開裂操作では十分に小さな断片に開裂しない場合には、開裂操作を複数回行うこともある。

【0004】

以上のようにして1段又は2以上の多段の開裂操作により得られたプロダクトイオンのマススペクトルパターンや上記プリカーサイオンのマススペクトルパターンを基に、例えばマトリックスサイエンス社が提供しているマスコット（MASCOT）等の検索エンジンを利用してアミノ酸配列データベースとの照合処理を実行する。これによって、被検ペプチドのアミノ酸配列を決定し、元のタンパク質を同定する。

【0005】

上記のようなデータベース検索エンジンとしては様々なものが知られ実用に供されているが、上述したマスコットは最も広く利用されているものの1つである（非特許文献1、2参照）。図4は、コンピュータのモニタ画面上に表示される、マスコットによるMS/MSイオンサーチの検索設定画面の一例である。

【 0 0 0 6 】

ユーザが設定すべき主な検索項目としては、照合に使用するデータベース (Database)、タンパク質の分解に使用した消化酵素の種類 (エンザイム: Enzyme)、確定的に起こる修飾の種類 (フィックスド・モディフィケーション: Fixed modification)、起こる可能性のある (非確定的な) 修飾の種類 (バリアブル・モディフィケーション: Variable modification)、質量分析の精度の許容値 (MS/MS tol.) などがある。これら検索項目のいくつかはプルダウンメニューとなっており、そのメニューの中からユーザが適宜に指定できるようになっている。ユーザがこれら検索条件を適宜設定した上で「スタートサーチ (Start Search)」ボタンをクリック操作して検索の実行を指示すると、コンピュータにおいて設定検索条件の下でデータベースの照合が実行され、最終的には検索上で確度が高いペプチド候補がリストアップされる。マスコットの場合には、ペプチド候補の確度はスコアと呼ばれるものである。

10

【 0 0 0 7 】

しかしながら、ユーザが常に正しく各検索項目を設定できるとは限らない。例えば、精度許容値「MS/MS tol.」を小さく見積もってしまったり正しいバリアブル・モディフィケーションを予想することが困難であったりすることは頻繁に起こり得る。このような問題を解決するため、予めタンパク質の種類を絞った上で検索条件の設定誤差を許容するエラートレナントサーチと呼ばれる方法も開発されている (非特許文献 3、4 参照)。

【 0 0 0 8 】

エラートレナントサーチ法を簡単に説明すると、まずユーザが図 4 に示したような検索設定画面において設定した検索条件の下で MS / MS イオンサーチを実行し、予め設定されたホモロジーレベルと呼ばれる閾値以上のスコアでヒットしたペプチドを含むタンパク質候補を抽出する。この MS / MS イオンサーチでは、バリアブル・モディフィケーションとしてユーザが設定したものを使用する。そうして抽出したタンパク質候補のペプチドにおいて、次の 4 つの (A)、(B)、(C)、(D) の検索条件の設定誤差を許容する条件の下でデータベース検索を試みる。

20

【 0 0 0 9 】

(A) エンザイムの検索項目を「セミ・スペシフィック (semi-specific)」に設定する。ここで、セミ・スペシフィックとは、酵素による特異的切断が片側のみで起こるものである。

30

(B) 「ミスド・クリベージ (missed cleavage)」の数を 1 だけ増加させる。ここで、ミスド・クリベージとは、本来は切断されるべきでない箇所が開裂により結合が切れることを意味する。

(C) 全てのバリアブル・モディフィケーションを 1 ずつ順番に設定する。

(D) 1 つのアミノ酸の変異 (substitution) を考慮する。

上記のようなデータベース検索によりヒットしたペプチド候補の中で、MS / MS イオンサーチ実行時よりもスコアが高く、且つ予め設定されたアイデンティファイレベル (上記ホモロジーレベルよりも高いレベル) と呼ばれる閾値以上のスコアでヒットしたペプチドを選択し、これを検索結果として表示する。

【 0 0 1 0 】

40

このエラートレナントサーチ法では、最初に MS / MS イオンサーチを実行しタンパク質を限定したうえで条件を緩和した検索を行うので、全てのタンパク質を検索対象とする通常の MS / MS イオンサーチよりも処理時間を短くすることが可能である。

【 0 0 1 1 】

しかしながら、上述の MS / MS イオンサーチ法及びエラートレナントサーチ法のいずれでも次のような問題がある。即ち、これら検索方法の良否の評価として、大別して、検索時間と、検索確度とがある。検索確度には、検索がヒットしないという見逃しの問題と、偽のペプチドが誤ってヒットしてしまうという擬陽性の問題、とがある。一般的に言えば、検索時間を短縮しようとするすると検索確度が低下し、検索確度を上げようとするすると検索時間が長くなる。

50

【 0 0 1 2 】

例えばエラートレラントサーチ法では、最初に検索対象のタンパク質を限定してしまうため、その時点で目的のタンパク質を除外してしまうと、上記手順によるデータベース検索の際に設定誤差を拡げてヒットすることは期待できない。その結果、検索時間は短いが、見逃しが多くなる。一方、MS/MSイオンサーチ法では、特定のタンパク質に絞る作業を行わないので、検出見逃しは相対的に少ないが、検索時間は長くなる。またマスコットでは、検索時間の増大と擬陽性の増加を防止するために、バリアブル・モディフィケーションの数を9に制限しているために、或る程度の見逃しが生じることは避けられない。

【 0 0 1 3 】

【非特許文献1】「マトリックス・サイエンス - マスコット - MS/MS・イオン・サーチ (Matrix Science - Mascot - MS/MS Ions Search)」、[online]、英国マトリックス・サイエンス社 (Matrix Science Ltd.)、[平成20年6月24日検索]、インターネット <URL : http://www.matrixscience.com/search_form_select.html >

【非特許文献2】飯田哲生、ほか3名、「LCMS - IT - TOFのプロテオーム解析への応用」、島津評論、Vol.63、No.1・2、p.19-28 (2006)

【非特許文献3】「エラー・トレラント・サーチ (Error Tolerant Search)」、[online]、英国マトリックス・サイエンス社 (Matrix Science Ltd.)、[平成20年6月24日検索]、インターネット <URL : http://www.matrixscience.com/help/error_tolerant_help.html >

【非特許文献4】デビッド (David M. Creasy) ほか1名、「エラー・トレラント・サーチング・オブ・アンインタプリテッド・タンデム・マス・スペクトロメトリー・データ (Error Tolerant Searching of Uninterpreted Tandem Mass Spectrometry Data)」、プロテオミクス (Proteomics)、Volume 2、Issue 10、p.1426-1434 (2002)

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 4 】

上述のようにMS/MSイオンサーチ法とエラートレラントサーチ法とを比較した場合、後者は検索時間の点で有利である反面、確度の点では劣り、前者は確度の点で有利である反面、検索時間では劣ると言える。但し、MS/MSイオンサーチ法でも見逃しの確率は比較的多く、同定されないタンパク質が残るという問題が顕在化している。本願発明者の検討により、その主な原因は、検索条件の1つであるバリアブル・モディフィケーションの数が9個以上であるが一定値に制限されている点にあることが明らかになった。また同時に、バリアブル・モディフィケーションを拡げれば見逃しの確率は下がるものの、検索時間が延びるのはもちろんのこと、擬陽性が増加し、全体として検索確度を上げることは難しいことも判明した。

【 0 0 1 5 】

本発明はこうした点に鑑みて成されたものであり、その目的とするところは、検索のための処理時間の増大をできるだけ抑えつつ、従来のMS/MSイオンサーチ法やエラートレラントサーチ法では見つけることができなかったペプチド・タンパク質を同定することができ、且つ、誤った同定も減らすことができる質量分析データ解析装置を提供することにある。

【課題を解決するための手段】

【 0 0 1 6 】

本願発明者は、MS/MSイオンサーチ法を始めとするデータベース検索の検索条件設定として、バリアブル・モディフィケーション (つまり非確定的な修飾) の数と、検索実行に際してのその検索条件の与え方に着目した。即ち、前述のように、従来の方法では、最大9種類の非確定的修飾の設定が可能であるが、検索に際しては複数の非確定的修飾の併存が許容されている。つまり、或る1つの種類の非確定的修飾と別の種類の非確定的修飾とが同時に起こり得ることも、検索の条件に組み込まれている。しかしながら、一般に

、或るペプチドのアミノ酸配列において複数箇所ですべて同時に修飾が起こることは皆無ではないが、確率としてはかなり低いと考えられる。

【 0 0 1 7 】

一方、本願発明者の検討によれば、アミノ酸配列の複数箇所ですべて同時に修飾・変異が起こるとした場合、実際のアミノ酸配列とは異なるにも拘わらず検索の際に高いスコアでヒットしてしまう確率、つまり擬陽性の確率が大幅に上がることが判明した。また、アミノ酸配列の複数箇所ですべて同時に修飾・変異が起こるという条件設定は、特に非確定的修飾・変異の数を増やす場合に検索回数的大幅な増大に繋がり、検索時間の点でも不利である。そうしたことを総合的に考慮すると、非確定的修飾・変異の複数の併存を認めない、つまり最大でも1箇所ですべて修飾・変異が生じないと仮定し、その代わりに選び得る非確定的修飾・変異の種類を大幅に増加させることは、擬陽性を減らしながら或いはその増加を抑えながら、見逃しを減らすのに有効であると考えられる。また、検索時間の点でも極端な増大は抑制できると考えられる。本発明はこうした知見に基づいてなされたものである。

10

【 0 0 1 8 】

即ち、上記課題を解決するために成された本発明は、MSⁿ分析(n ≥ 2)可能な質量分析計により収集されたスペクトルデータに基づいて被検試料中のペプチド・タンパク質を同定するためにデータベース検索を行う質量分析データ解析装置において、

a) スペクトルデータに基づいて被検試料に由来するピークのピーク情報を集めたピークリストを作成するピークリスト作成手段と、

b) アミノ酸配列に起こり得る修飾・変異に関する条件を任意の数だけユーザが指定するための第1条件指定手段と、

20

c) 該第1条件指定手段により指定される条件以外のデータベース検索のための検索条件をユーザが指定するための第2条件設定手段と、

d) 前記第1条件指定手段により指定された条件を順次1つずつ選択しつつ選択した条件と前記第2条件指定手段により指定された検索条件とを組み合わせた条件の下で、それぞれ前記ピークリスト作成手段により作成されたピークリストをデータベースと照合することによりペプチド・タンパク質の候補を抽出する、という処理を、前記第1条件指定手段により指定された条件の全てが終了するまで又は所定の終了条件が満たされるまで繰り返すデータベース検索実行手段と、

e) 前記データベース検索実行手段により抽出された各ペプチド・タンパク質の候補の確度に基づいた情報を出力する出力手段と、

30

を備え、非確定的な修飾・変異が最大でも1箇所ですべて生じないとの排他的条件の下に、前記第1条件指定手段により指定可能な非確定的修飾・変異の種類の数の制約をなくす又は広げるようにしたことにより、被検試料中のペプチド・タンパク質を同定する際の擬陽性の減少又はその増加の抑制と同定見逃しの減少とを両立するとともに検索時間の増大を抑制するようにしたことを特徴としている。

【 0 0 2 0 】

現在知られている非確定的修飾・変異の種類は600程度であるため、これらを一々、検索対象に含めるか否かをキー操作やマウス操作などで指定するのは実用的でない。そこで、本発明に係る質量分析データ解析装置の一態様として、好ましくは、アミノ酸配列の修飾・変異についての既知の全ての条件を格納しておく修飾・変異データベースを備え、前記第1条件指定手段は、修飾・変異データベースを利用して任意の数の条件の指定を可能とした構成とするとよい。

40

【 0 0 2 1 】

この構成では、一例として、第1条件指定手段は、修飾・変異データベースに登録されている全ての非確定的修飾・変異を検索条件として指定することができる。また、データベース中の全ての非確定的修飾・変異について起こり得る確率の高さ等に基づく優先度を定めておき、その優先度の順に適宜の数だけ、例えば、「優先度の高い順に100個」というように簡便に非確定的修飾・変異についての条件を設定することができる。

50

【 0 0 2 2 】

なお、第 2 条件設定手段により指定される検索条件とは、上述の従来の MS / MS イオンサーチにおいて設定可能なバリアブル・モディフィケーション以外の各項目とすることができ、例えばエンザイム、フィックスド・モディフィケーション、ミスド・クリベージなどを含む。

【 0 0 2 3 】

また、データベース検索実行手段は、第 1 条件指定手段により指定された非確定的修飾・変異の条件の全てが終了するまで検索を繰り返してもよいが、確度が相当に高いペプチド・タンパク質がヒットした時点で検索を終了するようにしておくことで、検索時間短縮の大きな効果が見込める。そこで、本発明に係る質量分析データ解析装置の一態様として、データベース検索実行手段は、抽出されたペプチド・タンパク質の候補の確度が所定の閾値以上であるものが探索された時点でデータベース検索の繰り返しを打ち切るようにしても構わない。

10

【 発明の効果 】

【 0 0 2 4 】

本発明に係る質量分析データ解析装置によれば、例えばマスコットの MS / MS イオンサーチ法などによるデータベース検索に比べて、タンパク質・ペプチドが同定できないというケースを大幅に減らすことができるとともに、誤った同定（擬陽性）も減らすことができる。一方で、一般的に検索時間は長くなるものの、従来のマスコットの MS / MS イオンサーチ法で単にバリアブル・モディフィケーションの制限を拡げる場合に比べれば、検索時間の増加を抑えることができ、通常のコンピュータの使用により許容できる程度の時間に検索時間を収めることが可能となる。さらにまた、適当に良好な検索結果が得られた時点で検索を打ち切るような処理を実施することで、検索時間の短縮が可能である。

20

【 発明を実施するための最良の形態 】

【 0 0 2 5 】

以下、本発明に係る質量分析データ解析装置を用いたタンパク質同定システムの一実施例について、添付の図面を参照して説明する。図 1 は本実施例のタンパク質同定システムの全体構成図、図 2 は本システムを用いたタンパク質同定の手順を示すフローチャートである。

【 0 0 2 6 】

本実施例のタンパク質同定システムは、大別して、質量分析部 1 と、コンピュータを中心に構成される制御・処理部 2 と、から成る。質量分析部 1 はイオントラップ飛行時間型質量分析計であり、目的試料中の分子や原子をイオン化するイオン化部 10 と、発生したイオンを一時的に捕捉し、必要に応じて m/z に応じたイオンの選別と CID によるイオンの開裂とを実行する 3 次元四重極型のイオントラップ 11 と、イオントラップ 11 から射出された各種イオンを m/z に応じて分離して検出する飛行時間型質量分析器 (TOFMS) 12 と、を備える。

30

【 0 0 2 7 】

生体試料を対象としたイオン化を行うイオン化部 10 としてはマトリックス支援レーザー脱離イオン化法 (MALDI) によるものが一般的であるが、イオン化法はこれに限るものではない。例えば、イオン化部 10 として、 $1 \mu m$ / 分未満の微流速の高速液体クロマトグラフ (HPLC) とこれに対応したナノESI (エレクトロスプレーイオン化) との組み合わせが有用である (非特許文献 2 参照)。

40

【 0 0 2 8 】

TOFMS 12 は、リフレクトロン電極により発生する電場によりイオンを折返し飛行させる飛行空間 13 と、該飛行空間 13 を飛行する間に m/z に応じて時間的に分離されたイオンを順次検出する検出器 14 と、を含む。

【 0 0 2 9 】

制御・処理部 2 は、質量分析部 1 の各部を制御する分析制御部 20、検出器 14 から得られる検出信号に基づいてマススペクトル、MS / MS スペクトルなどを作成する MS n

50

データ処理部 21 と、マスペクトル、MS/MS スペクトル中に現れているピークを抽出して各ピークの m/z や強度などの情報を含むピークリストを作成するピークリスト作成部 22 と、アミノ酸配列の非確定的修飾・変異の情報を格納した修飾・変異データベース (DB) 24 と、ペプチドのアミノ酸配列を推定するための同定用データベース (DB) 25 と、これらデータベースを用いてピークリストにヒットするタンパク質・ペプチドの検索を行う配列同定処理部 23 と、を機能ブロックとして含む。修飾・変異データベース 24 には、その時点で既知である全ての修飾・変異の情報、即ち、修飾・変異の名称、変化するアミノ酸の種類やペプチド部位、変化質量 (m/z) など、を予め登録しておくものとする。

【0030】

10

制御・処理部 2 の実体はコンピュータであって、該コンピュータにインストールされた専用の制御・処理ソフトウェアが動作することにより、前述の各種機能が達成される。配列同定処理部 23 に接続される検索条件入力部 28 は、検索条件を入力設定するためのものであり、具体的には、コンピュータに接続されるキーボードや、マウス等のポインティングデバイスである。また、表示部 29 は検索条件入力部 28 による検索条件入力設定画面を表示したり、同定結果を表示するためのものである。

【0031】

次に、本実施例のタンパク質同定システムを用いたタンパク質同定作業について、図 2 により説明する。なお、以下の説明ではマスコットの MS/MS イオンサーチ法と同じデータベース検索エンジンを用いた場合を例に挙げているが、データベース検索法はこれに限るものではなく、周知の他の方法を用いてもよいことは容易に理解できる。

20

【0032】

ユーザは目的とするタンパク質を適宜の酵素により消化してペプチド混合物を含む被検試料を調製する。この被検試料を質量分析部 1 により質量分析し、さらにそれにより得られるマスペクトル上に現れるピークをプリカーサとして選択して MS/MS 分析を実行する (ステップ S1)。プリカーサの選択方法は本発明には直接関係ないので詳細は省略するが、例えばマスペクトルに現れるピークに対し、同位体ピークの判定、多価イオンピークの判定などを行って同一成分由来のピークの重なりを除いた後に、ピーク強度を判定する等によりプリカーサとすべきイオンを決定する。

【0033】

30

次にピークリスト作成部 22 は、得られたマスペクトルや MS/MS スペクトルに現れるピークの情報収集してピークリストを作成する (ステップ S2)。ピークリストに掲載されるピーク情報は、ピークの m/z、強度を含む。

【0034】

次に、データベース検索を実行する前に、ユーザは検索条件入力部 28 より、非確定的修飾・変異 (Variable modification and substitution) を除く他の検索条件項目を設定する (ステップ S3)。即ち、図 4 に示したような検索条件設定画面を表示部 29 の画面上に表示し、データベース種類 (Database)、生物学的分類 (Taxonomy)、酵素の種類 (Enzyme)、誤開裂の許容数 (missed cleavage)、確定的修飾・変異 (Fixed modification and substitution、図 4 では「Fixed modifications」と記載の項目)、ペプチドの許容 m/z 範囲 (Peptide tol.)、MS/MS 分析時の許容 m/z 範囲 (MS/MS tol.)、ペプチド価数 (Peptide charge)、測定装置種類 (Instrument) などの各項目についてプルダウンメニューからの選択、数値や文字の直接的なキー入力、或いはチェックボックスのチェック記入などを行う。これらの検索条件は主として測定対象の試料の種類や測定装置 (質量分析部 1) の性能などに依存するものであり、ユーザにとっては既知であるか、或いは容易に予想できるものである。

40

【0035】

次にユーザはステップ S3 で設定しなかった (設定できなかった) 非確定的修飾・変異 (Variable modification and substitution、図 4 では「Variable modifications」と記載の項目) を検索条件入力部 28 により指定する (ステップ S4)。この際に、配列同定

50

処理部 23 は修飾・変異データベース 24 から読み出した登録情報のリストを表示部 29 の画面上に表示させ、そのリストに基づいてユーザが選択を行うようにすることができる。ここでは、指定可能な非確定的修飾・変異の数には制限がないため、最大、データベース 24 に登録されている全ての非確定的修飾・変異を検索条件として指定することができる。その場合に、ユーザは検索条件入力部 28 でリスト上の非確定的修飾・変異を全て選択するという指示を行えばよい。

【0036】

また、それ以外の場合に、リストに掲載されている非確定的修飾・変異を 1 つずつ選択して検索条件として指定することも可能であるが、そうした作業は面倒である。上述のように修飾・変異の総数は 600 程度であるが、実際には非常に低い確率でしか起こらない修飾・変異も数多く含まれる。そこで、起こる可能性の高さに基づく優先度や重要度などを各修飾・変異についてデータベース 24 に登録しておき、例えばこの優先度の高い順に所定個数の修飾・変異を選択して検索条件として指定する、といった簡便な指定方法を採用するとよい。

【0037】

上記のように検索に必要な検索条件を全て指定した上で、ユーザが検索の開始を指示すると（ステップ S5）、配列同定処理部 23 はまずステップ S4 で指定された非確定的修飾・変異のうちの 1 つを選択し、ステップ S3 で設定されたそれ以外の検索条件と組み合わせて検索条件とする（ステップ S6）。多数の非確定的修飾・変異の中から 1 つずつ修飾・変異を選択する際には、例えば上記のように各修飾・変異に与えられている優先度の順にするとよい。そして、この検索条件の下で、ピークリスト作成部 22 により作成されたピークリストに含まれるピーク情報に適合するアミノ酸配列を、同定用データベース 25 と照合することにより探索する（ステップ S7）。

【0038】

上記データベース検索により何らかのアミノ酸配列がヒットしたならば、その検索結果情報、例えばヒットしたペプチドのアミノ酸配列、タンパク質名、スコア（信頼性の指標値）などを図示しない記憶部に保存する（ステップ S8）。続いて、ステップ S4 で指定された全ての非確定的修飾・変異を検索条件とした検索を実行したか否かを判定し（ステップ S9）、未検索のものがあればステップ S6 へ戻る。そして、未だ選択されていない別の非確定的修飾・変異を 1 つ選択した上でステップ S7 以降へ進む。ステップ S6～S9 の繰り返しにより、ステップ S4 で指定された全ての非確定的修飾・変異を検索条件としたデータベース検索をそれぞれ実行し、各検索によりヒットしたものがある場合にはその結果が記憶部に保存される。したがって、例えば非確定的修飾・変異として 200 個（種類）が指定された場合、200 回のデータベース検索を繰り返すことになる。

【0039】

全ての非確定的修飾・変異を検索条件としたデータベース検索が終了すると、配列同定処理部 23 は、全ての検索結果を記憶部から読み出し、スコアなどを参考にした結果の妥当性の評価を行う（ステップ S10）。例えば一定値以上のスコアを有する結果、或いは、スコアの高い順に所定数の結果を抽出する。そして、抽出した結果、つまり妥当性の評価結果を表示部 29 の画面上に表示する（ステップ S11）。ユーザはこれを見て最終的に、例えば最もスコアの高い結果がペプチド・タンパク質の同定結果であると判断する。

【0040】

以上のように、本実施例のタンパク質同定システムでは、非確定的修飾・変異に関する検索条件の指定数の制約をなくす代わりに、起こり得る非確定的修飾・変異を 1 つに限定している。つまり、実際には起こる可能性がきわめて低い、2 つ以上の非確定的修飾・変異が同時に起こる場合については検索上で考慮しないようにしている。その結果、全てのペプチド・タンパク質に対し非確定的修飾・変異が生じたものを網羅的に検索することができるので、従来の MS / MS イオンサーチ法では同定できなかった、ユーザが予想しにくい修飾や変異が生じたタンパク質も同定できるようになる。また、非確定的修飾・変異に関する検索条件の指定数の制約はなくなったものの、2 つ以上の非確定的修飾・変異の

同時生起を考慮しないため、必要な検索回数の増大を抑え、ユーザが十分に許容できる時間内に検索時間を抑えることができる。さらに、2つ以上の非確定的修飾・変異の同時生起を考慮しないことで、検索における擬陽性の増加も抑えることができ、検索結果の信頼性を上げることができる。

【0041】

上述のように、修飾・変異データベース24に、非確定的な修飾や変異に対する発生頻度や重要度などの情報を登録しておくことで、ステップS4においてユーザは、発生頻度が高い又は優先度や重要度が高い修飾・変異を選択して指定することができる。このとき、発生頻度や優先度の閾値をユーザが設定できるようにし、その閾値以上の発生頻度や優先度が与えられた修飾・変異を非確定的修飾・変異の検索条件として自動的に選択するようにすると便利である。また、検索結果として妥当性の高い、つまり高いスコアを示す非確定的修飾・変異が見い出された場合には、非確定的修飾・変異の発生頻度を自動的に加算したり優先度を上げたりする、いわば学習機能を付加するようにしてもよい。これによって、分析するサンプルの種類、サンプル調製方法、測定方法などに伴い発生するユーザに特有の非確定的修飾・変異に関する情報を加えることができるので、より確度の高い検索が可能となる。

【0042】

なお、上記実施例では、指定された全ての非確定的修飾・変異を検索条件としたデータベース検索を実行していたが、検索の途中で十分に信頼に足る妥当な結果が得られた場合には、その時点でデータベース検索を打ち切るようにしてもよい。このように手順を変更した場合のフローチャートを図3に示す。図2と比較すれば明らかなように、ステップS8の後に、スコアが閾値以上であるか否かを判定し(ステップS12)、スコアが閾値未満であればステップS9へと進み、スコアが閾値以上であればステップS9をパスしてS10へと進むようにしている。後者の場合、それ以降の検索の繰り返しは実行されず、それまでに得られた検索結果のみを用いて妥当性を評価し、その評価結果を表示部29に出力する。これにより、不必要な検索を実行することがなくなり、タンパク質の同定に要する時間を短縮することができる。

【0043】

また、ステップS12でスコアが閾値以上であった場合でも、同じ優先度や発生頻度を示す非確定的修飾・変異が検索条件として残っている(未検索である)場合には、その検索条件についてのデータベース検索を実行した上でステップS10へと進むようにするとよい。これにより、スコアの高い結果を複数得ることも期待できる。

【0044】

本願発明者による検討によれば、通常のMS/MSイオンサーチ法やエラートレラントサーチ法では同定できなかった(ハイスコアの候補ペプチドがヒットしなかった)16個のデータの中で、本発明に係る上記のようなデータ処理方法を用いると12個のデータで正しい修飾ペプチドを同定することができた。次に、この同定の一例として、ウシ血清アルブミン(BSA=bovine serum albumin)のペプチドの同定例について説明する。

【0045】

BSAをトリプリン酵素消化して得たペプチドを含む試料を質量分析して得られるマススペクトルに現れるm/z1910のピークをプリカーサとしたMS/MS分析において、可能性が高いと予想される9種類のバリアブル・モディフィケーション(Acetyl (K)、Carbamidomethyl (C)、Carboxymethyl (C)、Dioxidation (M)、Methyl (DE)、Oxidation (M)、Phospho (ST)、Phospho (Y)、Propionamide (C))を指定してMS/MSイオンサーチを行うと、順位1位の候補は、スコア27でタンパク質「Oryza sativaのHypothetical protein」のペプチド「K.YGVASISTVDATCSTGER.E+Methyl (DE); Phospho (ST)」であった。この検索では、スコア39以上でホモロジーレベル、スコア64以上でアイデンティファイレベルであるので、上記順位1位の候補の信頼度は低い。なお、本検索結果の順位20位までの候補の中に、正しいタンパク質(BSA)はヒットしていなかった。

【0046】

これに対し、同じMS/MSデータに対して本発明に係る方法で、350種類のバリエーション・モディフィケーションを指定して検索を実行すると、順位1位の候補としてスコア62でタンパク質BSAのペプチド「K.CCAADDKEACFAVEGPK.L + Ammonia-loss (N-term C)」がヒットした。この検索では、スコア33以上でホモロジーレベル、スコア46以上でアイデンティファイレベルであるので、この候補の信頼度は高い。即ち、従来の方法では同定できなかったペプチドが高い確度で同定することができた。

【0047】

なお、上記実施例は本発明の一例にすぎず、本発明の趣旨の範囲で適宜変形、修正、追加等を行っても本願特許請求の範囲に包含されることは当然である。

【図面の簡単な説明】

10

【0048】

【図1】本発明の一実施例のタンパク質同定システムの全体構成図。

【図2】図1のタンパク質同定システムを用いたタンパク質同定手順を示すフローチャート。

【図3】変形したタンパク質同定手順を示すフローチャート。

【図4】コンピュータのモニタ画面上に表示されるマスコットによるMS/MSイオンサーチの検索設定画面の一例を示す図。

【符号の説明】

【0049】

1 ... 質量分析部

20

10 ... イオン化部

11 ... イオントラップ

12 ... TOFMS

13 ... 飛行空間

14 ... 検出器

2 ... 制御・処理部

20 ... 分析制御部

21 ... MS_nデータ処理部

22 ... ピークリスト作成部

23 ... 配列同定処理部

30

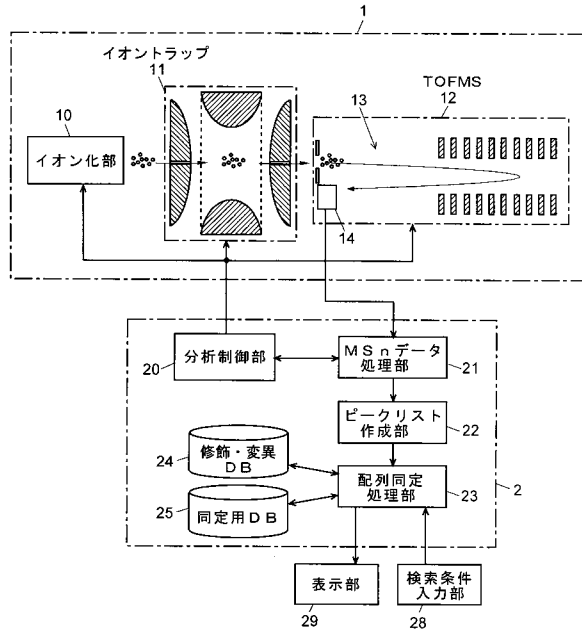
24 ... 修飾・変異データベース

25 ... 同定用データベース

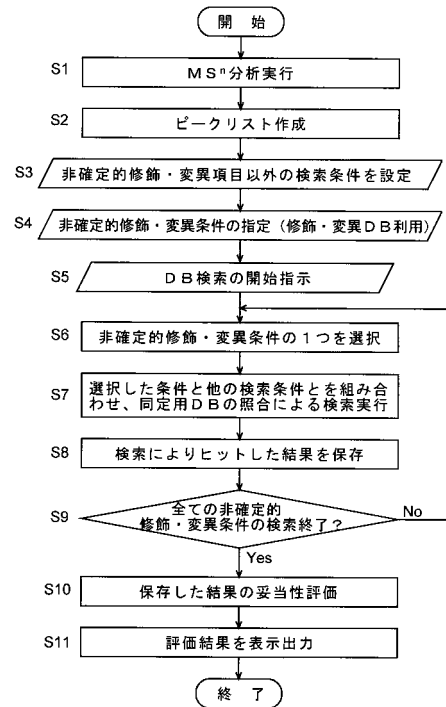
28 ... 検索条件入力部

29 ... 表示部

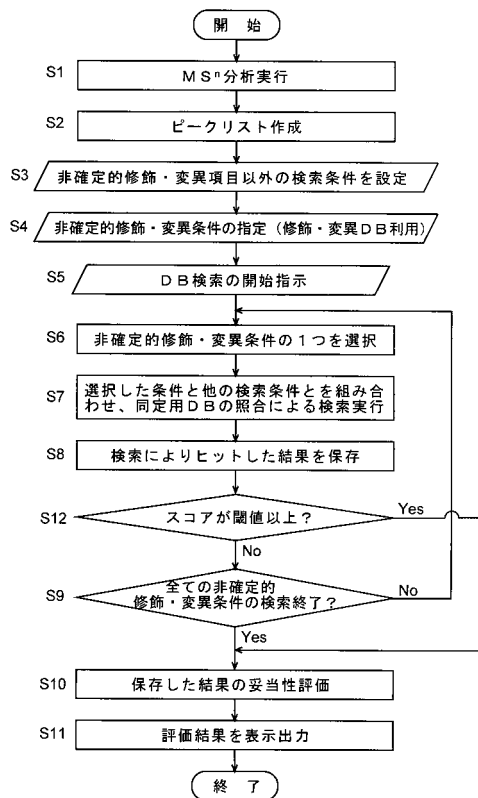
【図 1】



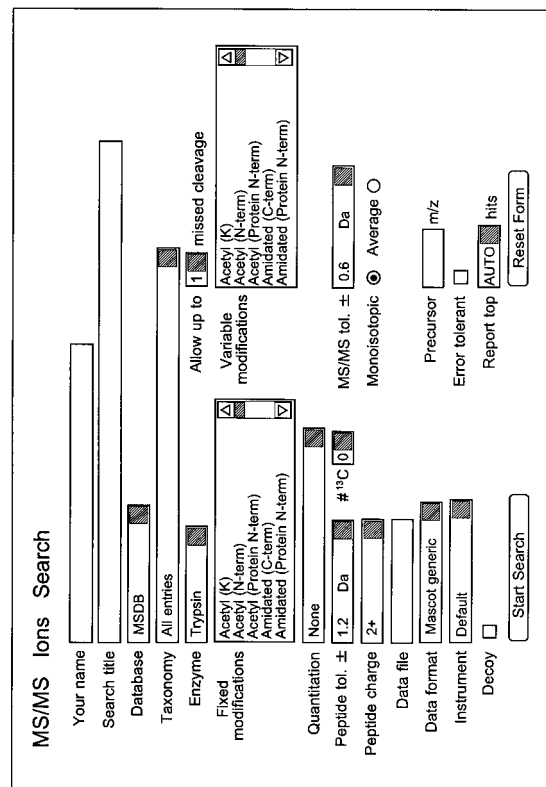
【図 2】



【図 3】



【図 4】



フロントページの続き

- (72)発明者 網澤 進
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 島 圭介
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 松尾 英一
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 山崎 雄三
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 飯田 哲生
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 鈴木 崇
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 藤本 宏隆
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
- (72)発明者 嶋田 崇史
京都市中京区西ノ京桑原町1番地 株式会社島津製作所内

審査官 大竹 秀紀

- (56)参考文献 特表2005-500990(JP,A)
特開2008-039608(JP,A)
特開2007-309695(JP,A)
特表2007-531874(JP,A)
特表2005-519284(JP,A)
特開平07-242700(JP,A)
国際公開第2007/137867(WO,A1)
国際公開第2006/004214(WO,A1)
国際公開第2005/100973(WO,A1)

- (58)調査した分野(Int.Cl., DB名)
G01N 27/62