

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
26. April 2018 (26.04.2018)



(10) Internationale Veröffentlichungsnummer
WO 2018/072862 A1

- (51) Internationale Patentklassifikation:
G06F 19/10 (2011.01) *H01J 49/00* (2006.01)
G01N 33/68 (2006.01)
- (21) Internationales Aktenzeichen: PCT/EP2017/001131
- (22) Internationales Anmeldedatum:
22. September 2017 (22.09.2017)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:
10 2016 012 302.8
17. Oktober 2016 (17.10.2016) DE
- (71) Anmelder: UNIVERSITÄT BREMEN (BCCMS)
[DE/DE]; Bibliothekstraße 1, 28359 Bremen (DE).
- (72) Erfinder: BOSKAMP, Tobias; Ostendorfer Strasse 39, 27726 Worpswede (DE).
- (74) Anwalt: MÖLLER, Friedrich et al.; Meissner Bolte Partnerschaft mbB, Hollerallee 73, 28209 Bremen (DE).
- (81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: METHOD FOR EVALUATING DATA FROM MASS SPECTROMETRY, MASS SPECTROMETRY METHOD, AND MALDI-TOF MASS SPECTROMETER

(54) Bezeichnung: VERFAHREN ZUM AUSWERTEN VON DATEN EINER MASSENSPEKTROMETRIE UND MASSENSPEKTROMETRISCHES VERFAHREN SOWIE EIN MALDI TOF MASSENSPEKTROMETER

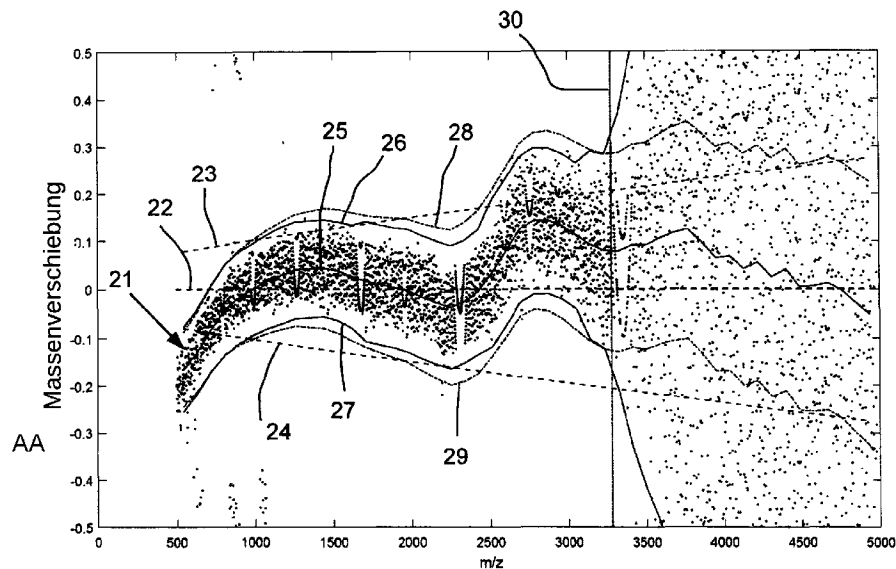
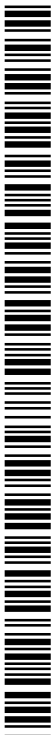


Fig. 3

AA Mass shift

(57) Abstract: The invention relates to a method for evaluating data from mass spectrometry for the analysis of peptides from biological samples, in particular from MALDI-TOF mass spectrometry, having the following steps: a) providing expected mass defects; b) determining measured mass defects, namely the mass defects resulting from the data from the mass spectrometry; c) comparing the measured mass defects with the expected mass defects.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF-Massenspektrometrie, mit folgenden Schritten: a) Bereitstellung



WO 2018/072862 A1

(84) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz 3)

Verfahren zum Auswerten von Daten einer Massenspektrometrie und massenspektrometrisches Verfahren sowie ein MALDI TOF Massenspektrometer

Beschreibung

Die Erfindung betrifft ein Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF-Massenspektrometrie gemäß der Ansprüche 1 und 8. Daneben betrifft die Erfindung ein massenspektrometrisches Verfahren zur Analyse von Peptiden aus biologischen Proben, insbesondere unter Verwendung eines MALDI-TOF-Massenspektrometers nach Anspruch 12. Schließlich ist die Erfindung gerichtet auf ein massenspektrometrisches Verfahren zur Analyse von Peptiden aus biologischen Proben gemäß den Ansprüchen 13 und 17 sowie ein MALDI-TOF-Massenspektrometer nach Anspruch 19.

Bei der sogenannten Matrix Assisted Laser Desorption/Ionization Time-of-flight Massenspektrometrie (MALDI-TOF-MS) wird eine biologische Gewebeprobe nach einer geeigneten Probenpräparation mit einer Matrixlösung überzogen und in Vakuum mit einem Laser beschossen. Dabei werden biologische Makromoleküle aus dem Gewebe herausgelöst und ionisiert, typischerweise mit einer einfach positiven Ladung. Die Ionen werden anschließend in einem elektrischen Feld beschleunigt und von einem Detektor registriert. Aus der Flugzeit lässt sich der m/z -Wert, d.h. das Verhältnis zwischen Masse und Ladung des Moleküls bestimmen. Das gemessene Massenspektrum stellt die relative Zahl der registrierten Ionen (spektrale Intensität) als Funktion ihrer m/z -Werte dar. Unter der Annahme einer einfach positiven Ionisierung ist der m/z -Wert äquivalent zur Masse m des ionisierten Moleküls. Nachfolgend wird zur Vereinfachung unter der Masse m des ionisierten Moleküls der m/z -Wert verstanden.

Der m/z-Wert bzw. die molekulare Masse wird in Dalton (Da) angegeben als ein Vielfaches der atomaren Masseneinheit (1 Da = 1 amu, atomic mass unit). Näherungsweise entspricht die Masse eines Moleküls in Da der Gesamtzahl der Protonen und Neutronen, aus denen die Atomkerne des Moleküls
5 zusammengesetzt sind. Die Differenz zwischen dieser ganzzahligen Nominalmasse und der tatsächlichen Masse wird als Massendefekt bezeichnet. Der Massendefekt eines Moleküls ist die Summe der Massendefekte der einzelnen Atome, die wiederum für jedes chemische Element bzw. Isotop verschieden sind.

10

Der Ausdruck "Massendefekt" wird in der Literatur nicht einheitlich verwendet. Eine erste Bedeutung des Begriffs bezieht sich auf die Differenz der Massen in der SI-Einheit kg. Eine zweite Bedeutung bezieht sich ebenfalls auf die Differenz der Massen, geht aber von der atomaren Masseneinheit u aus, die mit Bezug
15 zum Kohlenstoff-Isotop ^{12}C definiert ist. Aufgrund dieser Festlegung ist der Massendefekt des Kohlenstoff-Isotops ^{12}C gleich Null. Für diese zweite Bedeutung wird anstelle des Begriffs "Massendefekt" auch der Begriff "Massenüberschuss" verwendet, um den Unterschied zur oben genannten ersten Bedeutung herauszustellen. In der Biologie und Chemie, insbesondere im
20 Zusammenhang mit massenspektrometrischen Verfahren, wird aber trotzdem der Begriff "Massendefekt" verwendet, also im Sinne der zweiten Bedeutung, so auch hier.

Bei der Akquisition von MALDI-TOF-Massenspektrometriedaten von biologischen
25 Gewebeschnitten wird eine Vielzahl von Informationen über die proteomische Struktur der Gewebeproben gewonnen. Gleichzeitig unterliegt die Messung einer Reihe von potentiellen Störungen, die zu Verzerrungen und Verfälschungen der gewonnenen Informationen führen können. Aufgrund der hohen Komplexität der Daten ist eine objektive Beurteilung ihrer Qualität und Genauigkeit oftmals nicht
30 möglich. Aktuell existieren keine breit akzeptierten und leicht anwendbaren Maßstäbe, die eine Aussage darüber erlauben, welche Datenqualität eine Messung aufweist oder ob zwei Messungen Daten von vergleichbarer Qualität liefern.

Aus der US 2016/0003842 A1 ist ein massenspektrometrisches Verfahren zur Analyse von Peptiden bekannt. Ziel ist dabei die Erkennung sogenannter Glykopeptide. In den Fig. 2a, 2b der US 2016/0003842 A1 sind Massendefekte über nominalen Massen m/z aufgetragen. Unterschieden werden Bereiche mit
5 Peptiden einerseits und mit Glykopeptiden angereicherte Bereiche andererseits.

Eine häufig auftretende Verzerrung der gemessenen Daten besteht in einem systematischen Fehler in den gemessenen Massen, der auch unter sorgfältig kontrollierten experimentellen Bedingungen die messtechnisch bedingte
10 Toleranzgrenze überschreitet. Herkömmliche Methoden zur Korrektur solcher Massenverzerrungen sind in vielen Fällen entweder zu ungenau oder zu zeitaufwändig.

Aufgabe der vorliegenden Erfindung ist die Schaffung eines Verfahrens zur
15 Qualitätskontrolle von Massenspektrometriedaten (bei der Analyse von Peptiden aus biologischen Proben) bzw. ein massenspektrometrisches Verfahren, mit einer entsprechenden Kontrolle, bzw. ein massenspektrometrisches Verfahren mit einer Signalkorrektur.

20 Zur Lösung der Aufgabe weist das Verfahren die Merkmale des Anspruchs 1 auf. Vorgesehen ist demnach ein Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF- Massenspektrometrie, mit folgenden Schritten:

- 25 a) Bereitstellung erwarteter Massendefekte;
b) Bestimmung gemessener Massendefekte, nämlich der sich aus den Daten der Massenspektrometrie ergebenden Massendefekte;
c) Vergleich der gemessenen Massendefekte mit den erwarteten Massendefekten.

30

Je nach Größe der Abweichung der gemessenen Massendefekte von den erwarteten Massendefekten können die Daten bzw. eine zugrunde liegende Messung als fehlerhaft oder akzeptabel bewertet werden. Auch kann durch eine geeignete Signalquelle ein zur Bewertung der Daten korrespondierendes Signal
35 ausgegeben werden, etwa eine Anzeige auf einem Bildschirm.

5 Sofern die Daten bzw. die Messung als akzeptabel bewertet werden, wird mit den Daten weiter gearbeitet und/oder es werden weitere Messungen durchgeführt. Bei als fehlerhaft bewerteten Daten können diese beispielsweise für die weitere Verarbeitung verworfen werden und/oder die für die Durchführung der Massenspektrometrie verwendete Vorrichtung wird einer Überprüfung unterzogen.

10 Das Verfahren basiert unter anderem auf einer rechnergestützten Visualisierung der Massendefekte der in einem Spektrum aufgefundenen Peaks. Dabei wird ausgenutzt, dass eine Vielzahl der Peaks von Peptiden herrühren, deren Massendefekte einem charakteristischen Muster folgen. Durch Vergleiche der gemessenen Massendefekte mit einem theoretisch erwarteten Muster der Massendefekte / den erwarteten Massendefekten lassen sich rechnerisch und
15 visuell Rückschlüsse auf die Qualität der gemessenen Daten ziehen.

Die Massendefekte werden im massenspektrometrischen Verfahren natürlich nicht direkt gemessen, sondern aus den ermittelten Massen (welche in einem TOF-Massenspektrometer aus gemessenen Flugzeiten bestimmt werden) berechnet. Zur Vereinfachung wird aber der Begriff "gemessener Massendefekt" zur Abgrenzung gegenüber dem "erwarteten Massendefekt" verwendet. Letzterer ergibt sich anhand von Berechnungen aufgrund der besonderen Eigenschaften der Peptide.

25 In Fortbildung der Erfindung ist vorgesehen, dass die erwarteten Massendefekte berechnet werden aus

$$m_N r_p,$$

wobei m_N die Nominalmasse eines Peptids bezeichnet
30 und r_p vorzugsweise zwischen 10^{-3} und 10^{-4} liegt, insbesondere etwa $4,95 \times 10^{-4}$ beträgt.

Für die Berechnung der erwarteten Massendefekte werden somit für eine gegebene (ganzzahlige) Nominalmasse m_N die Differenz zu der sich aus dem

Produkt mit $1+r_p$ ergebenden Masse m berücksichtigt. r_p ist vorzugsweise der Faktor $4,95 \times 10^{-4}$. Auch davon abweichende Werte sind grundsätzlich möglich.

Nach einem weiteren Gedanken der Erfindung ist vorgesehen, dass der
5 Massendefekt zu einer gemessenen Masse m berechnet wird aus

$$m - \text{floor}\left(\frac{m}{1+r_p} + 0,5\right),$$

wobei die Funktion $\text{floor}(x)$ für ein beliebiges $x > 0$ den ganzzahligen Anteil von x
10 bezeichnet.

Der gemessene Massendefekt wird also als Differenz zwischen m und derjenigen Nominalmasse m_N bestimmt, deren zugehörige erwartete Peptidmasse $m_p = (1+r_p)m_N$ am dichtesten an m liegt.

15 In Fortbildung der Erfindung ist vorgesehen, dass die Diskrepanz δ_p zwischen dem gemessenen und dem erwarteten Massendefekt direkt aus der gemessenen Masse m berechnet wird als

$$\delta_p(m) = \varphi\left(\frac{m}{1+r_p} + 0,5\right) - 0,5$$

20

wobei die Funktion $\varphi(x) = x - \text{floor}(x)$ den Nachkommaanteil von x für ein beliebiges $x > 0$ bezeichnet.

Nach einem weiteren Gedanken der Erfindung ist vorgesehen, dass zum
25 Vergleich der gemessenen Massendefekte mit den erwarteten Massendefekten über Teilintervalle einer Massenachse jeweils der Median der gemessenen Massendefekte gebildet und mit dem erwarteten Massendefekt verglichen wird.

In Fortbildung der Erfindung werden bei weiterer Verwendung der Daten die
30 Messwerte (der ermittelten Massen) korrigiert, nämlich in Abhängigkeit von der Abweichung der gemessenen Massendefekte von den erwarteten Massendefekten. Im einfachsten Fall werden die ermittelten Massen um die Differenzen der Massendefekte korrigiert.

Nach einem weiteren Gedanken der Erfindung werden die gemessenen Massendefekte für lokale Maxima der spektralen Intensitäten berechnet. Jedes lokale Maximum wird als Peak einer bestimmten gemessenen Masse angesehen.

5

Ein weiteres erfindungsgemäßes Verfahren weist die Merkmale des Anspruchs 8 auf. Vorgesehen ist demnach ein Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF- Massenspektrometrie, vorzugsweise nach
10 einem der zuvor erörterten Verfahren, mit folgenden Schritten:

- a) Bereitstellung erwarteter Massendefekte;
- b) Bestimmung gemessener Massendefekte, nämlich der sich aus den Daten der Massenspektrometrie ergebenden Massendefekte;
- 15 c) Bestimmung der Diskrepanzen zwischen den gemessenen Massendefekten und den erwarteten Massendefekten.
- d) Bestimmung der Streuung der Diskrepanzen um deren Mittelwert;
- e) Vergleich der Streuung mit einer definierten zulässigen Streuung.

20 Je nach Abweichung der Streuung von der definierten zulässigen Streuung können die Daten bzw. eine zugrunde liegende Messung als fehlerhaft oder akzeptabel bewertet werden. Auch kann durch eine geeignete Signalquelle ein zur Bewertung der Daten korrespondierendes Signal ausgegeben werden, etwa eine Anzeige auf einem Bildschirm.

25

Sofern die Daten bzw. die Messung als akzeptabel bewertet werden, wird mit den Daten weiter gearbeitet und/oder es werden weitere Messungen durchgeführt. Bei als fehlerhaft bewerteten Daten können diese beispielsweise für die weitere Verarbeitung verworfen werden und/oder die für die Durchführung der
30 Massenspektrometrie verwendete Vorrichtung wird einer Überprüfung unterzogen.

In Fortbildung der Erfindung ist vorgesehen, dass zur Bestimmung einer Streuung der Massendefektdiskrepanzen Interquartilsabstände der ermittelten
35 Diskrepanzen über Teilintervalle der Massenachse bestimmt werden, und dass

bei Überschreitung eines Grenzwertes der Streuung insbesondere die Daten als fehlerhaft verworfen werden.

Nach einem weiteren Gedanken der Erfindung ist ein Korridor für zulässige
5 Streuungen gebildet durch die Grenzwerte

$$d_p^{1,2}(m) = \pm \mu \sqrt{v(m)}$$

wobei

$$v(m) = \sigma_p^2(m) + \frac{\Delta m(m)^2}{12},$$

10

und

$$\sigma_p(m_N) = \sigma_0 + s_p m_N, \text{ mit } \sigma_0 \approx 0,02 \text{ und } s_p \approx 2,0 \times 10^{-5},$$

und $\Delta m(m)$ die Breite von m/z-Bins an der Massenposition m bezeichnet,

15

und m/z-Bins die sich durch Diskretisierung der Massenachse ergebenden Intervalle repräsentieren,

und $\mu > 0$ einen Skalierungsfaktor angibt, vorzugsweise $\mu = 2$,

20

und wobei insbesondere die Daten als fehlerhaft verworfen werden, wenn die Streuung der Massendefektdiskrepanzen außerhalb des derart vorgegebenen Korridors liegt.

25 In Fortbildung der Erfindung ist vorgesehen, dass der Massebereich mit erkennbarem Peptidsignal bestimmt wird als die Gesamtheit aller Teilintervalle der Massenachse, für die der Quotient aus der tatsächlichen Streuung und der maximal zulässigen Streuung $d_p(m)$ einen festgelegten Schwellwert t nicht überschreitet, wobei vorzugsweise $t = 1,2$ festgelegt wird. Die untere bzw. obere
30 Grenze dieses Massebereiches gibt die Ausdehnung des Peptidsignalbereiches an.

Je nach Unterschreitung eines Toleranzwertes für die obere Grenze oder die Ausdehnung des Peptidsignalbereiches, bzw. Überschreitung eines
35 Toleranzwertes für die untere Grenze des Peptidsignalbereiches können die

Daten bzw. eine zugrunde liegende Messung als fehlerhaft oder akzeptabel bewertet werden. Auch kann durch eine geeignete Signalquelle ein zur Bewertung der Daten korrespondierendes Signal ausgegeben werden, etwa eine Anzeige auf einem Bildschirm.

5

Sofern die Daten bzw. die Messung als akzeptabel bewertet werden, wird mit den Daten weiter gearbeitet und/oder es werden weitere Messungen durchgeführt. Bei als fehlerhaft bewerteten Daten können diese beispielsweise für die weitere Verarbeitung verworfen werden und/oder die für die Durchführung der Massenspektrometrie verwendete Vorrichtung wird einer Überprüfung unterzogen.

Das erfindungsgemäße massenspektrometrische Verfahren weist die Merkmale des Anspruchs 12 auf. Zur Analyse von Peptiden aus biologischen Proben, insbesondere unter Verwendung eines MALDI-TOF-Massenspektrometers, sind folgende Schritte vorgesehen:

- a) Durchführung eines oder mehrerer massenspektrometrischer Untersuchungen an der biologischen Probe und Bereitstellung von sich aus den massenspektrometrischen Untersuchungen ergebenden Daten;
- 20 b) Durchführung eines der Verfahren zum Auswerten von Daten, wie zuvor dargestellt.

Eine weitere Lösung der eingangs genannten Aufgabe weist die Maßnahmen des Anspruchs 13 auf. Demnach ist ein Verfahren vorgesehen mit folgenden Schritten:

- a) Messung von m/z -Werten für verschiedene Peptide mittels einer Messapparatur, insbesondere unter Verwendung eines MALDI-TOF-Massenspektrometers,
- b) Zuordnung der gemessenen m/z -Werte zu entsprechenden m/z -Bins eines 2D-Histogramms,
- 30 c) Auftrag von spektralen Intensitäten der m/z -Bins in dem 2D-Histogramm, wobei auf einer Abszissenachse die m/z -Bins und auf einer Ordinatenachse die Diskrepanz zwischen gemessenem und erwartetem Massendefekt aufgetragen werden,

- d) wobei die durch die beiden Achsen aufgespannte Diagrammfläche in eine Vielzahl, vorzugsweise 20 bis 50, von Rechtecken aufgeteilt wird,
- e) wobei die gemessenen Werte auf eine m/z-Auflösung, die der gewählten Unterteilung der Ordinatenachse entspricht, hochinterpoliert werden, und
- 5 f) wobei für jedes Rechteck diejenigen Intensitätswerte des interpolierten Spektrums aufsummiert werden, deren Massendefektdiskrepanz in die jeweiligen Teilintervalle der Achsen fallen,
- g) wobei unterschiedliche Intensitätswerte unterschiedlich gekennzeichnet werden und gleiche Intensitätswerte gleich gekennzeichnet werden.

10

Bevorzugt kann es die Erfindung vorsehen, dass zu jedem Teilintervall der Abszissenachse, insbesondere der horizontalen Massenachse, die entsprechenden Intensitätswerte der Ordinatenachse statistisch ausgewertet werden, um Häufungspunkte und/oder Streuungswerte zu bestimmen.

15

Ein weiteres Ausführungsbeispiel kann es vorsehen, dass zur Beschreibung der Verteilung der aufsummierten Intensitätswerte in vertikaler Richtung zirkuläre Statistiken verwendet werden, insbesondere kann ein erstes zirkuläres Moment Z als (komplexwertige) Statistik herangezogen werden.

20

Insbesondere ist es denkbar, dass zur Bestimmung eines Massenshiftprofils (Vektor aller zirkulären Moment Z für alle Teilintervalle der Abszissenachse, insbesondere der horizontalen Massenachse) die Schritte zur Bildung des Histogramms und der Berechnung der zirkulären Momente zusammengesetzt

25 werden und gemäß der Formel

$$Z_k = \frac{1}{\int_{I_k} \tilde{S}(t) dt} \int_{I_k} \tilde{S}(t) e^{i\omega t} dt, \text{ mit } \omega = \frac{2\pi}{1 + r_p}$$

als Fourier-Integrale des kontinuierlich interpolierten Spektrums \tilde{S} über die
30 Teilintervalle I_k der Abszissenachse ausgedrückt werden.

Eine weitere Lösung der eingangs genannten Aufgabe weist die Maßnahmen des Anspruchs 17 auf. Demnach ist ein massenspektrometrisches Verfahren zur

Handhabung eines Einzelspektrums vorgesehen, bei dem für eine Massenshiftnormalisierung zu einem Ensemble für jedes Spektrum

- a) ein Massenshiftprofil ermittelt wird,
- b) aus allen einzelnen Massenshiftprofilen durch elementweise Bildung des arithmetischen Mittelwertes ein gemeinsames, mittleres Referenzprofil gebildet wird,
- c) jedes Spektrum modifiziert wird, so dass das Massenshiftprofil des modifizierten Spektrums dem Referenzprofil entspricht,

- 10 Darüber hinaus kann es vorgesehen sein, dass für die Normalisierung eines einzelnen Spektrums auf das Referenzprofil jeweils für die einzelnen Teilintervalle, für die die Massenshiftprofile berechnet werden, relative Verschiebungswerte bestimmt und über die gesamte Massenachse interpoliert werden, wobei die gemessenen Werte des Spektrums um diese interpolierten
- 15 Verschiebungswerte korrigiert werden.

Ein MALDI-TOF-Massenspektrometer zur Lösung der vorbezeichneten Aufgabe wird durch Anspruch 19 beschrieben. Demnach weist dieses Massenspektrometer eine Steuereinheit auf zur Durchführung des

20 erfindungsgemäßen Verfahren nach mindestens einem der Ansprüche 1 bis 18.

Weitere Merkmale der Erfindung ergeben sich aus der Beschreibung im Übrigen und aus den Ansprüchen. Vorteilhafte Ausführungsbeispiele der Erfindung werden nachfolgend anhand von Zeichnungen näher erläutert. Es zeigen:

25

Figur 1 einen Ausschnitt aus einem Mittelwertspektrum mit deutlich erkennbarer, wellenförmiger Grundlinie mit charakteristischer Wellenlänge von etwas mehr als 1 Da;

- 30 Figur 2 ein herkömmliches Massendefektdiagramm eines Mittelwertspektrums mit erkennbarem Peptidband;

Figur 3 ein Peptidmassendefektdiagramm (PMD) des Mittelwertspektrums zu Figur 2, nämlich mit Darstellung einer Massenverschiebung über der Masse (m/z), mit horizontaler Referenzlinie für einen erwarteten Massendefekt und

35

Referenzkorridor (gestrichelte Linien), Massendefektdiskrepanz und deren Streuung (durchgezogene Linien) sowie erwarteter Streuung (Strich-Punkt-Strich Linien) und oberer Grenze des Massebereichs mit erkennbarem Peptidsignal (gepunktete Linie);

5

Figur 4 ein PMD eines Mittelwertspektrums mit stark reduziertem Signal-Rauschverhältnis, erkennbar am verkürzten Peptidsignalebereich, der nur bis ca. 1300 Da reicht;

10 Figur 5 ein PMD eines Spektrums mit deutlicher Massenverschiebung um ca. 0,15 bis 0,35 Da;

Figur 6 ein PMD mit deutlich erkennbarer Linienstruktur aufgrund einer äquidistant gesampelten Massenachse;

15

Figur 7: ein PMD eines Mittelwertspektrums mit sehr grob gesampelter Massenachse; aufgrund des groben Samplings ist die Struktur des Peptidbandes kaum noch zu erkennen;

20 Figur 8 ein PMD eines Mittelwertspektrums mit fehlerhaft durchgeführter Kalibrierung, erkennbar am Bruch im Peptidband ab ca. $m/z=1700$;

Figur 9 ein Peptidmassendefekthistogramm (PMH) eines Mittelwertspektrums mit Massenverschiebung vergleichbar zu der in Figur 5 gezeigten;

25

Figur 10 Mittelwertspektren vor (oben) und nach (unten) Massenshiftnormalisierung.

Die in einer MALDI-Messung registrierten Moleküle beinhalten insbesondere
30 Metabolite und Peptide. Metabolite sind Stoffwechselprodukte und können unterschiedliche chemische Formen aufweisen, z.B. Lipide, Kohlehydrate oder Abbauprodukte von aus der Nahrung oder der Umwelt aufgenommenen Substanzen. Ihre Massen betragen typischerweise weniger als 1000 Da. Demgegenüber sind Peptide Verkettungen von Aminosäuren mit Massen von bis
35 zu 5000 Da und mehr.

Alle 23 in Proteinen vorkommende Aminosäuren – und damit alle Peptide – bestehen aus den fünf chemischen Elementen Kohlenstoff, Wasserstoff, Sauerstoff, Stickstoff und Schwefel. Das relative Verhältnis dieser Elemente zueinander ist bei allen Peptiden näherungsweise und unabhängig von ihrer Gesamtmasse gleich, so dass der Massendefekt eines Peptides im Wesentlichen durch dessen Nominalmasse bestimmt wird. Es ergibt sich ein nahezu linearer Zusammenhang zwischen der Masse m eines Peptids und seiner Nominalmasse m_N :

10

$$m \approx m_P(m_N) = (1+r_P) m_N, \text{ mit } r_P \approx 4,95 \times 10^{-4}.$$

Die Streuung der wahren Massen um den theoretischen Mittelwert m_P ist relativ klein, ihre Standardabweichung kann mit

15

$$\sigma_P(m_N) = \sigma_0 + s_P m_N, \text{ mit } \sigma_0 \approx 0,02, \text{ } s_P \approx 2,0 \times 10^{-5}$$

abgeschätzt werden.

Aufgrund der großen Zahl von unterschiedlichen Proteinen und daraus resultierenden Peptiden in biologischen Gewebezellen weist ein typisches MALDI-Spektrum Signalintensitäten an praktisch allen $m_P(m_N)$ für einen breiten Bereich von Nominalmassen m_N auf. Bildet man aus mehreren während der Vermessung einer Gewebeprobe gewonnen Spektren ein Summen- oder Mittelwertspektrum, siehe Figur 1, so ist darin eine charakteristische, gleichmäßige Wellenlinie als Grundlinie 20 zu erkennen, die insbesondere oberhalb von ca. 1000 Da deutlich hervortritt und eine Wellenlänge von näherungsweise $1+r_P$ Da aufweist.

Die Bestimmung der Masse eines Moleküls ist mit einem Fehler behaftet, der im Wesentlichen von zwei Ursachen herrührt: Zum einen kann die Flugzeit eines Moleküls nur mit einer gewissen Genauigkeit und in diskreten Intervallen gemessen werden, woraus sich eine Diskretisierung der Massenachse, also eine Einteilung in aufeinanderfolgende Intervalle (m/z -Bins) ergibt. Üblicherweise ist die Breite der m/z -Bins nicht konstant sondern nimmt zu höheren Massen hin zu.

35

Zum anderen hängt die Flugzeit des Moleküls nicht nur von seiner Masse ab sondern auch von dessen Anfangszustand innerhalb der Ionenwolke zu Beginn der Beschleunigung. Dieser Anfangszustand, insbesondere Geschwindigkeit und Bewegungsrichtung des Moleküls, sind weitestgehend unbekannt und führen zu
5 einem deutlichen Messfehler, der üblicherweise durch eine Kalibrierung nach der Messung korrigiert wird.

Zu den gängigen Kalibrierungsmethoden gehören die externe Kalibrierung und die statistische Peptidkalibrierung. Bei der externen Kalibrierung werden vor der
10 Messung mehrere Tropfen einer Lösung mit definierten Inhaltsstoffen neben der Gewebeprobe platziert. Die darin gemessenen Spektren werden nach der Messung mit den erwarteten Massen der bekannten Inhaltsstoffe verglichen und es wird eine Kalibrierungskurve für die m/z -Achse eines Spektrums bestimmt. Bei der Peptidkalibrierung wird der oben beschriebene Zusammenhang zwischen
15 wahrer Masse eines Peptids und dessen Nominalmasse ausgenutzt, um die mutmaßlich zu einem Peptid gehörenden Peakpositionen auf die theoretisch erwarteten m/z -Werte zu verschieben, siehe Wool A, Smilansky Z: Precalibration of matrix-assisted laser desorption/ionization-time of flight spectra for peptide mass fingerprinting. Proteomics 2002, 2, 1365–1373.

20 Beide Kalibrierungsmethoden können die Fehler in den m/z -Werten nicht vollständig korrigieren. Die externe Kalibrierung erfordert darüber hinaus eine manuelle Interaktion, während die Peptidkalibrierung sehr rechen- und zeitaufwändig ist.

25 Da bei dieser Methode der Ausgleich der Massenfehler global für alle Spektren einer Messung erfolgt, lassen sich unterschiedliche Fehler in den Spektren eines Datensatzes auf diese Weise nicht korrigieren. Als Alternative wird daher auch eine interne Kalibrierung angewandt, bei der die Kalibrierungslösung über die zu
30 messende Gewebeprobe verteilt wird und somit eine individuelle Korrektur für jedes Spektrum der Messung möglich ist.

Aus praktischen Gründen kann eine Kalibrierungslösung nur eine geringe Anzahl bekannter Substanzen beinhalten. Dies limitiert die Anzahl der Stützstellen, aus
35 denen die Kalibrierungskurve bestimmt wird, und damit die Genauigkeit der

Kalibrierung. Darüber hinaus erfordert diese Form der Kalibrierung eine manuelle Benutzerinteraktion.

Demgegenüber sind Methoden der statistischen Peptidkalibrierung (siehe Wool
5 A, Smilansky Z: Precalibration of matrix-assisted laser desorption/ionization-time
of flight spectra for peptide mass fingerprinting. *Proteomics* 2002, 2, 1365–1373;
Wolski WE, Lalowski M, Jungblut P, and Reinert K. Calibration of mass
spectrometric peptide mass fingerprint data without specific external or internal
calibrants. *BMC bioinformatics*, 6(1):203, 2005) vollautomatisch und kommen
10 ohne Kalibrierungslösung aus. Bei diesen Methoden erfolgt die Korrektur durch
Abgleich der im Gewebe gemessenen Massen mit einem theoretischen
Peptidmassenmodell (s.o.) sowie einer Peptidatenbank. Diese Methoden
erfordern ein vorheriges Peak Picking, also eine Identifikation relevanter Peaks in
einem Spektrum, sind sehr zeitaufwändig und können aufgrund falscher
15 Zuordnung zwischen Peak und Peptidatenbank zu fehlerhaften Ergebnissen
führen.

Zur Visualisierung der in einem Spektrum beobachteten Massendefekte werden
die m/z -Werte der in einem Spektrum gefundenen Peaks in einem Diagramm
20 aufgetragen, dessen horizontale Achse der Masse m (bzw. dem m/z -Wert)
entspricht, während auf der vertikalen Achse deren Nachkommaanteil $m\text{-floor}(m)$
abgetragen wird, siehe Figur 2. Solche Diagramme werden benutzt, um
unterschiedliche Zusammensetzungen komplexer Molekülmischungen sichtbar
zu machen. Eine Variante sind sog. Kendrick-Massendefektdiagramme, die zur
25 Charakterisierung von chemischen Verbindungen einer bestimmten Gruppe
verwendet werden, siehe Wikipedia: Kendrick mass.
https://en.wikipedia.org/wiki/Kendrick_mass. Im Zusammenhang mit
Untersuchungen an Peptiden können Massendefektdiagramme zur Abgrenzung
zwischen Peptiden und sog. Glykopeptiden herangezogen werden, siehe US
30 2016/0003842 A1 und Froehlich J et al.: A Classifier Based on Accurate Mass
Measurements to Aid Large Scale, Unbiased Glycoproteomics. *Mol. Cell.*
Proteomics 2013, 12, 1017–1025.

Darüber hinaus können Peptide gezielt chemisch so modifiziert werden, dass sie
35 einen vom Peptidmassenmodell (auch als Averagine-Modell bezeichnet)

signifikant abweichenden Massendefekt aufweisen und mittels dieser Abweichung von unmodifizierten Peptiden abgegrenzt werden können, siehe Chen X, Savickas P, Vestal M. Methods and systems for mass defect filtering of mass spectrometry data. US Patent 7,634,364, filed 2006-06-23, granted 2009-
5 12-15; Yao X, Diego P, Ramos AA, Shi Y. Averagine-scaling analysis and fragment ion mass defect labeling in peptide mass spectrometry. Anal. Chem. 2008 Oct 1;80(19):7383-91. doi: 10.1021/ac801096e; Sleno L. The use of mass defect in modern mass spectrometry. J. Mass. Spectrom. 2012, 47: 226–236. doi:10.1002/jms.2953

10

Bei dieser Methode der Massendefektfilterung wird demnach der zu einem spektralen Peak bestimmte Massendefekt genutzt, um das zugehörige Molekül chemisch näher zu charakterisieren. Eine ausreichend hohe Genauigkeit in der Massenbestimmung muss daher für diese Methode vorausgesetzt werden.

15

Zur grafischen Veranschaulichung der Massendefektfilterung wird gelegentlich eine von der üblichen Massenskala abweichende Darstellung verwendet, bei der in vertikaler Richtung anstelle des Massendefektes die Abweichung von der jeweils nächstgelegenen, dem Averagine-Modell entsprechenden Masse
20 abgetragen wird, siehe Yao X, Diego P, Ramos AA, Shi Y. Averagine-scaling analysis and fragment ion mass defect labeling in peptide mass spectrometry, Anal. Chem. 2008 Oct 1; 80(19):7383-91. doi: 10.1021/ac801096e.

Im jeweiligen Kontext dienen diese Diagramme lediglich der Illustration der
25 Methode mittels exemplarischer, synthetisch berechneter Peptidmassen. Eine Anwendung dieser Darstellungsform auf tatsächlich gemessene Daten ist nicht bekannt.

Aus einem über mehrere Spektren einer MALDI-Messung gebildeten
30 Mittelwertspektrum lässt sich ein Diagramm erstellen, das als Peptidmassendefektdiagramm (PMD) bezeichnet wird. Hierzu bestimmt man eine Liste aller lokalen Maxima und ihrer jeweiligen m/z -Werte, sowie für jeden m/z -Wert die Abweichung von der jeweils nächstgelegenen, dem theoretischen Peptidmassenmodell entsprechenden Masse. Unter der Annahme, dass die
35 gemessenen Signale auf Peptide zurückzuführen sind, wird nun für jeden m/z -

Wert m diejenige Nominalmasse m_N bestimmt, für die die Abweichung zwischen m und der nach dem theoretischen Peptidmassenmodell (Average-Modell, s. o.) erwarteten Masse $m_P(m_N)$ im Betrag minimiert wird (s.u.). Die minimale Abweichung $\delta_P(m)$, die Werte von -0,5 bis 0,5 annehmen kann, wird als
5 Peptidmodelldistanz bezeichnet. Die Peptidmodelldistanz entspricht der weiter oben beschriebenen Diskrepanz zwischen gemessenem und erwartetem Massendefekt.

Die Positionen aller lokalen Maxima werden nun in ein Diagramm eingetragen,
10 dessen horizontale Achse wiederum der Masse bzw. dem m/z-Wert entspricht, und auf dessen vertikaler Achse die oben bestimmte Abweichung vom Peptidmassenmodell abgetragen ist.

Gegenüber dem bekannten Massendefektdiagramm wird das PMD also durch
15 eine Transformation erzeugt, die die Positionen der theoretisch erwarteten Peptidmassendefekte auf die Referenzlinie 22 abbildet, die eine waagerechte Nulllinie ist. Darüber hinaus unterscheidet sich das PMD von bekannten Darstellungen dadurch, dass hierfür keine vorherige spezifische Signalanalyse durchgeführt wird, insbesondere keine Identifikation von signifikanten Peptidpeaks
20 (Peak Picking). Vielmehr spiegelt das PMD im Wesentlichen statistische Eigenschaften des spektralen Hintergrundsignals wieder (vgl. Abb. 1), von dem lediglich angenommen wird, dass es zu großen Anteilen von Peptidmolekülen herrührt. Die vertikale Achse erstreckt sich von -0,5 bis 0,5 und stellt die vorzeichenbehaftete Diskrepanz zwischen dem für ein Peptid erwarteten und
25 dem tatsächlich gemessenen Massendefekt dar. Zusätzlich zur Referenzlinie 22 kann ein Referenzkorridor mit Linien 23, 24 eingezeichnet werden, der die erwartete Streuung der Peptidmassendefekte um ihren Mittelwert darstellt, unter Berücksichtigung der für eine Messung gegebenen Diskretisierung der Massenachse.

30

An einem PMD lassen sich leicht folgende Qualitätseigenschaften eines Spektrums ablesen:

1. Massenbereich mit Peptidsignalen: Ein klar erkennbares Band
35 („Peptidband 21“) in der Nähe der Referenzlinie 22 lässt auf das

Vorhandensein von Peptidsignalen in dem jeweiligen Massenbereich schließen. Wo sich die Bandstruktur in einer unstrukturierten Punktwolke verliert (typischerweise erkennbar am oberen Ende der Massenachse, rechts von einer oberen Grenze 30), geht das Peptidsignal im Rauschen unter (Figuren 3,4).

2. Massenverschiebung: Weicht das Peptidband 21 erkennbar von der Referenzlinie 22 ab, so lässt dies auf eine Diskrepanz zwischen den wahren und den gemessenen Molekülmassen schließen. Die Größe der Diskrepanz entspricht dem vertikalen Versatz zwischen der Referenzlinie 22 und der Mittellinie 25 des Peptidbands 21. Der Referenzkorridor gibt dabei den Bereich an, in dem eine Verschiebung durch die jeweilige Diskretisierung der Massenachse erklärbar ist (Figuren 3,5).

3. Äquidistant oder zu grob gesamplete Massenachse: Ein MALDI-TOF Massenspektrometer diskretisiert die Massenachse typischerweise nicht äquidistant sondern mit zu hohen Massen hin zunehmender Binbreite. Bei der Nachverarbeitung von Spektraldaten werden die Daten häufig auf eine äquidistante Massenachse mit niedrigerer Auflösung resamplet, wodurch ein Genauigkeitsverlust entsteht. Eine äquidistante Massenachse ist im PMD deutlich durch eine lineare Struktur der aufgetragenen Punkte erkennbar (Figur 6). Ist die Auflösung der Daten nach dem Resampling zu niedrig, verliert sich die Struktur des Peptidbandes 21 (Figuren 6,7).

4. Fehlerhafte Kalibrierung der Massenachse: Bei der Kalibrierung der Massenachse können Fehler auftreten, die zu einer un stetigen Verzerrung der Massenachse führen. Solche Verzerrungen treten im PMD als ein Bruch bzw. Versatz im Peptidband 21 auf (Figur 8).

Im Vergleich zur Visualisierung mittels herkömmlicher Massendefektdiagramme sind in einem PMD die oben genannten Qualitätsmerkmale eines Spektrums sehr viel deutlicher erkennbar, insbesondere können auch kleinere oder auf Teilabschnitte der Massenachse beschränkte Massenverschiebungen als Abweichungen von der horizontalen Referenzlinie 22 leichter erfasst werden.

Statt für ein Mittelwertspektrum kann ein PMD auch für ein Einzelspektrum oder für das Maximumspektrum über mehrere Einzelspektren (sog. Skyline-Spektrum) gebildet werden, diese Darstellung ist jedoch weniger aussagekräftig.

5

Neben der reinen Visualisierung können die in einem PMD dargestellten Informationen auch wie folgt quantitativ ausgewertet werden (siehe auch mathematische Formulierung weiter unten):

10 1. Bestimmung der Diskrepanz zwischen gemessenem und erwartetem Peptidmassendefekt in Abhängigkeit von der Masse. Hierzu wird der Median der Massendefekte über Teilintervalle der Massenachse gebildet und mit dem erwarteten Wert verglichen.

15 2. Bestimmung der Streuung der Massendefekte um deren Mittelwert. Hierzu wird der Interquartilsabstand der Massendefekte über Teilintervalle bestimmt und in ein vorgegebenes Vielfaches der Standardabweichung einer angenommenen Normalverteilung umgerechnet.

20 3. Bestimmung des Massebereiches mit erkennbarem Peptidsignal. Hierzu wird die aus den Daten bestimmte Streuung der Massendefekte (Bereich zwischen Linien 26, 27) mit der Breite des Referenzkorridors (Bereich zwischen Linien 23, 24) verglichen und der Bereich ermittelt, in dem die Abweichung innerhalb einer gewählten Toleranz bleibt.

25

Diese quantitativen Informationen können sowohl im PMD dargestellt als auch numerisch angezeigt bzw. zur Qualitätsbewertung der Messung weiterverarbeitet werden.

30 Mit

$$S = (s_j, m_j)_{j=1 \dots n}, \text{ mit } n \in \mathbb{N}, 0 < m_1 < \dots < m_n$$

wird ein (Einzel-, Mittelwert- oder Skyline-) Spektrum bezeichnet, bestehend aus
35 den n Intensitäten $s_1 \dots s_n$ zu den m/z -Werten $m_1 \dots m_n$. Mit

$\text{floor}(x)$ für $x > 0$

wird der ganzzahligen Anteil einer positiven Zahl x bezeichnet, mit

$$5 \quad \varphi(x) = x - \text{floor}(x) \text{ für } x > 0$$

wird der Nachkommaanteil einer positiven Zahl x bezeichnet.

Das PMD der lokalen Maxima von S besteht aus der grafischen Darstellung der Punkte

10

$$\left\{ (m_i, \delta_P(m_i)) : s_i > \max_{0 < |j-i| \leq u} s_j \right\}$$

wobei $u \in \mathbb{N}$ den Radius der lokalen Umgebung bezeichnet, über die die lokalen Maxima gebildet werden, und die Funktion

15

$$\delta_P(m) = \varphi\left(\frac{m}{1+r_P} + 0,5\right) - 0,5$$

die vorzeichenbehaftete Diskrepanz zwischen dem für ein Peptid erwarteten und dem tatsächlich gemessenen Massendefekt beschreibt. Die obige Darstellung der Diskrepanz $\delta_P(m)$ ergibt sich wie folgt:

20

Der theoretisch erwartete Massendefekt eines Peptids mit Nominalmasse m_N beträgt

25

$$m_P - m_N = (1+r_P)m_N - m_N = r_P m_N.$$

Zu einer tatsächlich gemessenen Masse m eines Peptids wird dessen Nominalmasse als diejenige ganzzahlige Masse m_N angenommen, für die die absolute Differenz

30

$$|m - (1+r_P)m_N|$$

minimiert wird. Dies führt zu

$$m_N = \text{floor}\left(\frac{m}{1+r_p} + 0,5\right).$$

Die Diskrepanz $\delta_P(m)$ ergibt sich aus der Differenz zwischen gemessenem und erwarteten Massendefekt zu

5

$$\begin{aligned} \delta_P(m) &= \frac{(m - m_N) - (m_P - m_N)}{1 + r_p} = \frac{m - (1 + r_p) m_N}{1 + r_p} = \frac{m}{1 + r_p} - \text{floor}\left(\frac{m}{1 + r_p} + 0,5\right) \\ &= \varphi\left(\frac{m}{1 + r_p} + 0,5\right) - 0,5 \end{aligned}$$

Die Gewichtung der Differenz der Massendefekte mit $1/(1+r_p)$ dient der Normalisierung von $\delta_P(m)$ auf den Wertebereich $[-0,5 \dots 0,5]$.

10

Die Referenzlinie 22 der theoretisch erwarteten mittleren Massendefekte von Peptiden wird durch die Nulllinie $\delta_P = 0$ beschrieben. Zur Bestimmung des Referenzkorridors (Linien 23, 24) wird die erwartete Varianz $v(m)$ der Positionen der lokalen Maxima in Abhängigkeit von der Masse betrachtet, die durch die

Summe aus der Varianz der wahren Peptidmassen σ_P^2 und der von der Diskretisierung der Massenachse herrührenden Varianz abgeschätzt werden kann,

15

$$v(m) = \sigma_P^2(m) + \frac{\Delta m(m)^2}{12}.$$

20

Darin bezeichnet $\Delta m(m)$ die Breite der m/z -Bins an der Massenposition m . Der Referenzkorridor wird durch die Begrenzungslinien 23, 24 bzw.

$$d_P^{1,2}(m) = \pm \mu \sqrt{v(m)}$$

25

gebildet, wobei der Skalierungsfaktor $\mu > 0$ die Breite des Korridors als Vielfaches einer Standardabweichung angibt (typischerweise $\mu = 2$).

Zu dem Spektrum S sei eine Partitionierung I der Massenachse in paarweise disjunkte Intervalle I_k gegeben:

30

$$I = (I_k)_{k=1, \dots, K} \text{ mit } K \in \mathbb{N}, \cup_k I_k = [m_1, m_n].$$

Zu einem PMD, in dem die Punkte

$$(m_i, \delta_P(m_i))_{i \in L} \text{ mit } L = \left\{ i \in \{1 \dots n\} : s_i > \max_{0 < |j-i| \leq u} s_j \right\}$$

5

dargestellt sind, wird zur Bestimmung der Massendiskrepanz $E(m)$ für die Teilintervalle I_k die Diskrepanz

$$E_k = \text{median}\{\delta_P(m_i) : i \in L \cap I_k\}$$

10

gebildet. Die E_k werden als Punkte über den jeweiligen Mittelpunkten der zugehörigen Teilintervalle I_k dargestellt, dazwischen wird geeignet interpoliert (z.B. linear).

Die Streuung $e(m)$ der Massendefekte wird in analoger Weise aus den Interquartilsabständen (IQR) gebildet,

15

$$e_k = \mu \frac{\text{IQR}\{\delta_P(m_i) : i \in L \cap I_k\}}{2\sqrt{2}\text{erf}^{-1}(0,5)}$$

wobei der Skalierungsfaktor $\mu > 0$ wiederum die Breite des Korridors als Vielfaches einer Standardabweichung angibt, typischerweise $\mu = 2$, und erf die gaußsche Fehlerfunktion bezeichnet.

20

Als Massebereich mit erkennbarem Peptidsignal wird derjenige Teil der Massenachse bestimmt, für den das Verhältnis zwischen beobachteter (Linien 26, 27) und erwarteter Streuung (Linien 28, 29) unterhalb einer festgelegten

25 Toleranzschwelle t bleibt:

$$M_P = \left\{ m \in [m_1, m_n] : \frac{e(m)}{\mu\sqrt{v(m)}} \leq t \right\}$$

Ein typischer Toleranzwert ist $t = 1,2$. Die Positionen der äußeren Ränder von M_P können im PMD als vertikale Linien eingezeichnet werden.

30

Die zuvor beschriebene Darstellung eines Spektrums in einem PMD lässt sich prinzipiell sowohl auf Mittelwert- als auch auf Einzelspektren anwenden. Sie setzt

jedoch die Identifikation von lokalen Maxima in dem betreffenden Spektrum voraus und damit einen ausreichend hohen Signal-Rauschabstand, der bei Einzelspektren typischerweise nicht gegeben ist.

5 Durch die Darstellung der Spektren in einem Peptidmassendefekthistogramm (PMH) kann dieser Nachteil umgangen werden. Hierzu werden alle spektralen Intensitäten zu sämtlichen m/z -Bins eines Spektrums in einem 2D-Histogramm dargestellt, in dem wiederum die horizontale Achse der Massenachse entspricht und die vertikale Achse die Peptidmodellldistanz zur jeweiligen Masse darstellt
10 (s.u.). Beide Achsen werden gleichmäßig in vorher gewählte Anzahlen von Teilintervallen unterteilt (typischerweise 20-50, kann für beide Achsen unterschiedlich sein), woraus sich eine Partitionierung der Diagrammfläche in rechteckige Kacheln ergibt.

15 Das zu untersuchende Spektrum wird nun auf eine m/z -Auflösung, die der gewählten Unterteilung der Massendefektachse entspricht, hochinterpoliert. Für jede Kachel werden dann alle diejenigen Intensitätswerte des interpolierten Spektrums aufsummiert, deren Massen und Massendefekte in die jeweiligen Teilintervalle der horizontalen bzw. vertikalen Achse fallen.

20

Zur grafischen Darstellung können schließlich alle Kacheln mit einer geeignet gewählten Grauwert- oder Farbskala entsprechend der aufsummierten Intensitäten visualisiert werden. Wie im PMD werden zusätzlich die Referenzlinie 31 sowie der Referenzkorridor 32, 33 eingezeichnet (Fig. 9). Das PMH lässt sich
25 in gleicher Weise interpretieren wie das PMD.

Analog zur quantitativen Auswertung eines PMD lassen sich charakteristische Größen eines Spektrums auch aus einem PMH – und somit auch für Einzelspektren – berechnen. Hierzu wird zu jedem Teilintervall der horizontalen
30 Massenachse eine Auswertung jeweils der vertikal angeordneten, aufsummierten Intensitätswerte durchgeführt, um daraus Häufungspunkte und Streuungswerte zu bestimmen.

Es ist dabei zu beachten, dass die obere und die untere Randlinie eines PMH,
35 also die Punkte, die zu den extremen Distanzwerten $+0,5$ und $-0,5$ gehören, als

miteinander identifiziert aufgefasst werden können. Zur Beschreibung der Verteilung der aufsummierten Intensitätswerte in vertikaler Richtung sind daher zirkuläre Statistiken geeignet. Insbesondere kann das erste zirkuläre Moment Z als (komplexwertige) Statistik herangezogen werden (mathematische Formulierung s. u.). Die zirkulären Momente Z für alle Teilintervalle der Massenachse zusammengenommen werden hier als Massenshiftprofil des betrachteten Spektrums bezeichnen. Das komplexe Argument von Z entspricht (bis auf einen Faktor von $2\pi/(1+r_P)$) der Diskrepanz zwischen gemessenen und erwarteten Massen. Der Betrag von Z liefert ein reziprokes Maß für die Streuung der gemessenen Peptidmodellabstände: Der Wert $Z=0$ entspricht einer maximalen Streuung aller Messungen über das Intervall $[-0,5 .. 0,5]$, während im Extremfall einer minimalen Streuung, bei der alle Abstandswerte identisch sind, Z einen Wert mit Betrag 1 annimmt.

Für die tatsächliche Berechnung des Massenshiftprofils Z können die beiden Schritte der Bildung des 2D-Histogramms und der Berechnung der zirkulären Momente zusammengefasst und als Fourier-Integrale des Spektrums über die Teilintervalle der Massenachse ausgedrückt werden (s. u.). Numerisch lassen sich diese Integrale durch geeignete Integrationsformeln (z.B. Trapezformel oder Simpson'sche Regel) approximieren. Dabei kann auch auf eine feinere Diskretisierung und Interpolation des Spektrums verzichtet und direkt mit den diskreten spektralen Intensitäten in der ursprünglich vorliegenden Auflösung gerechnet werden.

Das Massenshiftprofil liefert eine Abschätzung der in einem Spektrum auftretenden Messfehler der gemessenen Massen gegenüber den wahren Massen. In der Praxis ist es häufig wünschenswert, diese Verschiebungen zu korrigieren und so zu einer höheren Genauigkeit der gemessenen Massen eines Spektrums zu gelangen.

30

Andererseits wird das Massenshiftprofil durch einen Abgleich der gemessenen Daten mit dem relativ einfachen, linearen Averagine-Modell (s. o.) gewonnen. Die Abschätzung der Massenfehler durch das Massenshiftprofil kann daher nicht genauer sein als die Genauigkeit des Modells selbst, das zumindest im unteren Massenbereich bis ca. 1000 Da für viele Anwendungen nicht ausreichend hoch

35

ist. Eine Korrektur der gemessenen Massen um die geschätzten Messfehler könnte daher dazu führen, dass Teile der Messung ungenauer werden.

5 Für viele Anwendungen ist jedoch nicht eine absolute Massengenauigkeit ausschlaggebend, sondern vielmehr eine möglichst gute Vergleichbarkeit zwischen einzelnen Spektren aus ein- und derselben oder aus mehreren Messungen. Der absolute Messfehler der gemessenen Massen eines Spektrums ist in diesen Fällen weniger relevant als die Differenzen der Messfehler innerhalb eines Ensembles von Spektren.

10

Das Verfahren der Massenshiftnormalisierung besteht darin, zu einem Ensemble zunächst für jedes Spektrum das jeweilige Massenshiftprofil zu ermitteln (s. o.), aus allen einzelnen Massenshiftprofilen ein gemeinsames, mittleres Referenzprofil zu bilden, und schließlich jedes Spektrum so zu modifizieren, dass
15 das Massenshiftprofil des modifizierten Spektrums dem Referenzprofil entspricht. Durch diese Angleichung wird die relative Abweichung zwischen den zu ein- und demselben Peptid gehörenden Signalpeaks der einzelnen Spektren verringert und die Vergleichbarkeit der Spektren wird erhöht (Fig. 10, unten).

20 Das Referenzprofil wird durch elementweise Bildung des arithmetischen Mittelwertes bestimmt (s. u.). Für die Normalisierung eines einzelnen Spektrums auf das Referenzprofil werden jeweils für die einzelnen Teilintervalle, für die die Massenshiftprofile berechnet wurden, relative Verschiebungswerte bestimmt und diese über die gesamte Massenachse interpoliert. Die gemessenen Massenwerte
25 des Spektrums werden dann um diese interpolierten Verschiebungswerte korrigiert.

Durch die Anwendung dieser Verschiebungen erhält jedes einzelne Spektrum eine eigene Massenachse. Für eine gemeinsame Auswertung eines Ensembles
30 von Spektren ist es meist wünschenswert, dass alle Spektren auf einer gemeinsamen Massenachse definiert sind. Dies lässt sich dadurch erreichen, dass eine gemeinsame Massenachse gebildet wird (z.B. durch Mittelung über alle einzelnen Massenachsen oder durch Auswahl einer beliebigen Massenachse als Referenzmassenachse) und anschließend jedes normalisierte Spektrum auf
35 die gemeinsame Massenachse interpoliert wird.

Peptidmassendefekthistogramm:

Mit

$$5 \quad S = (s_j, m_j)_{j=1 \dots n}, \text{ mit } n \in \mathbf{N}, 0 < m_1 < \dots < m_n$$

Wird wie oben ein (Einzel-, Mittelwert- oder Skyline-) Spektrum bezeichnet, bestehend aus den n Intensitäten $s_1 \dots s_n$ zu den m/z -Werten $m_1 \dots m_n$.

- 10 Zu einem Spektrum S sei wie oben eine Partitionierung I der Massenachse gegeben sowie eine weitere Partitionierung J des Intervalls $[-0,5 \dots 0,5]$,

$$I = (I_k)_{k=1 \dots K} \text{ mit } K \in \mathbf{N}, \bigcup_k I_k = [m_1, m_n],$$

$$J = (J_l)_{l=1 \dots L} \text{ mit } L \in \mathbf{N}, \bigcup_l J_l = \left[-\frac{1}{2}, \frac{1}{2}\right].$$

15

Zu den Partitionierungen I und J liefert

$$\Gamma_{k,l} = \{m \in I_k : \delta_{\mathbf{P}}(m) \in J_l\}$$

- 20 eine feinere Partitionierung der Massenachse, bei der die Teilintervalle $\Gamma_{k,l}$ den einzelnen Kacheln des 2D-Histogramms zugordnet sind. Weiter sei eine Interpolierende des Spektrums S durch eine kontinuierliche Funktion $\tilde{S}(m)$ gegeben,

$$25 \quad \tilde{S}: [m_1, m_n] \rightarrow \mathbb{R}, \text{ mit } \tilde{S}(m_j) = s_j, j = 1 \dots n.$$

Die Matrix $H(S) = (h_{k,l})$ sei durch

$$h_{k,l} = \frac{1 + \tau_{\mathbf{P}}}{\int_{I_k} \tilde{S}(t) dt} \int_{\Gamma_{k,l}} \tilde{S}(t) dt \quad \text{für } k = 1 \dots K, l = 1 \dots L$$

30

definiert. Zur numerischen Berechnung der $h_{k,l}$ kann z.B. \tilde{S} als lineare Interpolierende von S gewählt werden, die Integrale lassen sich dann exakt

auswerten. Zur Bildung des PMH wird die Matrix $H(S)$ als Grauwert- oder Falschfarbenbild dargestellt.

Massenshiftprofil:

- 5 Mit den Bezeichnungen und Definitionen des vorigen Abschnittes ist das erste zirkuläre Moment der Spalten von H gegeben durch die komplexwertigen Größen

$$Z_k = \sum_{i=1}^L h_{k,i} e^{2\pi i \bar{\delta}_i},$$

- 10 wobei $\bar{\delta}_i$ die Mittelwerte der Intervalle J_i bezeichnet. Im Grenzfall einer infinitesimal feinen Partitionierung J (d.h. $L \rightarrow \infty$) lassen sich die Z_k durch

$$Z_k = \frac{1}{\int_{I_k} \tilde{S}(t) dt} \int_{I_k} \tilde{S}(t) e^{i\omega t} dt, \text{ mit } \omega = \frac{2\pi}{1 + \tau_P}$$

- 15 darstellen. Zur konkreten numerischen Berechnung dieser Integrale können eine geeignete Integrationsformel (z.B. Trapezformel oder Simpson'sche Regel) und als Stützstellen von \tilde{S} gerade die diskreten Messpunkte $(m_j)_{j=1 \dots n}$ gewählt werden. Wegen $\tilde{S}(m_j) = s_j$ entfällt dann die Notwendigkeit, das Spektrum S explizit zu interpolieren.

Massenshiftnormalisierung:

- 20 Ein Ensemble von N Spektren S^i ($i = 1 \dots N$) sei gegeben, die eine gemeinsame Massenachse $(m_j)_{j=1 \dots n}$ aufweisen. Weiter sei eine Partitionierung I der Massenachse in K Teilintervalle wie oben gegeben. Die zu dieser Partitionierung für die einzelnen Spektren S^i berechneten Massenshiftprofile seien mit Z^i bezeichnet:

25

$$Z^i = (Z_k^i)_{k=1 \dots K} \text{ für } i = 1 \dots N$$

Zu den Massenshiftprofilen Z^i wird das Referenzprofil \bar{Z} durch elementweise arithmetische Mittelwertbildung berechnet:

30

$$\bar{Z} = (\bar{Z}_k)_{k=1 \dots K} \text{ mit } \bar{Z}_k = \frac{1}{N} \sum_{i=1}^N Z_k^i.$$

Sei nun $S = (s_j, m_j)_{j=1 \dots n}$ ein beliebiges, über derselben Massenachse definiertes Spektrum mit Massenshiftprofil $Z = (Z_k)_{k=1 \dots K}$. Zu jedem Teilintervall der Partitionierung I wird nun eine relative Verschiebung

5

$$\Delta_k = \frac{1}{2\pi} (\arg(Z_k) - \arg(\bar{Z}_k))$$

bestimmt, wobei $\arg(z) \in (-\pi, \pi]$ die komplexe Argumentfunktion bezeichnet. Die einzelnen Verschiebungen Δ_k werden den Mittelpunkten der Teilintervalle I_k zugeordnet und über die gesamte Massenachse interpoliert (typischerweise mittels linearer Interpolation). So wird ein Verschiebungsvektor $\Delta^* = (\Delta_j^*)_{j=1 \dots n}$ erhalten. Das normalisierte Spektrum S^* wird durch Anwendung der Verschiebungswerte auf die m/z -Werte des Spektrums S erhalten,

15

$$S^* = (s_j, m_j^*)_{j=1 \dots n}, \quad \text{mit } m_j^* = m_j + \Delta_j^*.$$

20

Bezugszeichenliste

20	Grundlinie
21	Peptidband
22	Referenzlinie
23	Linie für Referenzkorridor
24	Linie für Referenzkorridor
25	Mittellinie als Massendefektdiskrepanz
26	Linie für Massendefektstreuung
27	Linie für Massendefektstreuung
28	Linie für erwartete Streuung
29	Linie für erwartete Streuung
30	obere Grenze
31	Referenzlinie
32	Linie für Referenzkorridor
33	Linie für Referenzkorridor

Patentansprüche

1. Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF-Massenspektrometrie, mit folgenden Schritten:

- a) Bereitstellung erwarteter Massendefekte;
- b) Bestimmung gemessener Massendefekte, nämlich der sich aus den Daten der Massenspektrometrie ergebenden Massendefekte;
- c) Vergleich der gemessenen Massendefekte mit den erwarteten Massendefekten.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, dass die erwarteten Massendefekte berechnet werden aus

$$m_N r_p,$$

wobei m_N die Nominalmasse eines Peptids bezeichnet und

r_p vorzugsweise zwischen 10^{-3} und 10^{-4} liegt, insbesondere etwa $4,95 \times 10^{-4}$ beträgt.

3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet**, dass der Massendefekt zu einer gemessenen Masse m berechnet wird aus

$$m - \text{floor}\left(\frac{m}{1+r_p} + 0,5\right),$$

wobei die Funktion $\text{floor}(x)$ für ein beliebiges $x > 0$ den ganzzahligen Anteil von x bezeichnet.

4. Verfahren nach Anspruch 1 oder einem der weiteren Ansprüche, **dadurch gekennzeichnet**, dass die Diskrepanz δ_p zwischen dem gemessenen und dem erwarteten Massendefekt direkt aus der gemessenen Masse m berechnet wird als

$$\delta_p(m) = \varphi\left(\frac{m}{1+r_p} + 0,5\right) - 0,5$$

wobei die Funktion $\varphi(x) = x - \text{floor}(x)$ den Nachkommaanteil von x für ein beliebiges $x > 0$ bezeichnet.

5. Verfahren nach Anspruch 1 oder einem der weiteren Ansprüche, **dadurch gekennzeichnet**, dass zum Vergleich der gemessenen Massendefekte mit den erwarteten Massendefekten über Teilintervalle einer Massenachse jeweils der Median der gemessenen Massendefekte gebildet und mit dem erwarteten Massendefekt verglichen wird.
6. Verfahren nach Anspruch 1 oder einem der weiteren Ansprüche, **dadurch gekennzeichnet**, dass bei weiterer Verwendung der Daten die Messwerte korrigiert werden, nämlich in Abhängigkeit von der Abweichung der gemessenen Massendefekte von den erwarteten Massendefekten.
7. Verfahren nach Anspruch 1 oder einem der weiteren Ansprüche, **dadurch gekennzeichnet**, dass die gemessenen Massendefekte für lokale Maxima der spektralen Intensitäten berechnet werden.
8. Verfahren zum Auswerten von Daten einer Massenspektrometrie zur Analyse von Peptiden aus biologischen Proben, insbesondere einer MALDI-TOF-Massenspektrometrie, vorzugsweise nach einem der voranstehenden Ansprüche, mit folgenden Schritten:
 - a) Bereitstellung erwarteter Massendefekte;
 - b) Bestimmung gemessener Massendefekte, nämlich der sich aus den Daten der Massenspektrometrie ergebenden Massendefekte;
 - c) Bestimmung der Diskrepanzen zwischen den gemessenen Massendefekten und den erwarteten Massendefekten.
 - d) Bestimmung der Streuung der Diskrepanzen um deren Mittelwert;
 - e) Vergleich der Streuung mit einer definierten zulässigen Streuung.
9. Verfahren nach Anspruch 8, **dadurch gekennzeichnet**, dass zur Bestimmung einer Streuung der Massendefektdiskrepanzen Interquartilsabstände der ermittelten Diskrepanzen über Teilintervalle der Massenachse bestimmt werden, und dass bei Überschreitung eines Grenzwertes der Streuung insbesondere die Daten als fehlerhaft verworfen werden.
10. Verfahren nach Anspruch 8 oder 9, **dadurch gekennzeichnet**, dass ein Korridor für zulässige Streuungen gebildet ist durch Grenzwerte

$$d_P^{1,2}(m) = \pm \mu \sqrt{v(m)}$$

$$\text{wobei } v(m) = \sigma_P^2(m) + \frac{\Delta m(m)^2}{12}$$

und $\sigma_P(m_N) = \sigma_0 + s_P m_N$, mit $\sigma_0 \approx 0,02$ und $s_P \approx 2,0 \times 10^{-5}$,

und $\Delta m(m)$ die Breite von m/z-Bins an der Massenposition m bezeichnet,

und m/z-Bins die sich durch Diskretisierung der Massenachse ergebenden Intervalle repräsentieren,

und $\mu > 0$ einen Skalierungsfaktor angibt, vorzugsweise $\mu = 2$,

und dass insbesondere die Daten als fehlerhaft verworfen werden, wenn die Streuung der Massendefektdiskrepanzen außerhalb des derart vorgegebenen Korridors liegt.

11. Verfahren nach Anspruch 8 oder einem der weiteren Ansprüche, **dadurch gekennzeichnet**, dass der Massebereich mit erkennbarem Peptidsignal bestimmt wird als die Gesamtheit aller Teilintervalle der Massenachse, für die der Quotient aus der tatsächlichen Streuung und der maximal zulässigen Streuung $d_P(m)$ einen festgelegten Schwellwert t nicht überschreitet, wobei vorzugsweise $t = 1,2$ festgelegt wird, und dass bei Über- oder Unterschreitung eines Grenzwertes für die untere bzw. obere Grenze dieses Massebereiches oder dessen Ausdehnung insbesondere die Daten als fehlerhaft verworfen werden.

12. Massenspektrometrisches Verfahren zur Analyse von Peptiden aus biologischen Proben, insbesondere unter Verwendung eines MALDI-TOF-Massenspektrometers, mit folgenden Schritten:

- a) Durchführung eines oder mehrerer massenspektrometrischer Untersuchungen an der biologischen Probe und Bereitstellung von sich aus den massenspektrometrischen Untersuchungen ergebenden Daten;
- b) Durchführung des Verfahrens nach einem der voranstehenden Ansprüche.

13. Massenspektrometrisches Verfahren zur Handhabung eines Einzelspektrums von Peptiden aus biologischen Proben mit folgenden Schritten:

- a) Messung von m/z -Werte für verschiedene Peptide mittels einer Messapparatur, insbesondere unter Verwendung eines MALDI-TOF-Massenspektrometers,
- b) Zuordnung der gemessenen m/z -Werte zu entsprechenden m/z -Bins eines 2D-Histogramms,
- c) Auftrag von spektralen Intensitäten der m/z -Bins in dem 2D-Histogramm, wobei auf einer Abszissenachse die m/z -Bins und auf einer Ordinatenachse die Diskrepanz zwischen gemessenem und erwartetem Massendefekt aufgetragen werden,
- d) wobei die durch die beiden Achsen aufgespannte Diagrammfläche in eine Vielzahl, vorzugsweise 20 bis 50, von Rechtecken aufgeteilt wird,
- e) wobei die gemessenen Werte auf eine m/z -Auflösung, die der gewählten Unterteilung der Ordinatenachse entspricht, hochinterpoliert, und
- f) wobei für jedes Rechteck diejenigen Intensitätswerte des interpolierten Spektrums aufsummiert werden, deren Massendefektdiskrepanz in die jeweiligen Teilintervalle der Achsen fallen,
- g) wobei unterschiedliche Intensitätswerte unterschiedlich gekennzeichnet werden und gleiche Intensitätswerte gleich gekennzeichnet werden

14. Massenspektrometrisches Verfahren nach Anspruch 13, **dadurch gekennzeichnet**, dass zu jedem Teilintervall der Abszissenachse, insbesondere der horizontalen Massenachse, die entsprechenden Intensitätswerte der Ordinatenachse statistisch ausgewertet werden, um Häufungspunkte und/oder Streuungswerte zu bestimmen.

15. Massenspektrometrisches Verfahren nach Anspruch 14, **dadurch gekennzeichnet**, dass zur Beschreibung der Verteilung der Intensitätswerte in vertikaler Richtung zirkuläre Statistiken verwendet werden, insbesondere kann ein erstes zirkuläres Moment Z als (komplexwertige) Statistik herangezogen werden.

16. Massenspektrometrisches Verfahren nach Anspruch 15, **dadurch gekennzeichnet**, dass zur Bestimmung eines Massenshiftprofils (Vektor aller

zirkulären Moment Z für alle Teilintervalle der Abszissenachse, insbesondere der horizontalen Massenachse) die Schritte zur Bildung des Histogramms und der Berechnung der zirkulären Momente zusammengesetzt werden und gemäß der Formel

$$Z_k = \frac{1}{\int_{I_k} \tilde{S}(t) dt} \int_{I_k} \tilde{S}(t) e^{i\omega t} dt, \text{ mit } \omega = \frac{2\pi}{1 + r_p}$$

als Fourier-Integrale des kontinuierlich interpolierten Spektrums \tilde{S} über die Teilintervalle I_k der Abszissenachse ausgedrückt werden.

17. Massenspektrometrisches Verfahren zur Handhabung eines Einzelspektrums von Peptiden aus biologischen Proben, insbesondere nach einem der Ansprüche 13 bis 16, **dadurch gekennzeichnet**, dass für eine Massenshiftnormalisierung zu einem Ensemble für jedes Spektrum

- a) ein Massenshiftprofil ermittelt wird,
- b) aus allen einzelnen Massenshiftprofilen durch elementweise Bildung des arithmetischen Mittelwertes ein gemeinsames, mittleres Referenzprofil gebildet wird,
- c) jedes Spektrum modifiziert wird, so dass das Massenshiftprofil des modifizierten Spektrums dem Referenzprofil entspricht.

18. Massenspektrometrisches Verfahren nach Anspruch 17, **dadurch gekennzeichnet**, dass für die Normalisierung eines einzelnen Spektrums auf das Referenzprofil jeweils für die einzelnen Teilintervalle, für die die Massenshiftprofile berechnet werden, relative Verschiebungswerte bestimmt und über die gesamte Massenachse interpoliert werden und die gemessenen Werte des Spektrums um diese interpolierten Verschiebungswerte korrigiert werden.

19. MALDI-TOF-Massenspektrometer mit einer Steuereinheit zur Analyse von Peptiden aus biologischen Proben unter Anwendung eines Verfahrens nach einem der Ansprüche 1 bis 18.

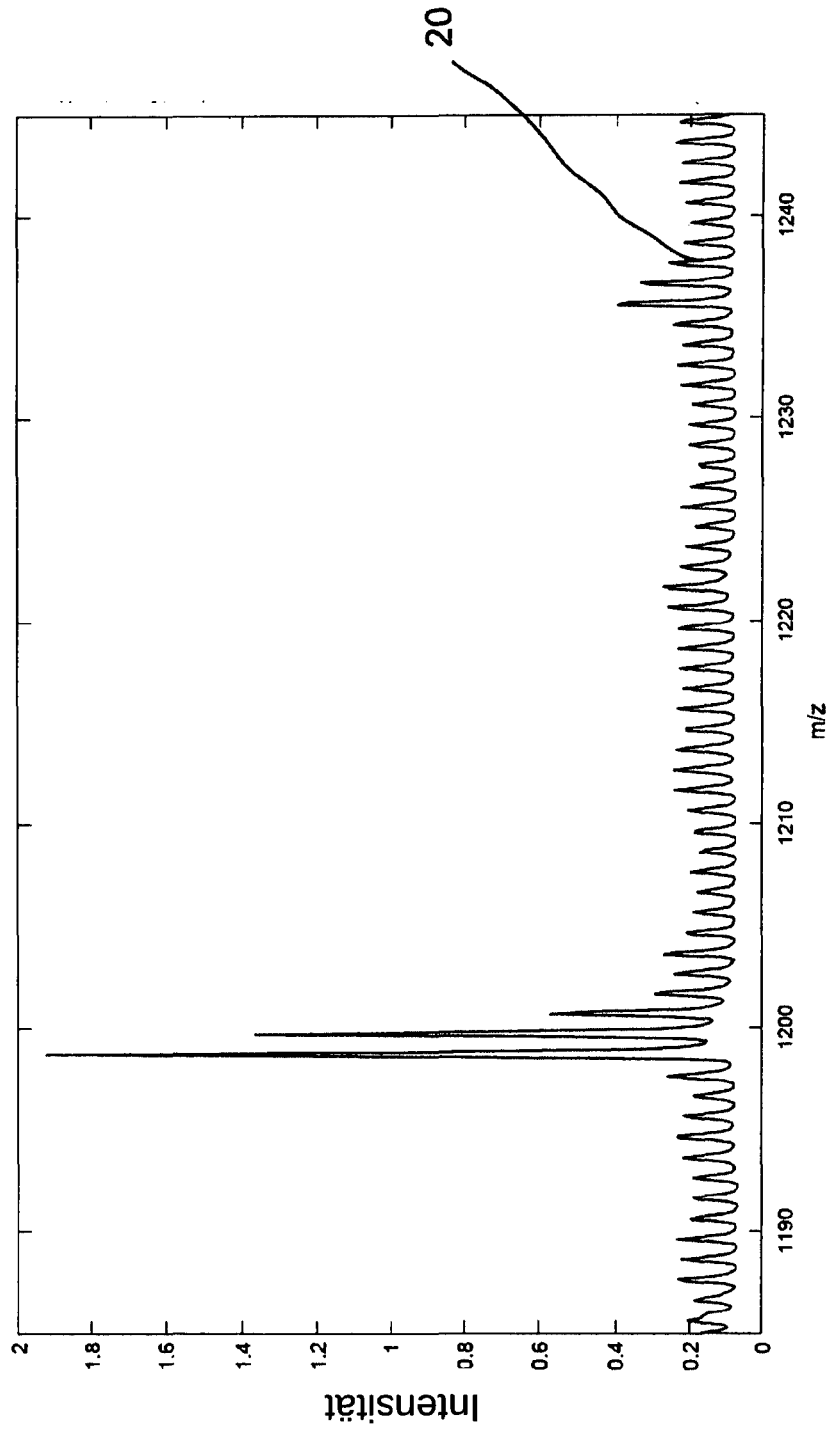


Fig. 1

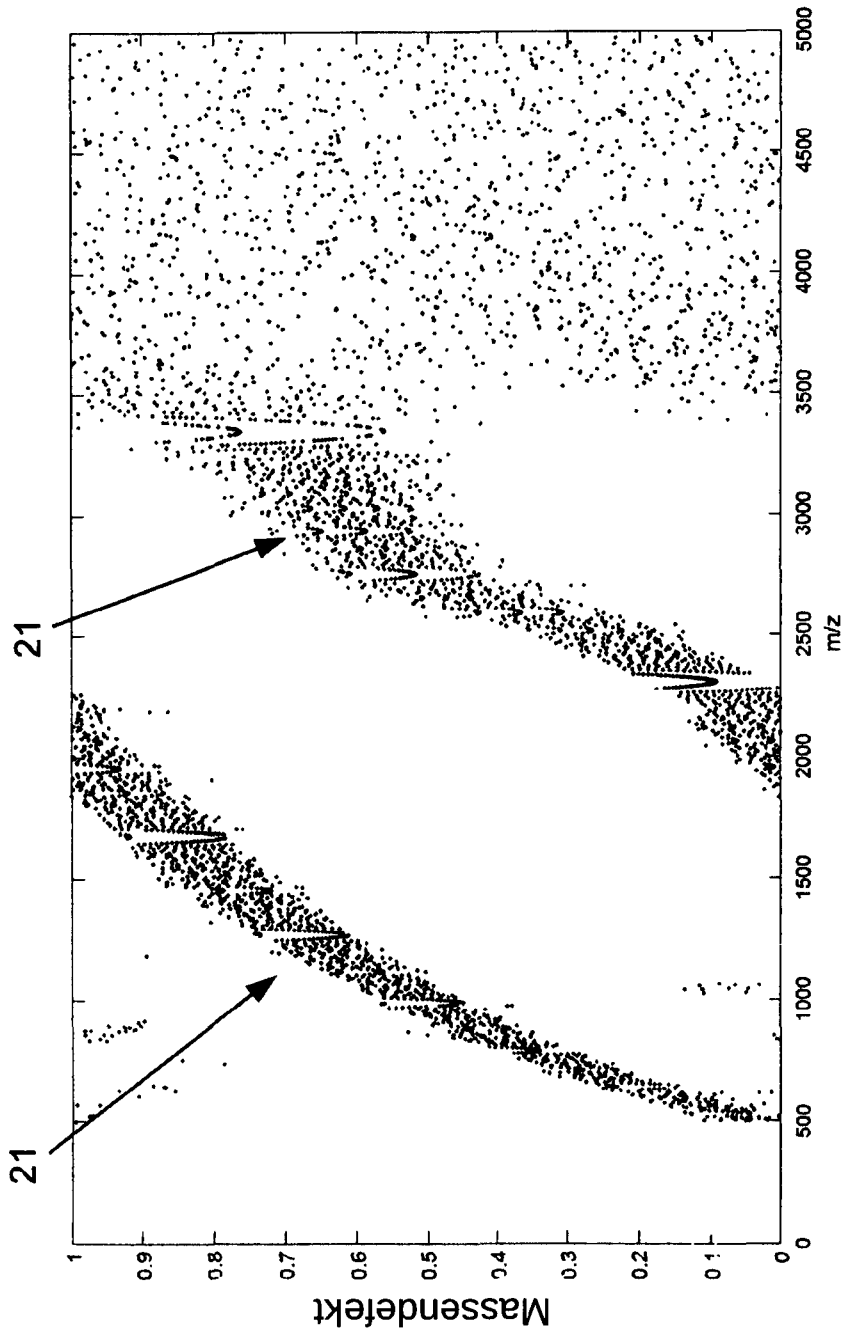


Fig. 2

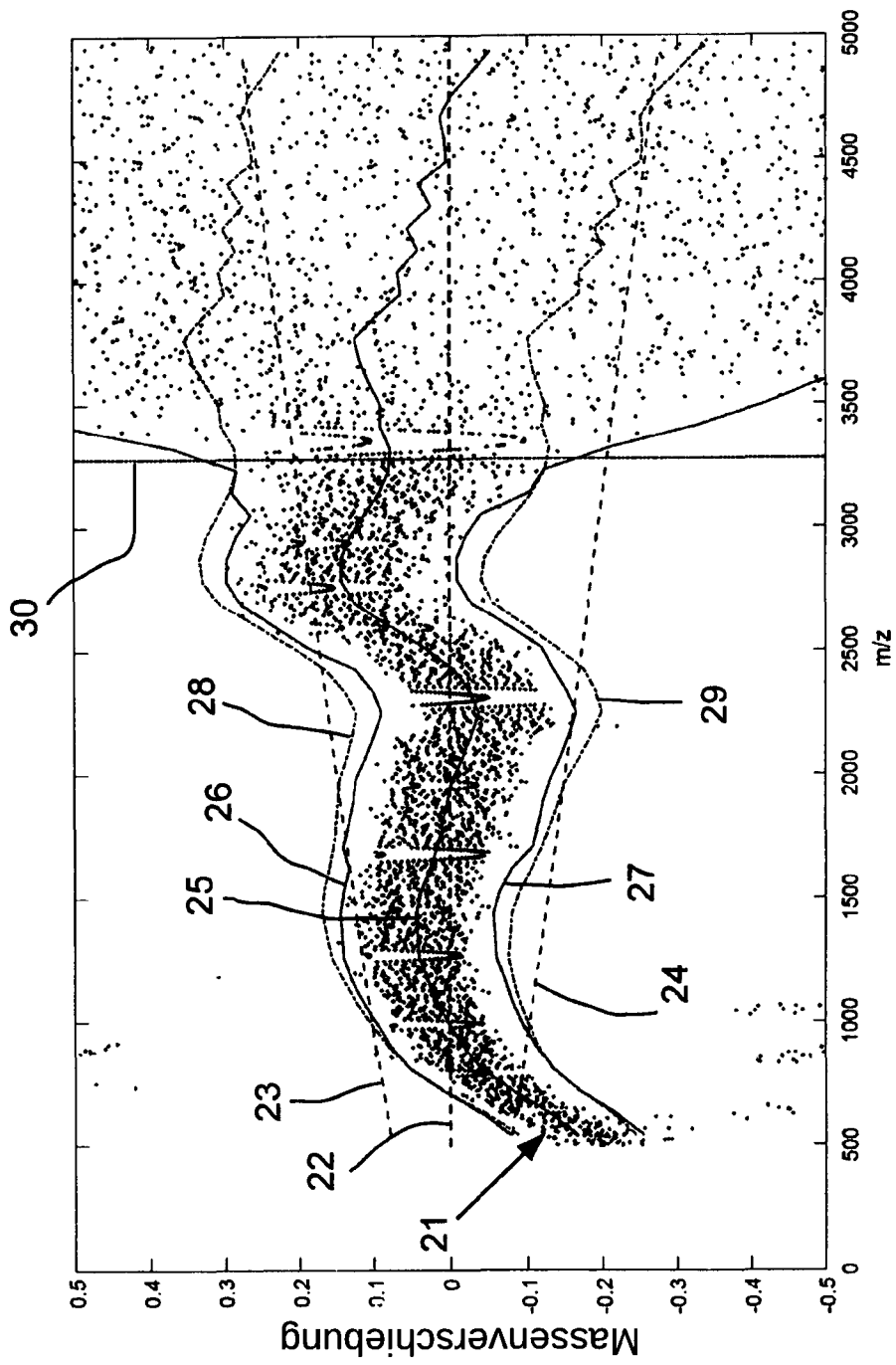


Fig. 3

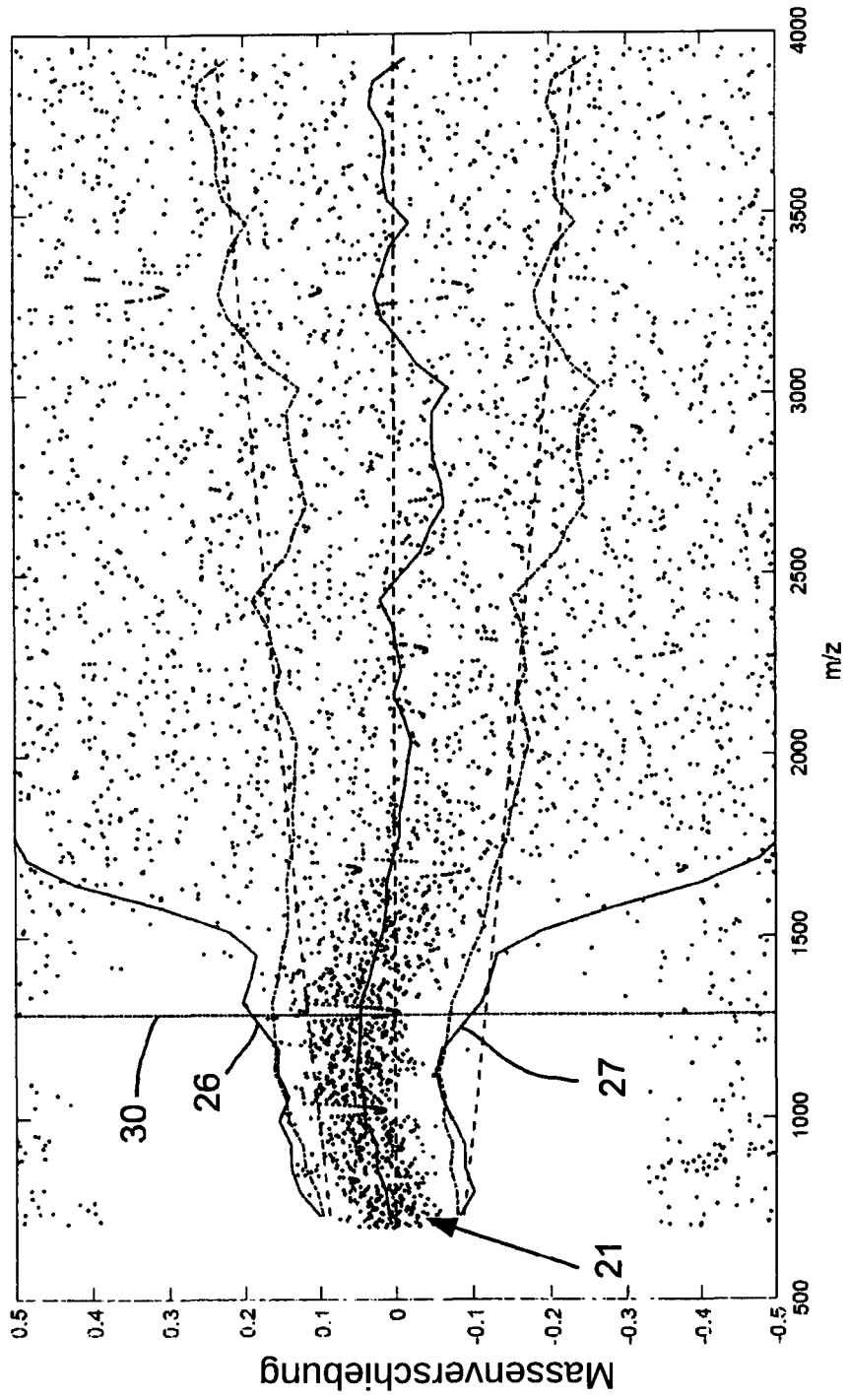


Fig. 4

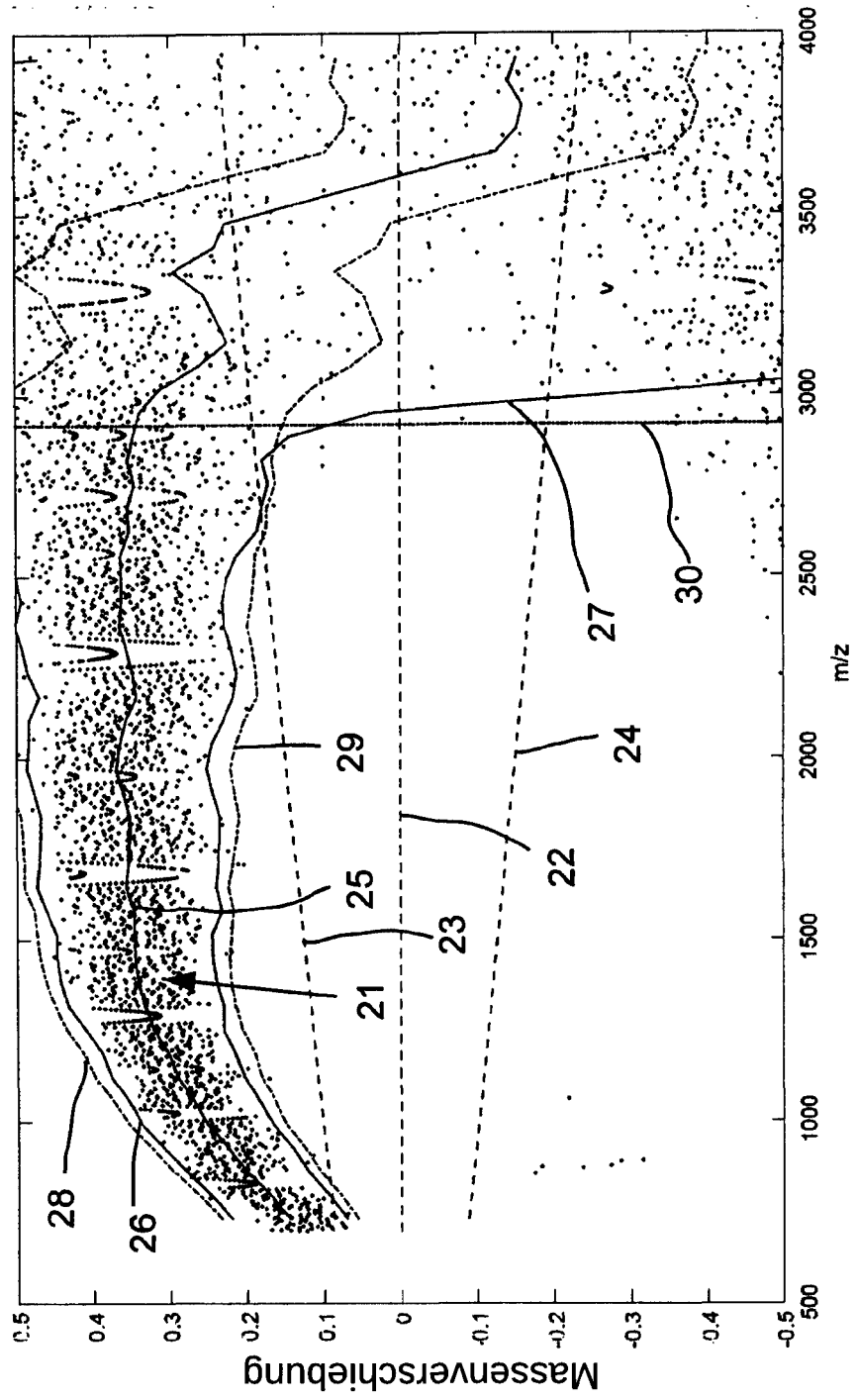


Fig. 5

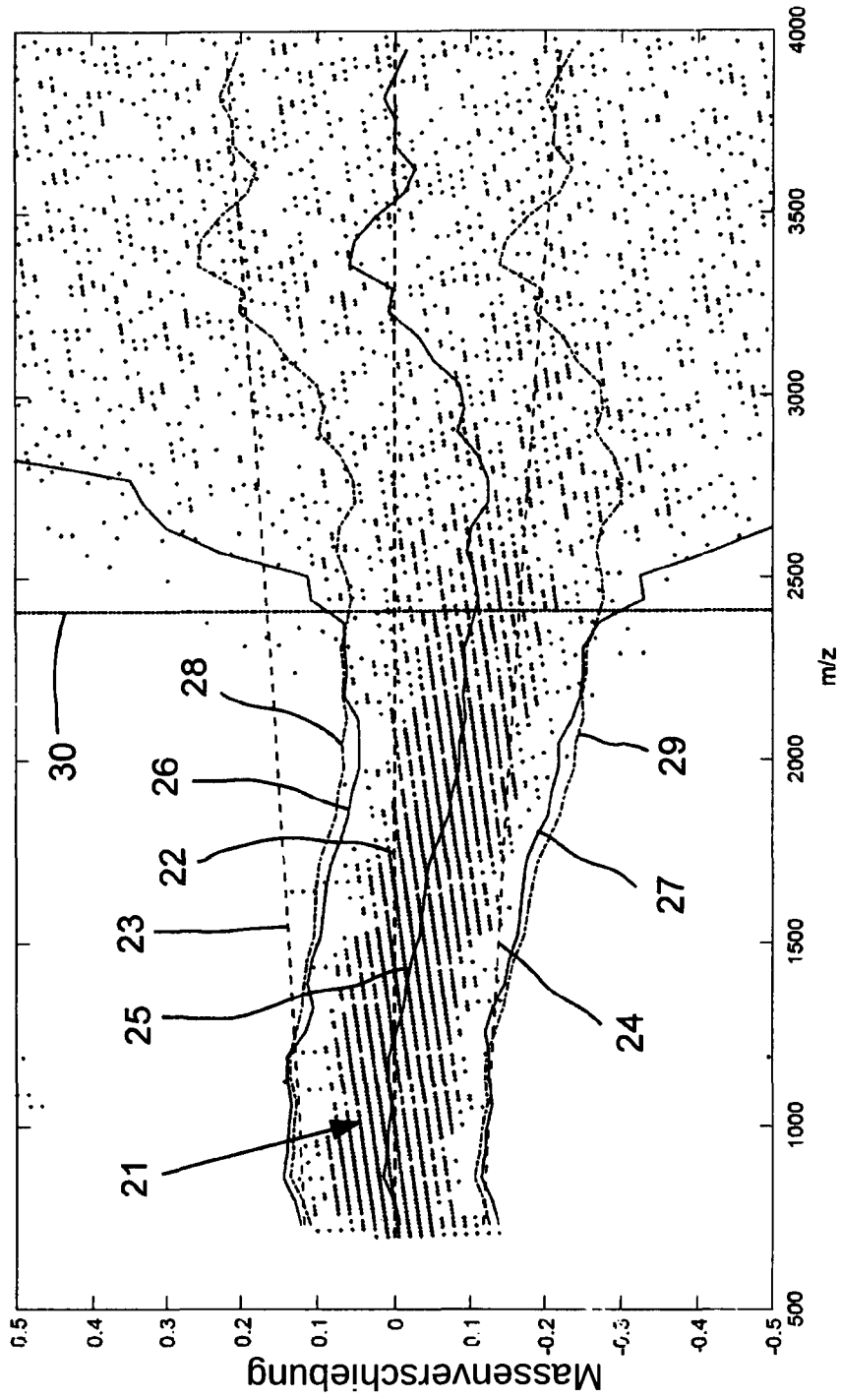


Fig. 6

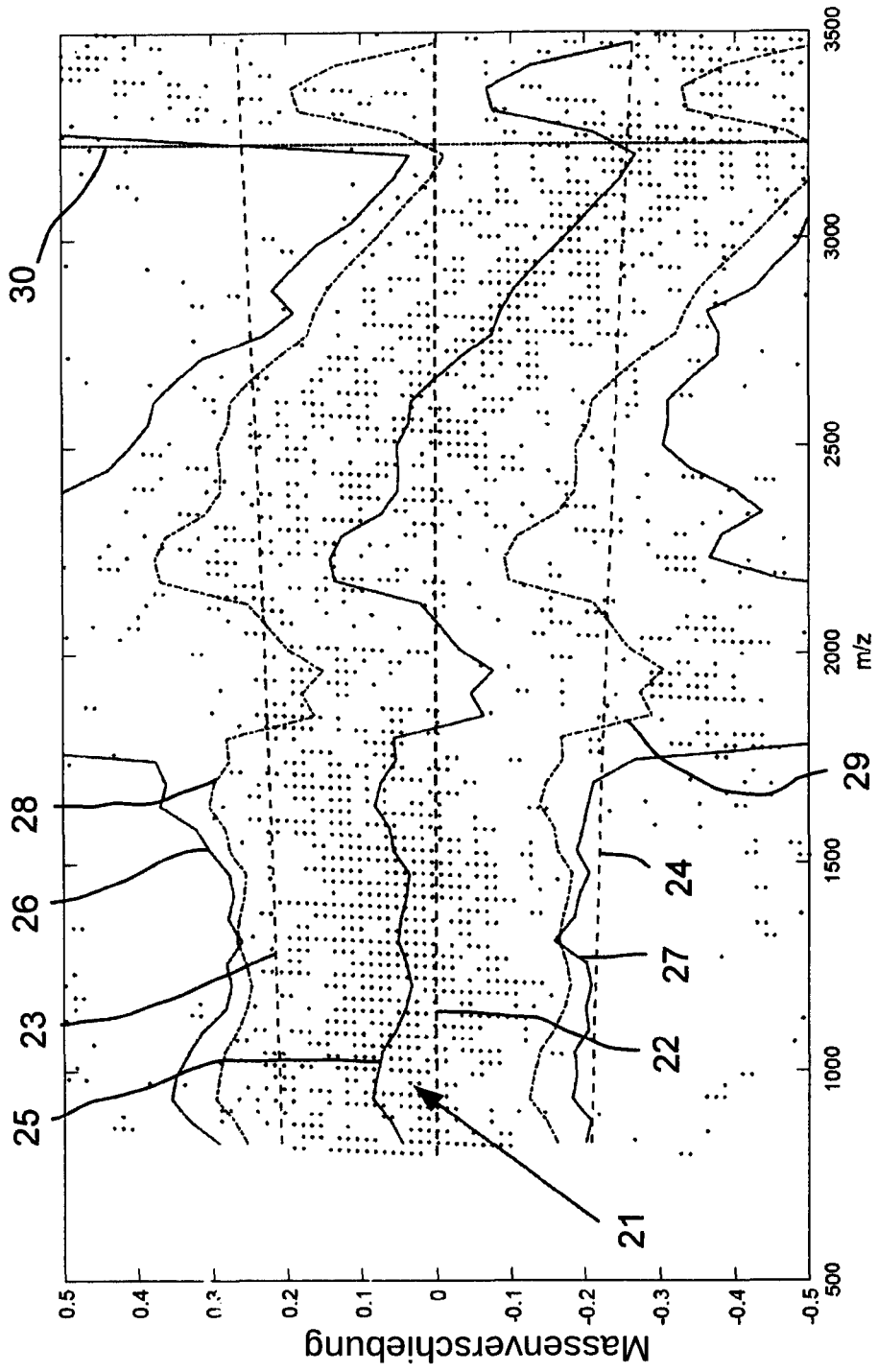


Fig. 7

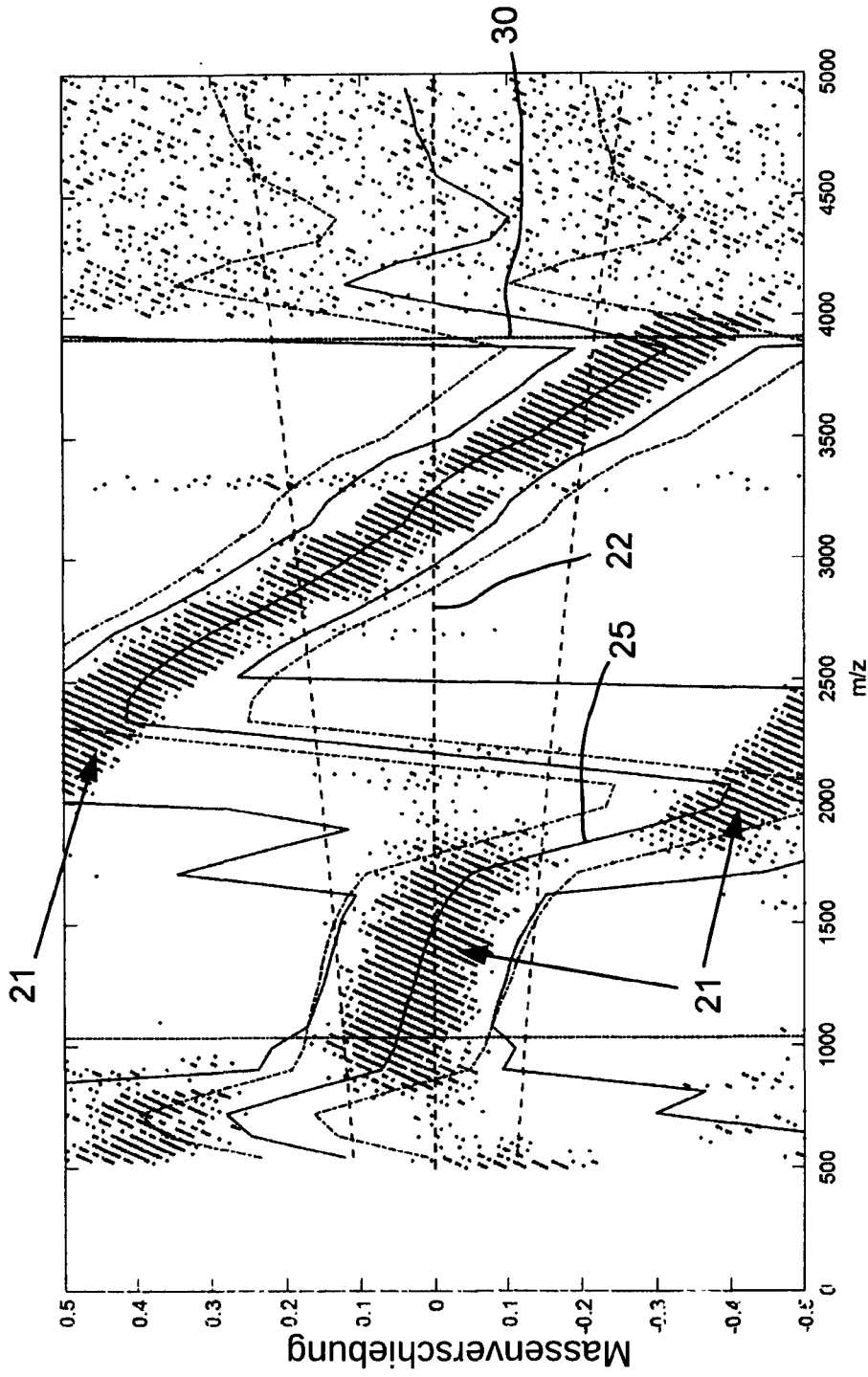
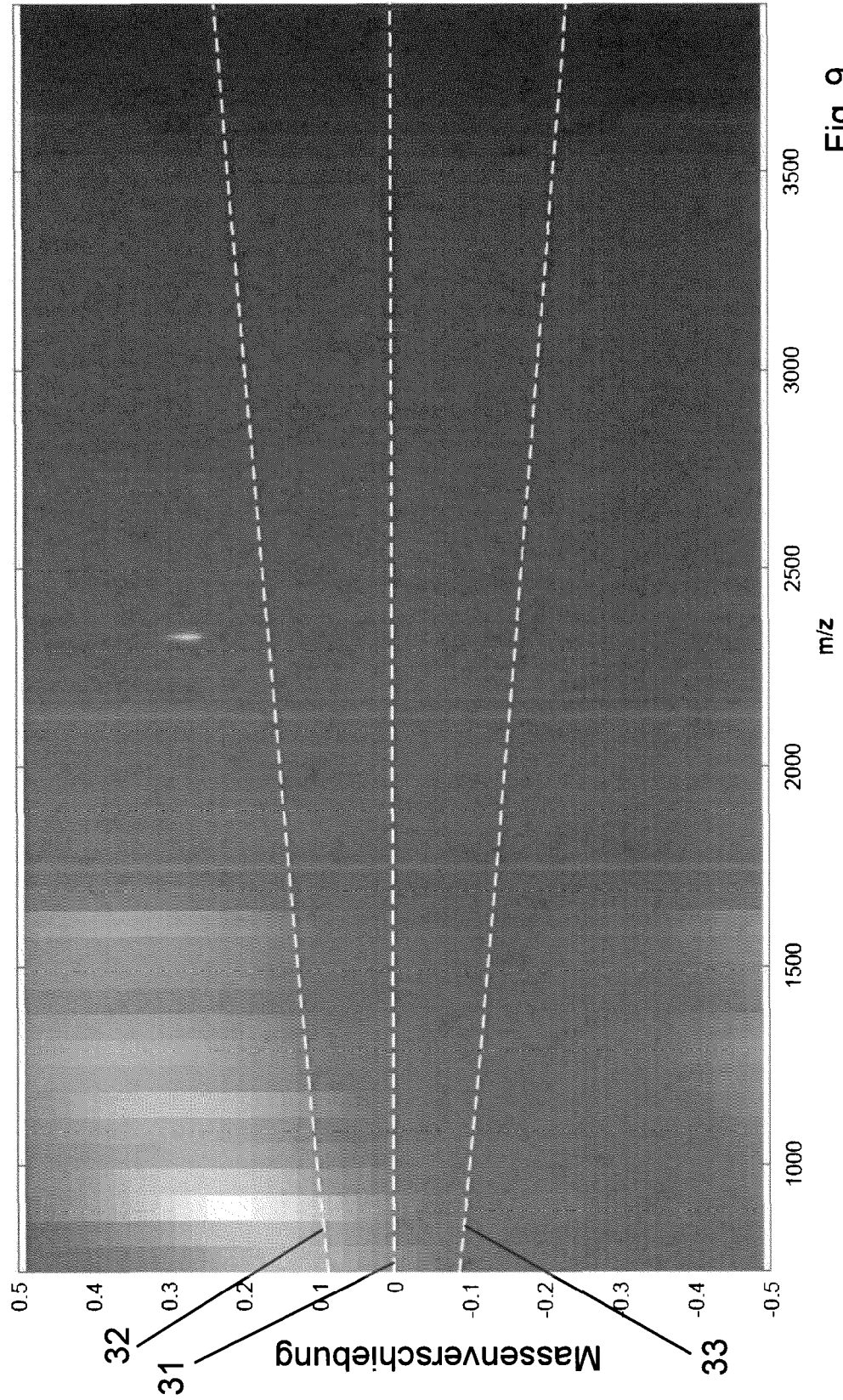


Fig. 8



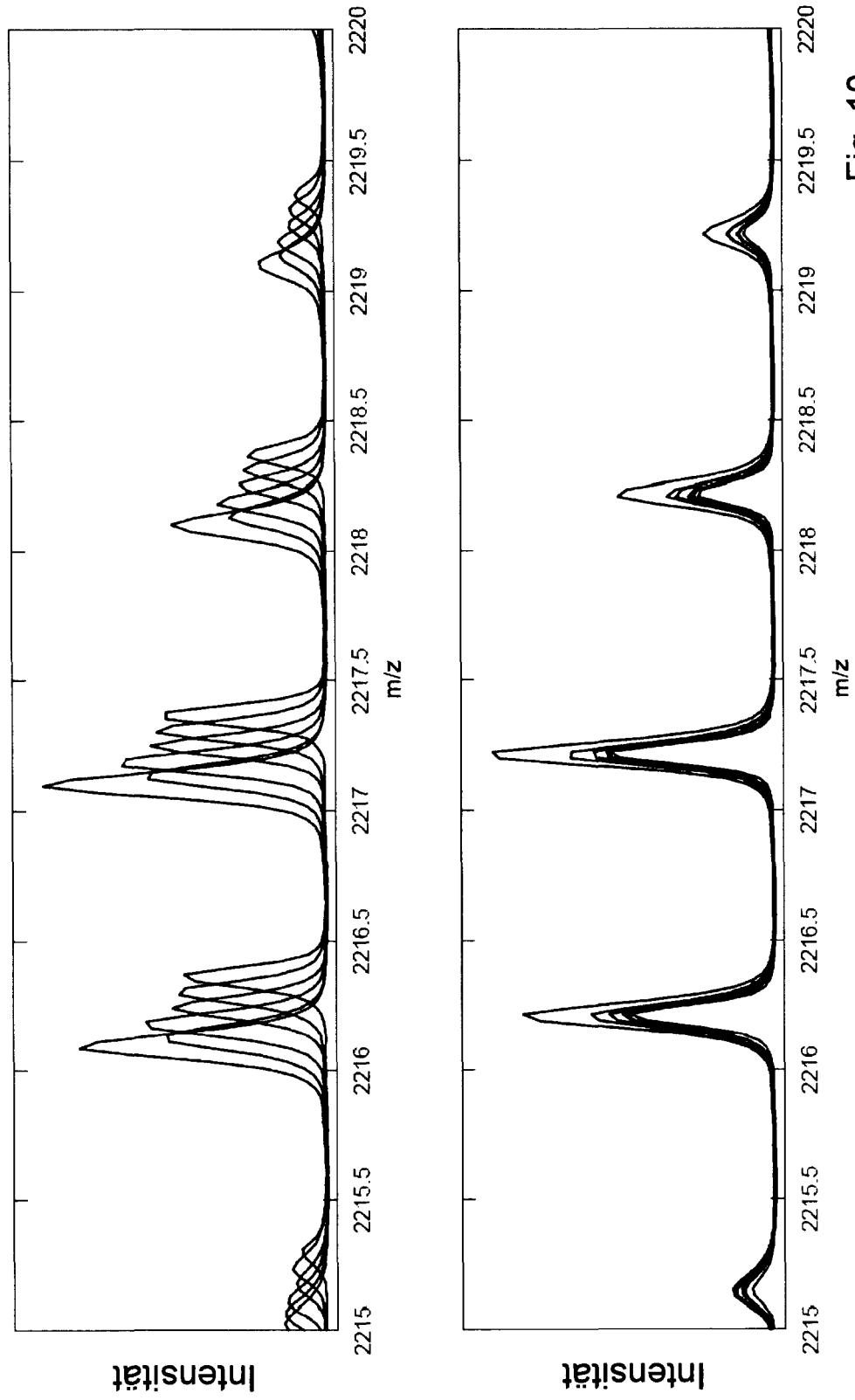


Fig. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/001131

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/10 G01N33/68 H01J49/00
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
G06F G01N H01J
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BAJRAMI B ET AL: "Shifting Unoccupied Spectral Space in Mass Spectrum of Peptide Fragment Ions", JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, ELSEVIER SCIENCE INC, US, vol. 20, no. 11, 1 November 2009 (2009-11-01), pages 2124-2134, XP026688360, ISSN: 1044-0305, DOI: 10.1016/J.JASMS.2009.07.005 [retrieved on 2009-07-14] the whole document abstract page 2124 - page 2123; figures 1-6; tables 1, 2 ----- -/--	1-19

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

18 December 2017

Date of mailing of the international search report

03/01/2018

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Boiangiu, Clara

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/001131

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MELINDA L. TOUMI ET AL: "Improving Mass Defect Filters for Human Proteins", JOURNAL OF PROTEOME RESEARCH., vol. 9, no. 10, 1 October 2010 (2010-10-01), pages 5492-5495, XP055435402, US ISSN: 1535-3893, DOI: 10.1021/pr100291q	1,12-19
Y	the whole document abstract page 1 - page 5	8-11
X	INDRANIL MITRA ET AL: "Improved Mass Defect Model for Theoretical Tryptic Peptides", ANALYTICAL CHEMISTRY, vol. 84, no. 6, 7 March 2012 (2012-03-07), pages 3026-3032, XP055435405, US ISSN: 0003-2700, DOI: 10.1021/ac203255e	1-19
Y	the whole document abstract pages 3-8	
Y	US 2012/232805 A1 (SADYGOV ROVSHAN G [US] ET AL) 13 September 2012 (2012-09-13) the whole document paragraph [0007] - paragraphs [0007], [0009] - [0011], [0013] - [0015], [0031], [0048] - [0053], [0063], [0097] - [0104]; claims 1-29	1-19
Y	HAIYING ZHANG ET AL: "Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry", JOURNAL OF MASS SPECTROMETRY., vol. 44, no. 7, 1 July 2009 (2009-07-01), pages 999-1016, XP055330389, GB ISSN: 1076-5174, DOI: 10.1002/jms.1610	1-19
Y	the whole document abstract page 1000, right-hand column - page 1014; figures 1-10; tables 1-6	
Y	CAN BRUCE ET AL: "Probabilistic Enrichment of Phosphopeptides by Their Mass Defect", ANALYTICAL CHEMISTRY, vol. 78, no. 13, 1 July 2006 (2006-07-01), pages 4374-4382, XP055435415, US ISSN: 0003-2700, DOI: 10.1021/ac060046w	1-19
	the whole document	
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/001131

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2007/076606 A1 (MDS INC DBA MDS SCIEIX [CA]; APPLERA CORP [US]; BLOOMFIELD NIC [CA]; LE) 12 July 2007 (2007-07-12) the whole document	1-19
Y	----- DEL PRETE EUGENIO ET AL: "Comparative Analysis of MALDI-TOF Mass Spectrometric Data in Proteomics: A Case Study", 31 July 2016 (2016-07-31), NETWORK AND PARALLEL COMPUTING; [LECTURE NOTES IN COMPUTER SCIENCE; LECT.NOTES COMPUTER], SPRINGER INTERNATIONAL PUBLISHING, CHAM, PAGE(S) 154 - 164, XP047351763, ISSN: 0302-9743 ISBN: 978-3-540-28012-5 [retrieved on 2016-07-31] the whole document	1-19
Y	----- WO 2014/130627 A1 (CHILDRENS MEDICAL CENTER [US]) 28 August 2014 (2014-08-28) the whole document	1-19
Y	----- US 2007/038387 A1 (CHEN XUNMING [US] ET AL) 15 February 2007 (2007-02-15) the whole document	1-19
Y	----- Assaf Wool ET AL: "Precalibration of matrix-assisted laser desorption/ionization-time of flight spectra for peptide mass fingerprinting", Proteomics, 1 January 2002 (2002-01-01), pages 1365-1373, XP055435027, Retrieved from the Internet: URL: <a href="http://onlinelibrary.wiley.com/store/10.1002/1615-9861(200210)2:10<1365::AID-PROT1365>3.0.CO;2-9/asset/1365ftp.pdf?v=1&t=jb5dxuss&s=bc88802156716333f237a20bd4e028cd99b57b3a">http://onlinelibrary.wiley.com/store/10.1002/1615-9861(200210)2:10<1365::AID-PROT1365>3.0.CO;2-9/asset/1365ftp.pdf?v=1&t=jb5dxuss&s=bc88802156716333f237a20bd4e028cd99b57b3a [retrieved on 2017-12-13] the whole document	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/001131

Patent document cited in search report	Publication date	Publication date	Patent family member(s)	Publication date
US 2012232805	A1	13-09-2012	NONE	

WO 2007076606	A1	12-07-2007	CA 2629203 A1	12-07-2007
			EP 1969615 A1	17-09-2008
			JP 5252205 B2	31-07-2013
			JP 2009522558 A	11-06-2009
			US 2007164207 A1	19-07-2007
			WO 2007076606 A1	12-07-2007

WO 2014130627	A1	28-08-2014	US 2016003842 A1	07-01-2016
			WO 2014130627 A1	28-08-2014

US 2007038387	A1	15-02-2007	NONE	

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES INV. G06F19/10 G01N33/68 H01J49/00 ADD.		
Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC		
B. RECHERCHIERTE GEBIETE		
Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole) G06F G01N H01J		
Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe) EPO-Internal, WPI Data		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	BAJRAMI B ET AL: "Shifting Unoccupied Spectral Space in Mass Spectrum of Peptide Fragment Ions", JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, ELSEVIER SCIENCE INC, US, Bd. 20, Nr. 11, 1. November 2009 (2009-11-01), Seiten 2124-2134, XP026688360, ISSN: 1044-0305, DOI: 10.1016/J.JASMS.2009.07.005 [gefunden am 2009-07-14] das ganze Dokument Zusammenfassung Seite 2124 - Seite 2123; Abbildungen 1-6; Tabellen 1, 2 ----- -/--	1-19
<input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen <input checked="" type="checkbox"/> Siehe Anhang Patentfamilie		
* Besondere Kategorien von angegebenen Veröffentlichungen : "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist "E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt) "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist "T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist "X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden "Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist		
Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts	
18. Dezember 2017	03/01/2018	
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter Boiangiu, Clara	

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	MELINDA L. TOUMI ET AL: "Improving Mass Defect Filters for Human Proteins", JOURNAL OF PROTEOME RESEARCH., Bd. 9, Nr. 10, 1. Oktober 2010 (2010-10-01), Seiten 5492-5495, XP055435402, US ISSN: 1535-3893, DOI: 10.1021/pr100291q	1,12-19
Y	das ganze Dokument Zusammenfassung Seite 1 - Seite 5	8-11
X	INDRANIL MITRA ET AL: "Improved Mass Defect Model for Theoretical Tryptic Peptides", ANALYTICAL CHEMISTRY, Bd. 84, Nr. 6, 7. März 2012 (2012-03-07), Seiten 3026-3032, XP055435405, US ISSN: 0003-2700, DOI: 10.1021/ac203255e	1-19
Y	das ganze Dokument Zusammenfassung Seiten 3-8	
Y	US 2012/232805 A1 (SADYGOV ROVSHAN G [US] ET AL) 13. September 2012 (2012-09-13) das ganze Dokument Absatz [0007] - Absätze [0007], [0009] - [0011], [0013] - [0015], [0031], [0048] - [0053], [0063], [0097] - [0104]; Ansprüche 1-29	1-19
Y	HAIYING ZHANG ET AL: "Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry", JOURNAL OF MASS SPECTROMETRY., Bd. 44, Nr. 7, 1. Juli 2009 (2009-07-01), Seiten 999-1016, XP055330389, GB ISSN: 1076-5174, DOI: 10.1002/jms.1610	1-19
Y	das ganze Dokument Zusammenfassung Seite 1000, rechte Spalte - Seite 1014; Abbildungen 1-10; Tabellen 1-6	
Y	CAN BRUCE ET AL: "Probabilistic Enrichment of Phosphopeptides by Their Mass Defect", ANALYTICAL CHEMISTRY, Bd. 78, Nr. 13, 1. Juli 2006 (2006-07-01), Seiten 4374-4382, XP055435415, US ISSN: 0003-2700, DOI: 10.1021/ac060046w	1-19
	das ganze Dokument	

	-/--	

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	WO 2007/076606 A1 (MDS INC DBA MDS SCIEX [CA]; APPLERA CORP [US]; BLOOMFIELD NIC [CA]; LE) 12. Juli 2007 (2007-07-12) das ganze Dokument	1-19
Y	----- DEL PRETE EUGENIO ET AL: "Comparative Analysis of MALDI-TOF Mass Spectrometric Data in Proteomics: A Case Study", 31. Juli 2016 (2016-07-31), NETWORK AND PARALLEL COMPUTING; [LECTURE NOTES IN COMPUTER SCIENCE; LECT.NOTES COMPUTER], SPRINGER INTERNATIONAL PUBLISHING, CHAM, PAGE(S) 154 - 164, XP047351763, ISSN: 0302-9743 ISBN: 978-3-540-28012-5 [gefunden am 2016-07-31] das ganze Dokument	1-19
Y	----- WO 2014/130627 A1 (CHILDRENS MEDICAL CENTER [US]) 28. August 2014 (2014-08-28) das ganze Dokument	1-19
Y	----- US 2007/038387 A1 (CHEN XUNMING [US] ET AL) 15. Februar 2007 (2007-02-15) das ganze Dokument	1-19
Y	----- Assaf Wool ET AL: "Precalibration of matrix-assisted laser desorption/ionization-time of flight spectra for peptide mass fingerprinting", Proteomics, 1. Januar 2002 (2002-01-01), Seiten 1365-1373, XP055435027, Gefunden im Internet: URL: <a href="http://onlinelibrary.wiley.com/store/10.1002/1615-9861(200210)2:10<1365::AID-PROT1365>3.0.CO;2-9/asset/1365ftp.pdf?v=1&t=jb5dxuss&s=bc88802156716333f237a20bd4e028cd99b57b3a">http://onlinelibrary.wiley.com/store/10.1002/1615-9861(200210)2:10<1365::AID-PROT1365>3.0.CO;2-9/asset/1365ftp.pdf?v=1&t=jb5dxuss&s=bc88802156716333f237a20bd4e028cd99b57b3a [gefunden am 2017-12-13] das ganze Dokument	1-19

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2017/001131

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 2012232805	A1	13-09-2012	KEINE

WO 2007076606	A1	12-07-2007	CA 2629203 A1 12-07-2007
			EP 1969615 A1 17-09-2008
			JP 5252205 B2 31-07-2013
			JP 2009522558 A 11-06-2009
			US 2007164207 A1 19-07-2007
			WO 2007076606 A1 12-07-2007

WO 2014130627	A1	28-08-2014	US 2016003842 A1 07-01-2016
			WO 2014130627 A1 28-08-2014

US 2007038387	A1	15-02-2007	KEINE
