



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 699 12 444 T3** 2010.05.06

(12) **Übersetzung der geänderten europäischen Patentschrift**

(97) **EP 1 105 517 B2**

(21) Deutsches Aktenzeichen: **699 12 444.1**

(86) PCT-Aktenzeichen: **PCT/US99/19415**

(96) Europäisches Aktenzeichen: **99 943 915.1**

(87) PCT-Veröffentlichungs-Nr.: **WO 2000/011208**

(86) PCT-Anmeldetag: **25.08.1999**

(87) Veröffentlichungstag
der PCT-Anmeldung: **02.03.2000**

(97) Erstveröffentlichung durch das EPA: **13.06.2001**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **29.10.2003**

(97) Veröffentlichungstag
des geänderten Patents beim EPA: **30.09.2009**

(47) Veröffentlichungstag im Patentblatt: **06.05.2010**

(51) Int Cl.⁸: **C12Q 1/00** (2006.01)

G01N 33/573 (2006.01)

G01N 33/53 (2006.01)

G01N 33/567 (2006.01)

G01N 24/00 (2006.01)

G01N 33/68 (2006.01)

Patentschrift wurde im Einspruchsverfahren geändert

(30) Unionspriorität:

97788 P **25.08.1998** **US**

99113 P **03.09.1998** **US**

(73) Patentinhaber:

University of Washington, Seattle, Wash., US

(74) Vertreter:

**Müller-Boré & Partner, Patentanwälte, European
Patent Attorneys, 81671 München**

(84) Benannte Vertragsstaaten:

**AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT,
LI, LU, MC, NL, PT, SE**

(72) Erfinder:

**AEBERSOLD, Hans, Rudolf, Mercer Island, US;
GELB, Michael H., Seattle, US; GYGI, P., Steven,
Seattle, US; SCOTT, Ronald, C., Seattle, US;
TURECEK, Frantisek, Seattle, US; GERBER, A.,
Scott, Seattle, US; RIST, Beate, Seattle, US**

(54) Bezeichnung: **SCHNELLE QUANTITATIVE ANALYSE VON PROTEINEN ODER PROTEINFUNKTIONEN IN KOMPLEXEN GEMISCHEN**

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

Hintergrund der Erfindung

[0001] Genomtechniken rückten zu einem Punkt vor, bei dem es prinzipiell möglich geworden ist, vollständige Genomsequenzen zu bestimmen und die mRNA-Niveaus für jedes Gen, das in einer Zelle exprimiert wird, quantitativ zu messen. Für einige Spezies ist die vollständige Genomsequenz nun bestimmt worden, und für einen Stamm der Hefe *Saccharomyces cerevisiae* sind die mRNA-Niveaus für jedes exprimierte Gen vorher unter verschiedenen Wachstumsbedingungen (Velculescu et al., 1997) quantifiziert worden. Vergleichende cDNA-Anordnungsanalyse und verwandte Technologien sind verwendet worden, um die verursachten Veränderungen bei der Genexpression bei dem mRNA-Niveau durch gleichzeitiges Beobachten des Expressionsniveaus einer großen Anzahl an Genen (in einigen Fällen alle Gene), die durch die untersuchte Zelle oder das Gewebe (Shalon et al., 1996) exprimiert werden, zu bestimmen. Außerdem sind biologische und Computertechniken verwendet worden, um die spezifische Funktion mit den Gensequenzen zu korrelieren. Die Interpretation der Daten, die durch diese Techniken im Rahmen der Struktur, Kontrolle und des Mechanismus der biologischen Systeme erhalten wurden, ist als beträchtliche Herausforderung erkannt worden. Insbesondere ist es sehr schwierig geworden, den Mechanismus von biologischen Verfahren durch Genomanalyse allein zu erklären.

[0002] Proteine sind wesentlich für die Kontrolle und Durchführung von so gut wie jedem biologischen Verfahren. Die Geschwindigkeit der Synthese und die Halbwertszeit der Proteine und daher ihr Expressionsniveau werden ebenso post-transkriptionell kontrolliert. Außerdem wird die Aktivität von Proteinen häufig durch post-translatorische Modifikationen, insbesondere Proteinphosphorylierung, und in Abhängigkeit der Assoziation des Proteins mit anderen Molekülen, einschließlich DNA und Proteine, moduliert. Weder das Niveau der Expression noch der Zustand der Aktivität der Proteine ist daher direkt aus der Gensequenz oder sogar des Expressionsniveaus des entsprechenden mRNA-Transkripts offensichtlich. Es ist daher wesentlich, daß eine vollständige Beschreibung eines biologischen Systems Messungen umfaßt, die die Identität, Menge und den Zustand der Aktivität der Proteine, die das System bilden, angeben. Die Großmaßstabs-Analyse (allumfassend) von Proteinen, die in einer Zelle oder Gewebe exprimiert werden, ist als Proteom-Analyse (Pennington et al., 1997) bezeichnet worden.

[0003] Gegenwärtig erreicht keine Proteinanalysetechnologie den Durchsatz und das Niveau der Automatisierung der Genomtechnologie. Die bekannteste Durchführung der Proteom-Analyse basiert auf der Trennung von Komplexproteinproben am bekanntesten durch zweidimensionale Gelelektrophorese (2DE) und die anschließende sequentielle Identifikation der getrennten Proteinspezies (Ducret et al., 1998; Garrels et al., 1997; Link et al., 1997; Shevchenko et al., 1996; Gygi et al., 1999; Boucherie et al., 1996). Dieser Ansatz ist durch die Entwicklung von starken massenspektrometrischen Techniken und die Entwicklung von Computeralgorithmen revolutioniert worden, die Protein- und Peptidmassenspektraldaten mit Sequenzdatenbanken in Beziehung setzen, und daher Proteine schnell und eindeutig identifizieren (Eng et al., 1994; Mann und Wilm, 1994; Yates et al., 1995). Diese Technologie erreichte ein Niveau an Empfindlichkeit, das nun die Identifikation von im wesentlichen jedem Protein ermöglicht, das durch konventionelle Proteinfärbeverfahren, einschließlich Silberfärben, nachweisbar ist (Figeys und Aebersold, 1998; Figeys et al., 1996; Figeys et al., 1997; Shevchenko et al., 1996). Jedoch begrenzt die sequentielle Weise, bei der Proben verarbeitet werden, den Probendurchsatz, die empfindlichsten Verfahren sind schwierig zu automatisieren gewesen, und Proteine mit geringer Häufigkeit, wie Regulatorproteine, entziehen sich dem Nachweis ohne vorherige Anreicherung, wodurch der dynamische Bereich der Technik effektiv eingeschränkt wird. In dem 2DE/(MS)N-Verfahren werden die Proteine durch Densitometrie von gefärbten Flecken in den 2DE-Gelen quantifiziert.

[0004] Die Entwicklung der Verfahren und Instrumentierung für automatisierte, datenabhängige Elektrosprühionisierungs-(ESI)-Tandem-Massenspektrometrie (MS^n) zusammen mit Mikrokapillarflüssigchromatographie (μ LC) und Datenbankforschung hat die Empfindlichkeit und Geschwindigkeit der Identifizierung von getrennten Proteinen bedeutend erhöht. Als Alternative zu dem 2DE/ MS^n -Ansatz der Proteom-Analyse ist die direkte Analyse durch Tandem-Massenspektrometrie von Peptidgemischen, die durch die Verdauung von Komplexproteinmischungen erzeugt werden, vorgeschlagen worden (Dongre et al., 1997). μ LC- MS/MS ist ebenso erfolgreich für die Großmaßstabs-Identifizierung von einzelnen Proteinen direkt aus den Gemischen ohne gelelektrophoretische Trennung (Link et al., 1999; Opitke et al., 1997) verwendet worden. Während diese Ansätze die Proteinidentifizierung dramatisch beschleunigen, können die Mengen der analysierten Proteine nicht leicht bestimmt werden, und diese Verfahren konnten das Problem des dynamischen Bereichs, das ebenso auf den 2DE/ MS/MS -Ansatz zutrifft, nicht wesentlich mindern. Deshalb sind Proteine mit geringer Häufigkeit in Komplexproben ebenso schwierig durch das μ LC/ MS/MS -Verfahren ohne deren vorherige Anreicherung zu

analysieren.

[0005] Es ist daher offensichtlich, daß gegenwärtige Technologien, obwohl sie geeignet sind, die Komponenten von Proteingemischen zu identifizieren, weder zum Messen der Menge noch des Zustands an Aktivität des Proteins in einem Gemisch fähig sind. Selbst evolutionäre Verbesserungen der gegenwärtigen Ansätze werden wahrscheinlich keine ausreichenden Fortschritte machen, die routinemäßige quantitative und funktionelle Proteom-Analyse in die Realität umzusetzen.

[0006] Diese Erfindung stellt Verfahren bereit, die bei der Proteom-Analyse eingesetzt werden können, wobei die in traditionellen Techniken inhärenten Einschränkungen überwunden werden. Der beschriebene grundlegende Ansatz kann zur quantitativen Analyse der Proteinexpression in Komplexproben (wie Zellen, Geweben und Fraktionen davon), zum Nachweis und zur Quantifizierung von speziellen Proteinen in Komplexproben eingesetzt werden.

[0007] In dieser Hinsicht ist eine Vielzahl an analytischen Techniken für klinische und diagnostische Assays derzeit erhältlich, die die Gegenwart, Abwesenheit, den Mangel oder Überschuß eines Proteins oder einer Proteinfunktion, in Verbindung mit einem normalen oder einem Krankheitszustand, nachweisen. Während diese Techniken ziemlich empfindlich sind, stellen sie nicht zwangsläufig eine chemische Speziation von Produkten bereit, und können infolgedessen schwer für das gleichzeitige Analysieren von mehreren Proteinen oder Enzymen in einer einzelnen Probe verwendet werden. Gegenwärtige Verfahren können nicht zwischen abweichender Expression von unterschiedlichen Enzymen oder deren Fehlfunktionen unterscheiden, die zu einer bekannten Gruppe von klinischen Symptomen führen. Die Verfahren hierin können in klinischen und diagnostischen Assays zum gleichzeitigen (multiplex) Überwachen von multiplen Proteinen eingesetzt werden.

[0008] US 5,650,270 offenbart Markierungsreagenzien, die als "Freisetzungsmarkierungen" bezeichnet werden. Diese Freisetzungsmarkierungen sind dreiteilige Moleküle, die durch die generische Formel Sg-Re-Rx dargestellt werden können, in der Sg eine "Signalgruppe" darstellt, Rx eine "Reaktivitätsgruppe" darstellt, enthaltend eine funktionelle Gruppe, die mit einer zu markierenden Substanz reagiert, wodurch die Freisetzungsmarkierung angelagert wird, und Re eine "Freisetzungsgruppe" darstellt, bei der die Spaltung bei einer entsprechenden Zeit und unter entsprechenden Bedingungen auftreten kann, um die Signalgruppe Sg in einer Form, die zur Bestimmung geeignet ist, freizusetzen.

Zusammenfassung der Erfindung

[0009] Diese Erfindung stellt Massenspektrometrie-basierende Verfahren zur schnellen und quantitativen Analyse von Proteinen in Gemischen von Proteinen bereit. Das analytische Verfahren kann zur qualitativen und insbesondere zur quantitativen Analyse von globalen Proteinexpressionsprofilen in Zellen und Geweben verwendet werden, d. h. die quantitative Analyse von Proteomen. Das Verfahren kann ebenso zum Screenen und Identifizieren von Proteinen eingesetzt werden, deren Expressionsniveau in Zellen, Gewebe oder biologischen Flüssigkeiten durch einen Stimulus (beispielsweise Verabreichung eines Arzneimittels oder Kontakt mit einem möglicherweise toxischen Material), durch eine Veränderung der Umwelt (beispielsweise Nährstoffgehalt, Temperatur, Verlauf der Zeit) oder durch eine Veränderung im Zustand oder Zellzustand (beispielsweise Krankheitszustand, Malignität, ortsgerichtete Mutation, gezielte Genzerstörung) der Zelle, des Gewebes oder Organismus, aus dem die Probe stammt, beeinflusst wird. Die Proteine, die in einem derartigen Screen identifiziert werden, können als Marker für den veränderten Zustand fungieren. Beispielsweise können die Vergleiche von Proteinexpressionsprofilen von normalen und malignen Zellen zur Identifizierung von Proteinen führen, deren Gegenwart oder Abwesenheit charakteristisch und diagnostisch für die Malignität sind.

[0010] In einer beispielhaften Ausführungsform können die Verfahren hierin eingesetzt werden, um hinsichtlich Veränderungen bei der Expression von speziellen Proteinen zu screenen. Diese Veränderungen können durch eine Vielzahl an Chemikalien, einschließlich pharmazeutische Agonisten oder Antagonisten oder möglicherweise schädlichen oder toxischen Materialien, hervorgerufen werden. Das Wissen derartiger Veränderungen kann zum Diagnostizieren Enzym-basierender Krankheiten und zum Untersuchen des komplexen, regulativen Netzwerkes in Zellen nützlich sein.

[0011] Die Verfahren hierin können ebenso zum Durchführen einer Vielzahl an klinischen und diagnostischen Analysen verwendet werden, um die Gegenwart, Abwesenheit, den Mangel oder Überschuß eines gegebenen Proteins in einer biologischen Flüssigkeit (beispielsweise Blut) oder in Zellen oder Gewebe nachzuweisen. Das Verfahren ist bei der Analyse von Komplexgemischen aus Proteinen, d. h. denen, die 5 oder mehr unterschiedliche Proteine oder Proteinfunktionen enthalten, besonders nützlich.

[0012] Das erfindungsgemäße Verfahren setzt affinitätsmarkierte reaktive Proteinreagenzien ein, die die selektive Isolation von Peptidfragmenten aus den komplexen Gemischen ermöglicht. Die isolierten Peptidfragmente sind für die Gegenwart eines Proteins charakteristisch in solchen Gemischen. Isolierte Peptide werden durch Massenspektrometrie-(MS-)techniken charakterisiert. Insbesondere kann die Sequenz von isolierten Peptiden unter Verwendung der Tandem-MS-(MSⁿ-)techniken bestimmt werden, und durch die Anwendung von Sequenzdatenbankforschungstechniken kann das Protein, aus dem das sequenzierte Peptid stammt, identifiziert werden. Die Reagenzien stellen ebenso unterschiedliches Isotopenmarkieren von isolierten Peptiden bereit, die die quantitative Bestimmung durch Massenspektrometrie der relativen Mengen an Proteinen in unterschiedlichen Proben erleichtert. Ebenso erleichtert die Verwendung von unterschiedlich isotopenmarkierten Reagenzien als interne Standards die quantitative Bestimmung der absoluten Mengen von einem oder mehreren Proteinen, die in der Probe vorliegen.

[0013] Die affinitätsmarkierten reaktiven Proteinreagenzien dieser Erfindung weisen drei Teile auf: eine Affinitätsmarkierung (A), die an eine reaktive Proteingruppe (PRG) durch eine Linkergruppe (L) kovalent gebunden ist:

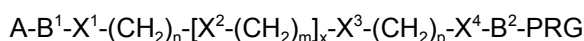
A-L-PRG

[0014] Der Linker kann unterschiedlich isotopenmarkiert werden, beispielsweise durch Substitution von einem oder mehreren Atomen in dem Linker mit einem stabilen Isotop davon. Beispielsweise können Wasserstoffe mit Deuteriums oder C¹² mit C¹³ substituiert werden.

[0015] Die Affinitätsmarkierung A fungiert als molekularer Griff, welcher selektiv an ein Abfangreagenz (CR) kovalent oder nicht-kovalent bindet. Das Binden an CR erleichtert die Isolation von Peptiden, Substraten oder Reaktionsprodukten, die mit A gekennzeichnet oder markiert wurden. In speziellen Ausführungsformen ist A Streptavidin oder Avidin. Nach der Affinitätsisolierung von affinitätsmarkierten Materialien, wobei einige von denen isotopenmarkiert werden können, wird die Wechselwirkung zwischen A und dem Abfangreagenz unterbrochen oder gestört, um die MS-Analyse der isolierten Materialien zu ermöglichen. Die Affinitätsmarkierung kann durch das Abfangreagenz durch Addition von Verdrängungsliganden, die freies A oder ein Derivat von A sein können, oder durch Ändern der Lösungsmittel- (beispielsweise Lösungsmitteltyp oder pH) oder Temperaturbedingungen verdrängt werden, oder der Linker kann chemisch, enzymatisch, thermisch oder photochemisch gespalten werden, um die isolierten Materialien für die MS-Analyse freizusetzen.

[0016] Die hierin bereitgestellten PRG-Gruppen reagieren selektiv mit einer funktionellen Proteingruppe, um eine kovalente oder nicht-kovalente Bindung zu bilden, die das Protein an speziellen Stellen markiert. In speziellen Ausführungsformen ist PRG eine Gruppe mit spezifischer Reaktivität für bestimmte Proteingruppen, wie die Spezifität für Sulfhydrylgruppen, und ist im allgemeinen zum selektiven Markieren von Proteinen in komplexen Gemischen nützlich. Ein spezielles Sulfhydrylreagenz markiert Proteine, die Cystein enthalten.

[0017] Reagenzien in Übereinstimmung mit der Erfindung weisen die allgemeine Formel auf:



worin:

A eine Affinitätsmarkierung ist;

PRG die reaktive Proteingruppe ist;

X¹, X², X³ und X⁴ unabhängig voneinander und X² unabhängig von anderem X² in der Linkergruppe aus O, S, NH, NR, NRR⁺, CO, COO, COS, S-S, SO, SO₂, CO-NR', CS-NR', Si-O, Aryl- oder Diarylgruppen ausgewählt sein können, oder X¹ bis X⁴ nicht vorhanden sein können, aber vorzugsweise mindestens eines von X¹ bis X⁴ vorliegt;

B¹ und B² unabhängig voneinander optionale Komponenten sind, die das Binden der A- oder PRG-Gruppe an den Linker erleichtern oder unerwünschte Spaltung von diesen Gruppen aus dem Linker verhindern können, und beispielsweise aus COO, CO, CO-NR', CS-NR' ausgewählt werden können und eine oder mehrere CH₂-Gruppen allein oder in Kombination mit anderen Gruppen, beispielsweise (CH₂)_q-CONR', (CH₂)_q-CS-NR' oder (CH₂)_q, enthalten können;

n, m, p und q ganze Zahlen sind, die Werte von 0 bis etwa 100 aufweisen können, wobei vorzugsweise eines von n, m, p oder q nicht 0 ist, und x ebenso eine ganze Zahl ist, die zwischen 0 und 100 liegen kann, wobei die Summe von n + xm + p + q vorzugsweise weniger als etwa 100 und stärker bevorzugt weniger als etwa 20 beträgt;

R eine Alkyl-, Alkenyl-, Alkynyl-, Alkoxy- oder Arylgruppe ist; und

R' ein Wasserstoffatom, eine Alkyl-, Alkenyl-, Alkynyl-, Alkoxy- oder Arylgruppe ist.

[0018] Eine oder mehrere der CH₂-Gruppen des Linkers können gegebenenfalls mit kleinen C₁-C₆-Alkyl-, -Alkenyl- oder -Alkoxygruppen, einer Arylgruppe substituiert werden oder können mit funktionellen Gruppen, die die Ionisierung fördern, wie sauren oder basischen Gruppen, welche eine bleibende positive oder negative Ladung tragen, substituiert werden. Eine oder mehrere Einzelbindungen, die die CH₂-Gruppen in dem Linker verbinden, können durch eine Doppel- oder eine Dreifachbindung ersetzt werden. Bevorzugte R- und R'-Alkyl-, Alkenyl-, Alkynyl- oder Alkoxygruppen sind klein mit 1 bis etwa 6 Kohlenstoffatomen.

[0019] Ein oder mehrere der Atome in dem Linker können mit einem stabilen Isotop substituiert werden, um ein oder mehrere im wesentlichen chemisch identische, aber isotopen-unterscheidbare Reagenzien zu erzeugen. Beispielsweise können ein oder mehrere Wasserstoffatome in dem Linker mit Deuterium substituiert werden, um isotopenschwere Reagenzien zu erzeugen.

[0020] In einer beispielhaften Ausführungsform enthält der Linker Gruppen, die gespalten werden können, um die Affinitätsmarkierung zu entfernen. Wenn eine spaltbare Linkergruppe eingesetzt wird, wird sie typischerweise gespalten, nachdem affinitätsmarkierte Peptide, Substrate oder Reaktionsprodukte unter Verwendung der Affinitätsmarkierung zusammen mit dem CR isoliert worden sind. In diesem Fall bleibt jede Isotopenmarkierung in dem Linker vorzugsweise an das Protein, Peptid, Substrat oder Reaktionsprodukt gebunden.

[0021] Linkergruppen umfassen unter anderem: Ether, Polyether, Etherdiamine, Polyetherdiamine, Diamine, Amide, Polyamide, Polythioether, Disulfide, Silylether, Alkyl- oder Alkenylketten (geradkettig oder verzweigt, und Teile, die cyclisch sein können), Aryl-, Diaryl- oder Alkyl-Aryl-Gruppen. Arylgruppen in Linkern können ein oder mehrere Heteroatome (beispielsweise N-, O- oder S-Atome) enthalten.

[0022] Quantitative relative Mengen an Proteinen in unterschiedlichen Proben, die Proteingemische (beispielsweise biologische Flüssigkeiten, Zell- oder Gewebelysate, usw.) enthalten, können unter Verwendung chemisch identischer, affinitätsmarkierter und unterschiedlich isotopenmarkierter Reagenzien bestimmt werden, um die Proteine in den unterschiedlichen Proben affinitätszumarkieren und unterschiedlich isotopenzu-markieren. In diesem Verfahren wird jede zu vergleichende Probe mit einem unterschiedlich isotopenmarkierten Reagenz behandelt, um bestimmte Proteine darin mit der Affinitätsmarkierung zu markieren. Die behandelten Proben werden dann vorzugsweise in gleichen Mengen kombiniert, und die Proteine in der kombinierten Probe werden enzymatisch aufgeschlossen, wenn notwendig, um Peptide zu erzeugen. Einige der Peptide werden affinitätsmarkiert und außerdem werden markierte Peptide, die aus unterschiedlichen Proben stammen, unterschiedlich isotopenmarkiert. Wie oben beschrieben, werden affinitätsmarkierte Peptide isoliert, aus dem Abfangreagenz freigesetzt und durch (LC/MS) analysiert. Peptide, die für ihren Proteinursprung charakteristisch sind, werden unter Verwendung von MSⁿ-Techniken sequenziert, was die Identifizierung von Proteinen in den Proben ermöglicht. Die relativen Mengen eines gegebenen Proteins in jeder Probe werden durch Vergleichen der relativen Häufigkeit der Ionen, die aus unterschiedlich markierten Peptiden, aus denen dieses Protein stammt, erzeugt werden, bestimmt. Das Verfahren kann verwendet werden, um die relativen Mengen an bekannten Proteinen in unterschiedlichen Proben einzuschätzen. Da außerdem das Verfahren kein vorheriges Wissen über den Typ an Proteinen, die in den Proben vorliegen können, benötigt, kann es verwendet werden, um Proteine zu identifizieren, die bei unterschiedlichen Niveaus in den überprüften Proben vorliegen. Insbesondere kann das Verfahren angewendet werden, um Proteine zu screenen und zu identifizieren, die unterschiedliche Expression in Zellen, Gewebe oder biologischen Flüssigkeiten aufweisen.

[0023] Es ist ebenso möglich, die Niveaus an spezifischen Proteinen in mehrfachen Proben in einer Einzelanalyse (Multiplexing) zu quantifizieren. In diesem Fall können die Affinitätsmarkierungsreagenzien, die verwendet werden, um Proteine zu derivatisieren, die in unterschiedlichen affinitätsmarkierten Peptiden aus unterschiedlichen Proben vorliegen, selektiv durch Massenspektrometrie quantifiziert werden.

[0024] In diesem Aspekt der Erfindung stellt das Verfahren die quantitative Messung von spezifischen Proteinen in biologischen Flüssigkeiten, Zellen oder Geweben bereit, und kann angewendet werden, um die globalen Proteinexpressionsprofile in unterschiedlichen Zellen und Geweben zu bestimmen. Dieselbe allgemeine Strategie kann erweitert werden, um die Proteom-breite, qualitative und quantitative Analyse des Zustands der Modifikation von Proteinen durch Einsetzen von Affinitätsreagenzien mit unterschiedlichen Spezifitäten für die Reaktion mit Proteinen zu erreichen. Das Verfahren und die Reagenzien dieser Erfindung können verwendet werden, um Proteine mit geringer Häufigkeit in komplexen Gemischen zu identifizieren, und können verwendet werden, um spezielle Gruppen oder Klassen an Proteinen, wie Membran- oder Zelloberflächenproteine, oder Proteine, die innerhalb von Organellen, sub-Zellfraktionen oder biochemischen Fraktionen, wie Immunfällun-

gen, enthalten sind, selektiv zu analysieren. Außerdem können diese Verfahren angewendet werden, um die Unterschiede in exprimierten Proteinen in unterschiedlichen Zellzuständen zu analysieren. Beispielsweise können die Verfahren und Reagenzien hierin in diagnostischen Assays zum Nachweis der Gegenwart oder Abwesenheit von ein oder mehreren Proteinen, die für einen Krankheitszustand, wie Krebs, bezeichnend sind, eingesetzt werden.

Ausführliche Beschreibung der Erfindung

[0025] Die erfindungsgemäßen Verfahren setzen affinitätsmarkierte reaktive Proteinreagenzien ein, bei denen die Affinitätsmarkierung an eine reaktive Proteingruppe durch einen Linker kovalent gebunden wird. Der Linker ist isotoptenmarkiert, um Paare oder Gruppen von Reagenzien zu erzeugen, die im wesentlichen chemisch identisch sind, aber die durch Masse unterscheidbar sind. Beispielsweise können irgendeines oder mehrere der Wasserstoff-, Stickstoff-, Sauerstoff- oder Schwefelatome in dem Linker durch deren isotop stabile Isotope: ^2H , ^{13}C , ^{15}N , ^{17}O , ^{18}O oder ^{34}S ersetzt werden.

[0026] Geeignete Affinitätsmarkierungen binden selektiv entweder kovalent oder nicht-kovalent und mit hoher Affinität an ein Abfangreagenz (CR). Die CR-A-Wechselwirkung oder -Bindung sollte nach ausgedehnten und mehrfachen Waschungen mit einer Vielzahl an Lösungen intakt bleiben, um nicht-spezifisch gebundene Komponenten zu entfernen. Die Affinitätsmarkierung bindet minimal oder vorzugsweise überhaupt nicht an die Komponenten in dem Assaysystem, außer CR, und bindet nicht signifikant an die Oberflächen der Reaktionsbehälter. Jede nicht-spezifische Wechselwirkung der Affinitätsmarkierung mit anderen Komponenten oder Oberflächen sollte durch mehrfaches Waschen, das CR-A intakt läßt, unterbrochen werden. Außerdem muß es möglich sein, die Wechselwirkung von A und CR zu unterbrechen, um Peptide, Substrate oder Reaktionsprodukte beispielsweise durch Zugabe eines Verdrängungsliganden oder durch Ändern der Temperatur- oder Lösungsmittelbedingungen freizusetzen. Vorzugsweise reagiert weder CR noch A mit anderen Komponenten in dem Assaysystem chemisch, und beide Gruppen sollten über den Zeitraum eines Assays oder Experiments chemisch stabil sein. Die Affinitätsmarkierung unterliegt vorzugsweise nicht der peptidähnlichen Fragmentierung während der (MS)ⁿ-Analyse. Die Affinitätsmarkierung ist vorzugsweise in der zu analysierenden Probenflüssigkeit löslich, und das CR sollte in der Probenflüssigkeit, auch wenn es an ein unlösliches Harz, wie Agarose, gebunden wird, löslich bleiben. Bei CR bedeutet der Ausdruck löslich, daß CR ausreichend hydratisiert oder anderweitig solvatisiert wird, so daß es zum Binden an A richtig fungiert. CR oder CR-enthaltende Konjugate sollten nicht in der zu analysierenden Probe vorliegen, außer wenn sie zugegeben werden, um A einzufangen.

[0027] Beispiele von A- und CR-Paaren umfassen:

d-Biotin oder strukturell modifizierte Biotin-basierende Reagenzien, einschließlich d-Iminobiotin, das an Proteine von Avidin/Streptavidin bindet, das beispielsweise in den Formen Streptavidin-Agarose, oligomeres-Avidin-Agarose oder monomeres Avidin-Agarose verwendet werden kann;
irgendein 1,2-Diol, wie 1,2-Dihydroxyethan ($\text{HO}-\text{CH}_2-\text{CH}_2-\text{OH}$), und andere 1,2-Dihydroxyalkane, einschließlich denen der cyclischen Alkane, beispielsweise 1,2-Dihydroxycyclohexan, das an eine Alkyl- oder Arylborsäure oder -Borsäureester bindet, wie Phenyl-B(OH)₂ oder Hexyl-B(OEthyl)₂, die mittels der Alkyl- oder Arylgruppe an ein festes Trägermaterial, wie Agarose, gebunden werden können;
Maltose, das an Maltose-bindendes Protein bindet (sowie irgendein anderes Zucker/Zucker-bindendes Proteinpaar oder allgemeiner an Ligand/Ligand-bindende Proteinpaare, die die oben erläuterten Eigenschaften aufweisen);
ein Hapten, wie Dinitrophenylgruppe, für jeden Antikörper, wo das Hapten an einen Anti-Hapten-Antikörper bindet, der das Hapten erkennt, beispielsweise wird die Dinitrophenylgruppe an ein Anti-Dinitrophenyl-IgG binden;
einen Liganden, der an ein Übergangsmetall bindet, beispielsweise wird ein Oligomerhistidin an Ni(II) binden, das Übergangsmetall-CR kann in Form eines harzgebundenen, chelatisierten Übergangsmetalls, wie Nitrilotriessigsäure-chelatisiertes Ni(II) oder Iminodiessigsäure-chelatisiertes Ni(II), verwendet werden;
Glutathion, das an Glutathion-S-Transferase bindet.

[0028] Im allgemeinen wird irgendein A-CR-Paar zur Affinitätsanreicherung, die den oben erläuterten Erkennungskriterien entspricht, allgemein verwendet. Biotin und Biotin-basierende Affinitätsmarkierungen werden bevorzugt. Von besonderem Interesse sind strukturell modifizierte Biotine, wie d-Iminobiotin, die aus Avidin- oder Streptavidinsäulen unter Lösungsmittelebedingungen, kompatibel mit ESI-MS-Analyse, wie verdünnte Säuren, enthaltend 10 bis 20% organisches Lösungsmittel, eluiert werden. Es wird erwartet, daß d-Iminobiotin-markierte Verbindungen in Lösungsmitteln unter pH 4 eluieren werden. d-Iminobiotin-markierte reaktive Proteinreagenzien können durch Verfahren, die für die entsprechenden Biotin-markierten Reagenzien hierin beschrieben werden, synthetisiert werden.

[0029] Ein Verdrängungsligand, DL, wird gegebenenfalls verwendet, um A durch CR zu ersetzen. Geeignete DLs liegen typischerweise nicht in den Proben vor, es sei denn sie werden zugegeben. DL sollte in der zu analysierenden Probe chemisch und enzymatisch stabil sein, und sollte nicht mit den Komponenten (anders als CR) in den Proben reagieren oder daran binden, oder nicht-spezifisch an die Wände des Reaktionsbehälters binden. DL unterliegt vorzugsweise nicht der peptidähnlichen Fragmentierung während der MS-Analyse, und seine Gegenwart in einer Probe sollte die Ionisierung von markierten Peptid-, Substrat- oder Reaktionsprodukt-konjugaten nicht signifikant unterdrücken.

[0030] DL selbst wird vorzugsweise während der massenspektrometrischen Analyse minimal ionisiert, und die Bildung von Ionen, die aus DL-Clustern bestehen, ist vorzugsweise minimal. Die Wahl von DL hängt von den A- und CR-Gruppen ab, die eingesetzt werden. Im allgemeinen wird DL gewählt, um A durch CR in einer vernünftigen Zeitspanne, höchstens innerhalb einer Woche seiner Zugabe, aber stärker bevorzugt innerhalb weniger Minuten oder bis zu einer Stunde, zu ersetzen. Die Affinität von DL für CR sollte vergleichbar oder stärker als die Affinität der markierten Verbindungen, die A enthalten, für CR sein. Außerdem sollte DL in dem Lösungsmittel, das während der Elution von markierten Verbindungen, die A enthalten, aus CR, verwendet wird, löslich sein. DL ist vorzugsweise freies A oder ein Derivat oder strukturelle Modifikation von A. Beispiele von DL umfassen d-Biotin oder d-Biotin-Derivate, insbesondere die, die Gruppen enthalten, die die Clusterbildung unterdrücken oder die Ionisierung in der MS unterdrücken.

[0031] Die Linkergruppe (L) sollte in der zu analysierenden Probenflüssigkeit löslich sein, und sie sollte in bezug auf die chemische Reaktion, beispielsweise im wesentlichen chemisch inert, mit Komponenten der Proben sowie den A- und den CR-Gruppen stabil sein. Der Linker sollte, wenn er an A gebunden ist, nicht mit der spezifischen Wechselwirkung von A mit CR interferieren oder mit der Verdrängung von A aus CR durch einen Verdrängungsliganden oder durch eine Veränderung der Temperatur oder des Lösungsmittels interferieren. Der Linker sollte minimal oder vorzugsweise überhaupt nicht an andere Komponenten in dem System, an Reaktionsbehälteroberflächen oder CR binden. Jede der nicht-spezifischen Wechselwirkungen des Linkers sollte nach mehrfachen Waschungen, die den A-CR-Komplex intakt lassen, unterbrochen werden. Die Linker unterliegen vorzugsweise nicht der peptidähnlichen Fragmentierung während der (MS)ⁿ-Analyse. Höchstens einige der Atome in den Linkergruppen sollten ohne weiteres durch stabile Schweratomisotope ersetzbar sein. Der Linker enthält vorzugsweise Gruppen oder Komponenten, die die Ionisierung der affinitätsmarkierten Reagenzien, Peptide, Substrate oder Reaktionsprodukte erleichtern.

[0032] Um die Ionisierung zu fördern, kann der Linker saure oder basische Gruppen, beispielsweise COOH, SO₃H, primäre, sekundäre oder tertiäre Aminogruppen, Stickstoff-heterocyclen, Ether oder Kombinationen von diesen Gruppen enthalten. Der Linker kann ebenso Gruppen mit einer bleibenden Ladung, beispielsweise Phosphoniumgruppen, quartäre Ammoniumgruppen, Sulfoniumgruppen, chelatisierte Metallionen, Tetralkyl- oder Tetrarylborat oder stabile Carbanionen, enthalten.

[0033] Die kovalente Bindung des Linkers an A oder PRG sollte typischerweise nicht vorsätzlich durch chemische oder enzymatische Reaktionen während des Assays gespalten werden. In einigen Fällen kann es wünschenswert sein, den Linker aus der Affinitätsmarkierung A oder aus dem PRG abzuspalten, um die Freisetzung aus einer Affinitätssäule zu erleichtern. Daher kann der Linker beispielsweise durch chemische, thermische oder photochemische Reaktion spaltbar sein. Photospaltbare Gruppen in dem Linker können die 1-(2-Nitrophenyl)-ethylgruppe umfassen. Thermisch labile Linker können beispielsweise ein doppelsträngiger Duplex sein, der aus zwei komplementären Strängen aus Nukleinsäure, einem Strang einer Nukleinsäure mit einem komplementären Strang einer Peptidnukleinsäure oder zwei komplementären Peptidnukleinsäuresträngen, die sich beim Erwärmen aufspalten werden, gebildet wird. Spaltbare Linker umfassen ebenso die mit Disulfidbindungen, sauren oder basischen labilen Gruppen, einschließlich unter anderem Diarylmethyl- oder Trime-thylarylmethylgruppen, Silylether, Carbamate, Oxyester, Thioester, Thionoester und α -fluorierte Amide und Ester. Enzymatisch spaltbare Linker können beispielsweise Protease-empfindliche Amide oder Ester, β -Lactamase-empfindliche β -Lactamanaloga und Linker enthalten, die Nuklease-spaltbar oder Glycosidase-spaltbar sind.

[0034] Die reaktive Proteingruppe (PRG) ist eine Gruppe, die mit bestimmten funktionellen Proteingruppen selektiv reagiert. Jede selektiv reaktive Proteingruppe sollte mit einer funktionellen Gruppe von Interesse, die in mindestens einem Teil der Proteine in einer Probe vorliegt, reagieren. Die Reaktion von PRG mit funktionellen Gruppen an dem Protein sollte unter Bedingungen stattfinden, die nicht zum wesentlichen Abbau der Verbindungen in der zu analysierenden Probe führen. Beispiele von selektiv reaktiven PRGs, die zur Verwendung in affinitätsmarkierten Reagenzien dieser Erfindung geeignet sind, umfassen die, die mit Sulfhydrylgruppen reagieren, um Proteine, die Cystein enthalten, zu markieren, die, die mit Aminogruppen, Carboxylatgruppen, Es-

tergruppen, reaktiven Phosphatgruppen und reaktiven Aldehyd- und/oder Ketongruppen oder nach der Fragmentierung mit CNBr mit Homoserinlaktat reagieren.

[0035] Reaktive Thiolgruppen umfassen Epoxide, α -Halogenacylgruppen, Nitrile, sulfonierte Alkyl- oder Arylthiole und Maleimide. Reaktive Aminogruppen markieren Aminogruppen in Proteinen und umfassen Sulfonylhalogenide, Isocyanate, Isothiocyanate, aktive Ester, einschließlich Tetrafluorphenylester, und N-Hydroxysuccinimidylester, Säurehalogenide und Säureanhydride. Außerdem umfassen reaktive Aminogruppen Aldehyde oder Ketone in Gegenwart oder Abwesenheit von NaBH_4 oder NaCNBH_3 .

[0036] Reaktive Carbonsäuregruppen umfassen Amine oder Alkohole in Gegenwart eines Kopplungsmittels, wie Dicyclohexylcarbodiimid oder 2,3,5,6-Tetrafluorphenyltrifluoracetat, und in Gegenwart oder Abwesenheit eines Kopplungskatalysators, wie 4-Dimethylaminopyridin; und Übergangsmetall-Diamin-Komplexe, einschließlich Cu(II) phenanthrolin.

[0037] Reaktive Estergruppen umfassen Amine, die beispielsweise mit Homoserinlaktat reagieren.

[0038] Reaktive Phosphatgruppen umfassen chelatisiertes Metall, wobei das Metall, beispielsweise Fe(III) oder Ga(III) , mit beispielsweise Nitrilotriessigsäure oder Iminodiessigsäure chelatisiert wird.

[0039] Reaktive Aldehyd- oder Ketongruppen umfassen Amin plus NaBH_4 oder NaCNBH_3 , oder diese Reagenzien, nachdem zunächst ein Kohlenhydrat mit Periodat behandelt wurde, um ein Aldehyd oder Keton zu erzeugen.

[0040] Interne Standards, die geeignet isotoptenmarkiert werden, können in den erfindungsgemäßen Verfahren eingesetzt werden, um die absoluten quantitativen Mengen an Proteinen in den Proben zu messen. Interne Standards sind von besonderer Verwendung in Assays, die beabsichtigen, die affinitätsmarkierten Produkte von enzymatischen Reaktionen zu markieren. In dieser Anmeldung ist der interne Standard zu dem markierten enzymatischen Produkt chemisch identisch, das durch die Wirkung des Enzyms auf das affinitätsmarkierte Enzymsubstrat erzeugt wurde, aber trägt Isotoptenmarkierungen, die ^2H , ^{13}C , ^{15}N , ^{17}O , ^{18}O oder ^{34}S umfassen können, was es ermöglicht, durch die MS-Techniken unabhängig voneinander nachgewiesen zu werden. Interne Standards zur Verwendung in dem Verfahren hierin, um ein oder mehrere Proteine in einer Probe zu quantifizieren, werden durch die Reaktion von affinitätsmarkierten reaktiven Proteinreagenzien mit einem bekannten Protein hergestellt, um die affinitätsmarkierten Peptide zu erzeugen, die durch den Aufschluß des markierten Proteins gebildet werden. Affinitätsmarkierte Peptide-interne Standards sind im wesentlichen zu den entsprechenden affinitätsmarkierten Peptiden, die durch den Aufschluß des affinitätsmarkierten Proteins gebildet werden, chemisch identisch, außer daß sie unterschiedlich isotoptenmarkiert werden, um ihren unabhängigen Nachweis durch MS-Techniken zu ermöglichen.

[0041] Das erfindungsgemäße Verfahren wird angewendet, um die relativen Mengen von ein oder mehreren Proteinen in zwei oder mehr Proteinproben zu bestimmen, wobei die Proteine in jeder Probe mit affinitätsmarkierenden Reagenzien, die im wesentlichen chemisch identisch, aber unterschiedlich isotoptenmarkiert sind, umgesetzt werden. Die Proben werden kombiniert und als eins verarbeitet. Die relative Menge von jedem markierten Peptid, die die relative Menge des Proteins, aus dem das Peptid stammt, widerspiegelt, wird durch die Messung der jeweiligen Isotoptenpeaks durch Massenspektrometrie bestimmt.

[0042] Die erfindungsgemäßen Verfahren können auf die Analyse oder den Vergleich der mehrfach unterschiedlichen Proben angewendet werden. Proben, die durch die erfindungsgemäßen Verfahren analysiert werden können, umfassen Zellhomogenate; Zellfraktionen; biologische Flüssigkeiten, einschließlich Urin, Blut und Hirn-Rückenmarks-Flüssigkeit; Gewebekomponenten; Tränen; Fäkalien; Speichel; Spülungsflüssigkeit, wie Lungen- oder Peritonealspülungen; Gemische aus biologischen Molekülen, einschließlich Proteine, Lipide, Kohlenhydrate und Nukleinsäuren, die durch teilweise oder vollständige Fraktionierung von Zell- oder Gewebekomponenten erzeugt werden.

[0043] Die erfindungsgemäßen Verfahren setzen MS- und $(\text{MS})^n$ -Verfahren ein. Während eine Vielzahl an MS und $(\text{MS})^n$ verfügbar ist und in diesen Verfahren verwendet werden können, werden matrixunterstützte Laser-desorption/ionisierungs-MS-(MALDI/MS-) und Elektrosprühionisierungs-MS-(ESI/MS-)-Verfahren bevorzugt.

Quantitative Proteom-Analyse

[0044] Dieses Verfahren wird in Schema 1 unter Verwendung eines Biotin-markierten reaktiven Sulfhydrylre-

agenz für quantitative Proteinprofilmessungen in einem Probenproteingemisch und einem Referenzproteingemisch schematisch dargestellt. Das Verfahren umfaßt die folgenden Schritte:

Reduktion. Disulfidbindungen von Proteinen in den Proben- und Referenzgemischen werden zu freien SH-Gruppen reduziert. Das bevorzugte Reduktionsmittel ist Tri-n-butylphosphin, das unter Standardbedingungen verwendet wird. Alternative Reduktionsmittel umfassen Mercaptoethylamin und Dithiothreitol. Wenn erforderlich, kann diese Reaktion in Gegenwart von Lösungsvermittlern, einschließlich hoher Konzentrationen an Harnstoff, und Detergenzien durchgeführt werden, um die Proteinlöslichkeit aufrechtzuerhalten. Die zu vergleichenden Referenz- und Probenproteingemische werden separat verarbeitet, wobei identische Reaktionsbedingungen angewendet werden;

Derivatisierung von SH-Gruppen mit einer Affinitätsmarkierung. Freie SH-Gruppen werden mit dem biotinylierenden Reagenz Biotinyl-iodoacetylamidyl-4,7,10-trioxatridecandiamin derivatisiert, dessen Synthese nachstehend beschrieben wird. Das Reagenz wird in unterschiedlich isotoopenmarkierten Formen durch Substitution von Linkeratom mit stabilen Isotopen hergestellt, und jede Probe wird mit einer unterschiedlich isotoopenmarkierten Form des Reagenz derivatisiert. Die Derivatisierung von SH-Gruppen wird vorzugsweise unter leicht basischen Bedingungen (pH 8,5) für 90 min bei RT durchgeführt. Für die quantitative, vergleichende Analyse von zwei Proben wird jeweils eine Probe (bezeichnet als Referenzprobe und Probe), wie in Schema 1 dargestellt, mit der isotoopenleichten bzw. mit der isotoopenschweren Form des Reagenz derivatisiert. Für die vergleichende Analyse von mehreren Proben wird eine Probe als Referenz bezeichnet, auf die sich die anderen Proben beziehen. Typischerweise wird die Referenzprobe mit dem isotoopenschweren Reagenz markiert und die experimentellen Proben werden mit der isotoopenleichten Form des Reagenz markiert, obwohl die Wahl der Reagenzien willkürlich ist. Diese Reaktionen sind ebenso mit der Gegenwart von hohen Konzentrationen an Lösungsvermittlern kompatibel;

Kombination von markierten Proben. Nach Beendigung der Affinitätsmarkierungsreaktion werden definierte Aliquote der Proben, die mit den isotoopen-unterschiedlichen Reagenzien (beispielsweise schwere und leichte Reagenzien) markiert wurden, kombiniert, und alle nachfolgenden Schritte werden auf den zusammengefaßten Proben durchgeführt. Die Kombination von unterschiedlich markierten Proben zu diesem frühen Zeitpunkt der Verfahrensweise beseitigt die Variabilität aufgrund der anschließenden Reaktionen und Manipulationen. Vorzugsweise gleiche Mengen jeder Probe werden kombiniert;

Entfernung von überschüssigem affinitätsmarkiertem Reagenz. Überschüssiges Reagenz wird beispielsweise durch Zugabe eines Überschusses an SH-enhaltenden Kügelchen zu dem Reaktionsgemisch adsorbiert, nachdem die Protein-SH-Gruppen vollständig derivatisiert sind. Die Kügelchen werden zu der Lösung zugegeben, um etwa einen 5fachen molaren Überschuß an SH-Gruppen gegenüber dem Reagenz, das zugegeben und für 30 min bei RT inkubiert wurde, zu erreichen. Nach der Reaktion werden die Kügelchen durch Zentrifugation entfernt;

Proteinaufschluß. Die Proteine in dem Probengemisch werden typischerweise mit Trypsin aufgeschlossen. Alternative Proteasen sind ebenso mit der Verfahrensweise kompatibel, die tatsächlich chemische Fragmentierungsverfahren sind. In Fällen, in denen die vorhergehenden Schritte in Gegenwart von hohen Konzentrationen an denaturierenden Lösungsvermittlern durchgeführt wurden, wird das Probengemisch verdünnt, bis die Denaturierungsmittelkonzentration mit der Wirkung der verwendeten Proteasen kompatibel ist. Dieser Schritt kann bei der Analyse von kleinen Proteinen weggelassen werden;

Affinitätsisolation der affinitätsmarkierten Peptide durch Wechselwirkung mit einem Abfangreagenz. Die biotinylierten Peptide werden auf Avidin-Agarose isoliert. Nach dem Aufschluß wird der pH der Peptidproben auf 6,5 verringert, und die biotinylierten Peptide werden auf Kügelchen, die mit monomerem Avidin (Pierce) beschichtet wurden, immobilisiert. Die Kügelchen werden gründlich gewaschen. Das letzte Waschlösungsmittel umfaßt 10% Methanol, um restliches SDS zu entfernen. Biotinylierte Peptide werden aus Avidin-Agarose mit beispielsweise 0,3% Ameisensäure bei einem pH von 2 eluiert;

Analyse der isolierten, derivatisierten Peptide durch μ LC-MSⁿ oder CE-MSⁿ mit datenabhängiger Fragmentierung. Verfahrens- und Instrumentenkontrollprotokolle, die im Stand der Technik bekannt sind, und beispielsweise in Ducret et al., 1998; Figeys und Aebersold, 1998; Figeys et al., 1996; oder Haynes et al., 1998 beschrieben werden, werden verwendet.

[0045] In diesem letzten Schritt kann sowohl die Menge als auch die Sequenzidentität der Proteine, aus denen die markierten Peptide stammen, durch automatisierte Mehrstufen-MS bestimmt werden. Dies wird durch den Betrieb des Massenspektrometers in einem Dualmodus erreicht, in dem er sich mit aufeinanderfolgenden Scans zwischen dem Messen der relativen Mengen an Peptiden, die aus der Kapillarsäule eluieren, und dem Aufzeichnen der Sequenzinformationen der gewählten Peptide abwechselt. Peptide werden durch das Messen in der MS-Weise der relativen Signalintensitäten für Paare von Peptidionen von identischer Sequenz, die mit der isotoopenleichten bzw. -schweren Form des Reagenz markiert werden, quantifiziert, und die sich daher in der Masse durch den Massenunterschied, der innerhalb des affinitätsmarkierten Reagenz codiert ist, unterscheiden. Die Peptidsequenzinformation wird automatisch durch die Wahl der Peptidionen eines besonderen

Masse-zu-Ladungs-(m/z)-Verhältnisses für stoßinduzierte Dissoziation (CID) in dem Massenspektrometer, das in der MSⁿ-Weise arbeitet, erzeugt. (Link, A. J. et al. (1997), Gygi, S. P. et al. (1999) und Gygi, S. P. et al. (1999)). Die resultierenden CID-Spektren werden dann automatisch mit den Sequenzdatenbanken in Beziehung gesetzt, um das Protein, aus dem das sequenzierte Peptid stammt, zu identifizieren. Die Kombination der Ergebnisse, die durch MS- und MSⁿ-Analysen von affinitätsmarkierten und unterschiedlich markierten Peptidproben erzeugt wurden, bestimmt daher die relativen Mengen sowie die Sequenzidentitäten der Komponenten von Proteingemischen in einem einzelnen, automatisierten Verlauf.

[0046] Die Ergebnisse der Anwendung dieses Verfahrens unter Verwendung des biotinylierten Sulfhydrylreagenz und der quantitativen Analyse von synthetischen Peptidproben, der relativen Quantifizierung der Peptide in einem Proteinaufschluß und die Tandem-Massenspektroalanalyse eines derivatisierten Peptids werden in **Fig. 1**, **Tabelle 1** bzw. **Fig. 2** gezeigt.

[0047] Dieses Verfahren kann ebenso unter Verwendung anderer Affinitätsmarkierungen und anderer reaktiver Proteingruppen, einschließlich reaktiver Aminogruppen, reaktiver Carboxylgruppen oder Gruppen, die mit Homoserinlaktone reagieren, praktiziert werden.

[0048] Der Ansatz, der hierin zur quantitativen Proteom-Analyse eingesetzt wird, basiert auf zwei Prinzipien. Zunächst enthält eine kurze Sequenz von angrenzenden Aminosäuren von einem Protein (5 bis 25 Reste) ausreichende Informationen, um nur dieses Protein zu identifizieren. Die Proteinidentifizierung durch MSⁿ wird erreicht durch das in Beziehung setzen der Sequenzinformation, die in dem CID-Massenspektrum enthalten ist, mit den Sequenzdatenbanken unter Verwendung hoch entwickelter Computerforschungsalgorithmen (Eng, J. et al. (1994); Mann, M. et al. (1994); Qin, J. et al. (1997); Clauser, K. R. et al. (1995)). Zweitens sind Paare von identischen Peptiden, die mit den leichten bzw. schweren affinitätsmarkierten Reagenzien markiert wurden (oder bei der Analyse von mehr als zwei Proben, Gruppen von identisch markierten Peptiden, bei denen jedes Gruppenmitglied unterschiedlich isotoopenmarkiert wird), chemisch identisch und dienen daher als gegenseitige interne Standards zur genauen Quantifizierung. Die MS-Messung unterscheidet ohne weiteres zwischen Peptiden, die aus unterschiedlichen Proben stammen, welche beispielsweise unterschiedliche Zellzustände darstellen, wegen des Unterschieds zwischen isotoopen-unterschiedlichen Reagenzien, die an die Peptide gebunden sind. Die Verhältnisse zwischen den Intensitäten der unterschiedlichen Gewichtskomponenten dieser Paare oder Gruppen an Peaks stellen ein genaues Maß der relativen Häufigkeit der Peptide (und daher der Proteine) in den ursprünglichen Zellpools bereit, da die MS-Intensitätsantwort eines gegebenen Peptids unabhängig von der isotoopen Zusammensetzung der Reagenzien ist (De Leenheer, A. P. et al (1992)). Die Verwendung von isotoopenmarkierten internen Standards ist die Standardpraxis bei der quantitativen Massenspektrometrie und ist vorteilhafter beispielsweise bei der genauen Quantifizierung von Arzneimitteln und Stoffwechselprodukten in Körperflüssigkeiten (De Leenheer, A. P. et al. (1992)) genutzt worden.

[0049] In einer anderen Darstellung des Verfahrens wurden zwei Gemische, die aus denselben sechs Proteinen bei bekannten, aber unterschiedlichen Konzentrationen bestehen, hergestellt und analysiert. Die Proteingemische wurden markiert, kombiniert und behandelt, wie in Schema 1 schematisch dargestellt, welches nicht Bestandteil der Beschreibung ist. Die isolierten, markierten Peptide wurden quantifiziert und in einem einzeln kombinierten μ LC-MS- und μ LC-MSⁿ-Experiment auf einem ESI-Ion-Trap-Massenspektrometer sequenziert. Alle sechs Proteine wurden unzweideutig identifiziert und genau quantifiziert (**Tabelle 2**). Mehrfach markierte Peptide wurden für jedes Protein zusammengestoßen. Die Unterschiede zwischen den beobachteten und den erwarteten Mengen für die sechs Proteine lag zwischen 2 und 12%.

[0050] Das Verfahren wird außerdem für ein einzelnes Peptidpaar in den **Fig. 3A** bis C dargestellt. Ein einzelner Scann des Massenspektrometers, das in der MS-Weise arbeitet, wird in **Fig. 3A** gezeigt. Vier Paare von Peptidionen, die durch den Massenunterschied, der das affinitätsmarkierte Reagenz codiert, charakterisiert werden, werden in diesem Scann nachgewiesen und mit ihren jeweiligen m/z-Werten angegeben. Der gezeigte Scann wurde in 1,3 s erreicht. Während des Verlaufs des einstündigen chromatographischen Elutionsgradienten wurden mehr als 1200 dieser Scanns automatisch aufgezeichnet. **Fig. 3B** zeigt eine erweiterte Ansicht des Massenspektrums um das Ionenpaar mit m/z-Verhältnissen von 993,8 bzw. 977,7. Coelution und ein nachgewiesener Massenunterschied von vier Einheiten identifiziert möglicherweise die Ionen als Paar von zweifach geladenen affinitätsmarkierten Peptiden von identischer Sequenz (Massenunterschied von acht und ein Ladungszustand von zwei). **Fig. 3C** zeigt die rekonstruierten Ionenchromatogramme für diese zwei Spezies. Die relativen Mengen wurden durch Integrieren der Kontur der jeweiligen Peaks bestimmt. Das Verhältnis (leicht/schwer) wurde als 0,54 (**Tabelle 1**) bestimmt. Die Peaks in den rekonstruierten Ionenchromatogrammen erscheinen gezackt, da in jedem zweiten Scann das Massenspektrometer zwischen der MS- und der MSⁿ-Weisen umschaltet, um Sequenzinformationen (CID-Massenspektrum) eines gewählten Peptidions zu sammeln.

Diese CID-Spektren wurden verwendet, um das Protein, aus dem die markierten Peptide stammen, zu identifizieren. [Fig. 4A](#) zeigt das CID-Spektrum, das aus dem Peptidion mit $m/z = 998$ (markiert mit einem Pfeil in [Fig. 3A](#)) aufgezeichnet wurde. Datenbankforschung mit diesem CID-Spektrum identifizierte das Protein als Glyceraldehyd-3-phosphatdehydrogenase ([Fig. 4B](#)), was ein Element des Proteingemisches war.

[0051] Mehrere nützliche Merkmale dieses Verfahrens sind offensichtlich. Zunächst wurden mindestens zwei Peptide aus jedem Protein in dem Gemisch nachgewiesen. Daher kann sowohl die Quantifizierung als auch die Proteinidentifizierung überflüssig sein. Zweitens enthielten die identifizierten Peptide alle mindestens einen markierten Cysteinylrest. Die Gegenwart des relativ seltenen Cysteinylrestes in einem Peptid fügt einen zusätzlich starken Zwang zur Datenbankforschung hinzu (Sechi, S. et al. (1998)). Drittens verringert das Markieren und die selektive Anreicherung von Cystein-enthaltenden Peptiden signifikant die Komplexität des Peptidgemisches, das durch den gleichzeitigen Aufschluß der sechs Proteine erzeugt wurde. Für dieses Proteingemisch wurde die Komplexität von 293 potentiellen Trypsinpeptiden auf 44 Trypsinpeptide, die mindestens einen Cysteinylrest enthalten, verringert. Viertens sind die Peptidproben, die aus der Avidinaffinitätsäule eluierten, mit der Analyse durch μ LC-MSⁿ direkt kompatibel.

Quantitative Analyse der Proteinexpression in unterschiedlichen Zellzuständen

[0052] Die reaktive Proteinaffinitätsreagenzstrategie wurde angewendet, um die Unterschiede bei der stationären Proteinexpression in Hefe, *S. cerevisiae*, in zwei nicht-Glukose unterdrückten Zuständen (Tabelle 3) zu untersuchen. Zellen wurden aus der Hefe geerntet, die in der log-Phase wächst, wobei entweder 2% Galaktose oder 2% Ethanol als Kohlenstoffquelle genutzt werden. Einhundert μ g des löslichen Hefeproteins aus jedem Zellzustand wurden unabhängig voneinander mit den isotopen-unterschiedlichen affinitätsmarkierten Reagenzien markiert. Die markierten Proben wurden vereinigt und der Strategie, die in Schema 1 beschrieben wird, unterzogen. Ein Fünffzigstel (Äquivalent von ungefähr 2 μ g Protein aus jedem Zellzustand) der Probe wurde analysiert.

[0053] Die Glukoserepression verursacht große Anzahlen von Proteinen mit Stoffwechselfunktionen, die für das Wachstum auf anderen Kohlenstoffquellen, die minimal exprimiert werden sollen (Ronne, H. (1995); Hodges, P. E. et al. (1999)) signifikant sind. Das Wachstum auf Galaktose oder Ethanol ohne Glukose führt zur Expression von Glukose-unterdrückten Genen. Tabelle 3 stellt eine Auswahl von 34 Hefegenen dar, die in der Analyse zusammentreffen, aber sie enthält jede bekannten Glukose-unterdrückten Gene, die identifiziert wurden (Mann, M. et al. (1994)). Jedes dieser Gene würde minimal in der Hefe, die auf Glukose wuchs, exprimiert werden. Gene, die sowohl für das Wachstum auf Galaktose (GAL1, GAL10) als auch das Wachstum auf Ethanol (ADH2, ACH1) spezifisch sind, wurden nachgewiesen und quantifiziert.

[0054] Die quantitative Beschaffenheit des Verfahrens ist offensichtlich fähig, die kleinen Veränderungen der relativen Proteinniveaus genau zu messen. Der Beweis der Genauigkeit der Messungen kann aus der ausgezeichneten Übereinstimmung hervorgehen, die durch das Überprüfen der Verhältnisse für Proteine, für die mehrfache Peptide quantifiziert wurden, festgestellt wurde. Beispielsweise wiesen die fünf Peptide, gefunden aus PCK1, ein mittleres Verhältnis von $\pm 95\%$, Konfidenzintervalle von $1,57 \pm 0,15$ auf, und der Prozentfehler betrug $< 10\%$. Außerdem trafen die beobachteten Veränderungen auf die erwarteten Veränderungen aus der Literatur zu (Ronne, H. 1995; Hodges, P. E. et al. (1999)). Schließlich stimmen die beobachteten Veränderungen mit den Veränderungen in der Farbintensität für dieselben Proteine, die nach der zweidimensionalen Gelelektrophorese (Daten werden nicht gezeigt) untersucht wurden, überein.

[0055] Die Alkoholdehydrogenasefamilie von Isoenzymen in Hefe erleichtert das Wachstum auf entweder Hexosezuckern (ADH1) und Ethanol (ADH2). Das Gen ADH2 kodiert ein Enzym, das sowohl Glukose- als auch Galaktose-unterdrückt wird, und ermöglicht einer Hefezelle, vollständig auf Ethanol durch deren Umwandeln in Acetaldehyd, das in den TCA-Kreislauf eindringt ([Fig. 5A](#)), zu wachsen. In Gegenwart von Zucker führt ADH1 die Umkehrreaktion durch, die Acetaldehyd in Ethanol umwandelt. Die Regulierung dieser Isoenzyme ist der Schlüssel zur Kohlenstoffnutzung in Hefe (Ronne, H. (1995)). Die Fähigkeit, die Unterschiede in der Genexpression über Familien der Isoenzyme genau zu messen, ist unter Verwendung von cDNA-Anordnungs-techniken aufgrund der Kreuzhybridisierung (DeRisi, J. L. et al. (1997)) manchmal schwierig. Dem erfindungsgemäßen angewendeten Verfahren, wie in [Fig. 1](#) dargestellt, gelang die Messung der Genexpression für jedes Isoenzym, selbst wenn sich ADH1 und ADH2 93% Aminosäure-(88% Nukleotid)-sequenzähnlichkeit teilen. Dies war so, weil sich die affinitätsmarkierten Peptide von jedem Isoenzym durch einen einzelnen Aminosäurerest (Valin oder Threonin) unterschieden, der die Retentionszeit um mehr als 2 min und die Masse um 2 Dalton für die ADH2-Peptide (Schema 5B) verschob. ADH1 wurde bei ungefähr den 2fach hohen Niveaus exprimiert, wenn Galaktose die Kohlenstoffquelle im Vergleich zu Ethanol war. Die Ethanol-Induktion von ADH2-Ex-

pression führte im Vergleich zur Galaktose-Induktion zu mehr als 200fachen Erhöhungen.

[0056] Die oben beschriebenen Ergebnisse zeigen, daß das erfindungsgemäße Verfahren die quantitative Analyse von Proteingemischen und die Identifizierung der Proteinkomponenten darin in einem einzelnen, automatisierten Vorgang bereitstellt.

[0057] Das Verfahren, wie unter Verwendung eines reaktiven Sulfhydrylreagenz angewendet, verringert signifikant die Komplexität der Peptidgemische, da affinitätsmarkierte Cystein-enthaltende Peptide selektiv isoliert werden. Beispielsweise stellte ein theoretischer Trypsinaufschluß des gesamten Hefeproteoms (6113 Proteine) 344.855 Peptide her, aber nur 30.619 von diesen Peptiden enthalten einen Cysteinylnrest. Daher wird die Komplexität des Gemisches verringert, während die Proteinquantifizierung und -identifizierung noch erreicht werden. Die chemische Reaktion des Sulfhydrylreagenz mit Protein kann in Gegenwart von Harnstoff, Natriumdodecylsulfat (SDS), Salzen und anderen Chemikalien, die keine reaktive Thiolgruppe enthalten, durchgeführt werden. Daher können die Proteine mit starken Lösungsvermittlern, bis sie enzymatisch aufgeschlossen werden, in Lösung gehalten werden. Die Empfindlichkeit des μ LC-MSⁿ-Systems hängt von der Probenqualität ab. Insbesondere sind allgemein verwendete Proteinlösungsvermittler schlecht kompatibel oder inkompatibel mit MS. Die Affinitätsreinigung der markierten Peptide entfernt vollständig die Kontaminanten, die mit MS inkompatibel sind. Die Quantifizierung und Identifizierung von Proteinen mit geringer Häufigkeit durch konventionelle Verfahren erfordern große Mengen (Milligramm) an Ausgangsproteinlysate und umfassen einige Arten der Anreicherung für diese Proteine mit geringer Häufigkeit. Die oben beschriebenen Assays beginnen mit etwa 100 μ g Protein und verwendeten keine Fraktionierungstechniken. Davon wurde ungefähr 1/50 des Proteins in einem einzelnen μ LC-MSⁿ-Experiment analysiert. Dieses System weist eine Nachweisgrenze von 10 bis 20 fmol pro Peptid auf (Gygi, S. P. et al. (1999)). Aus diesem Grund werden die beschriebenen Assays, die μ LC-MSⁿ einsetzen, nur häufige Proteine nachweisen. Jedoch sind die erfindungsgemäßen Verfahren mit jedem biochemischen, immunologischen oder zellbiologischen Fraktionierungsverfahren kompatibel, die die Gemischkomplexität verringern und Proteine mit geringer Häufigkeit anreichern, während die Quantifizierung aufrechterhalten wird. Dieses Verfahren kann sowohl bei der Quantifizierung als auch bei der Identifizierung überflüssig sein, wenn Cysteine nachgewiesen werden. Es gibt einen dynamischen Bereich, der mit der Fähigkeit des Verfahrens, die Unterschiede in Expressionsniveaus von affinitätsmarkierten Peptiden zu quantifizieren, verbunden ist, was sowohl von der Intensität der Peaks, die auf das Peptidpaar (oder Gruppe) zutreffen, als auch von der Gemischkomplexität abhängt. Außerdem wird dieser dynamische Bereich für jeden Typ des verwendeten Massenspektrometers unterschiedlich sein. Der Ion-Trag wurde in den hierin beschriebenen Assays aufgrund seiner Fähigkeit, beeindruckende Mengen an Sequenzinformationen (tausend Proteine können möglicherweise identifiziert werden) in einer datenabhängigen Weise zu sammeln, selbst wenn er einen begrenzteren dynamischen Quantifizierungsbereich bietet, eingesetzt. Der dynamische Bereich des Ion-Traps (basiert auf Signal-Rausch-Verhältnis) variierte in Abhängigkeit der Signalintensität des Peptidpaares und der Komplexität des Gemisches, aber Unterschiede von bis zum 100fachen waren im allgemeinen nachweisbar und sogar größere Unterschiede konnten für häufigere Peptide bestimmt werden. Außerdem identifizieren Proteinexpressionsniveauveränderungen von mehr als 100- bis 200fach noch diese Proteine als wichtige potentielle Beteiligte an den phenotypischen Unterschieden zwischen den zwei originalen Zellzuständen. Das Verfahren kann erweitert werden, um die Reaktivität gegen andere funktionelle Gruppen zu umfassen. Ein kleiner Prozentsatz an Proteinen (8% für *S. cerevisiae*) enthält keine Cysteinylnreste und werden daher durch die Analyse unter Verwendung von Reagenzien mit Sulfhydrylgruppenspezifität (d. h. Thiolgruppenspezifität) übersehen. Affinitätsmarkierte Reagenzien mit Spezifitäten gegen funktionelle Gruppen anders als Sulfhydrylgruppen machen ebenso Cystein-freie Proteine für die Analyse empfindlich.

[0058] Die erfindungsgemäßen Verfahren können auf die Analyse von Proteinen mit geringer Häufigkeit und Klassen an Proteinen mit besonderen physikalisch-chemischen Eigenschaften, einschließlich schlechte Löslichkeit, große und kleine Größe und extreme pI-Werte, angewendet werden.

[0059] Die prototypische Anwendung der Chemie und des Verfahrens ist die Herstellung von quantitativen Profilen von komplexen Proteinproben und schließlich gesamten Lysaten von Zellen und Geweben nach dem bevorzugten Verfahren, das oben beschrieben wird. Außerdem weisen die Reagenzien und erfindungsgemäßen Verfahren Anwendungen auf, die über die Bestimmung der Proteinexpressionsprofile hinaus gehen. Derartige Anwendungen umfassen die folgenden:

Anwendung von Amino-reaktiven und Sulfhydryl-reaktiven, unterschiedlich isotoopenmarkierten, affinitätsmarkierten Reagenzien zur quantitativen Analyse von Proteinen in Immunfällungskomplexen. In der bevorzugten Version dieser Technik werden Proteinkomplexe aus Zellen, die unterschiedliche Zustände darstellen (beispielsweise unterschiedliche Zustände an Aktivierung, unterschiedliche Krankheitszustände, unterschiedliche Zustände an Differenzierung), mit einem spezifischen Reagenz, vorzugsweise einem Antikörper ausgefällt. Die

Proteine in dem ausgefällten Komplex werden dann derivatisiert und wie oben analysiert.

[0060] Anwendung von Amino-reaktiven, unterschiedlich isotoopenmarkierten, affinitätsmarkierten Reagenzien, um die Orte von induzierter Proteinphosphorylierung zu bestimmen. In einer bevorzugten Version dieses Verfahrens werden gereinigte Proteine (beispielsweise immunausgefällt aus Zellen unter unterschiedlichen Anregungsbedingungen) fragmentiert und, wie oben beschrieben, derivatisiert. Phosphopeptide werden in dem resultierenden Peptidgemisch durch Fragmentierung in der Ionenquelle des ESI-MS-Geräts identifiziert, und deren relative Häufigkeiten werden durch Vergleichen der Ionensignalintensitäten der experimentellen Probe mit der Intensität eines eingeschlossenen isotoopenmarkierten Standards bestimmt.

[0061] Amino-reaktive, unterschiedlich isotoopenmarkierte, affinitätsmarkierte Reagenzien werden verwendet, um die N-terminalen Ionenreihen in MSⁿ-Spektren zu identifizieren. In einer bevorzugten Version dieser Anwendung werden die zu analysierenden Peptide mit einem 50:50-Gemisch eines isotoopenleichten und -schweren Reagenz, das für Aminogruppen spezifisch ist, derivatisiert. Die Fragmentierung der Peptide durch CID stellt daher zwei N-terminale Ionenreihen her, die sich in der Masse genau durch den Massenunterschied der verwendeten Reagenzspezies unterscheiden. Diese Anwendung verringert dramatisch die Schwierigkeit beim Bestimmen der Aminosäuresequenz des derivatisierten Peptids.

Quantitative Analyse von Oberflächenproteinen in Zellen und Gewebe

[0062] Die Zelloußenmembran und ihre assoziierten Proteine (Zelloberflächenproteine) nehmen am Abtasten äußerer Signale und Reagieren auf Umweltsignale teil. Veränderungen der Häufigkeit von Zelloberflächenproteinen können einen speziellen Zellzustand oder die Fähigkeit einer Zelle, auf ihre sich verändernde Umwelt zu reagieren, widerspiegeln. Daher kann die umfassende, quantitative Charakterisierung der Proteinkomponenten der Zelloberfläche Markerproteine oder Konstellationen von Markerproteinen, die für einen speziellen Zellzustand charakteristisch sind, identifizieren, oder die Molekularbasis für Zellantworten auf äußere Reize erklären. Tatsächlich sind Veränderungen bei der Expression einer Anzahl von Zelloberflächenrezeptoren, wie Her2/neu-, erbB-, IGFI-Rezeptor und EGF-Rezeptor, in die Krebsentstehung hineingezogen worden, und ein derzeitiger immunologischer, therapeutischer Ansatz für Brustkrebs basiert auf der Infusion eines Antikörpers (Herceptin, Genentech, Palo Alto, CA), der speziell Her2/neu-Rezeptoren erkennt.

[0063] Zelloberflächenproteine sind ebenso experimentell zugänglich. Diagnostische Assays zur Zellklassifizierung und präparative Isolation von speziellen Zellen durch Verfahren, wie Zellsortieren oder Panning, basieren auf Zelloberflächenproteinen. Daher kann die Differentialanalyse von Zelloberflächenproteinen zwischen normalen und kranken (beispielsweise Krebs) Zellen wichtige diagnostische oder therapeutische Targets identifizieren. Während die Wichtigkeit von Zelloberflächenproteinen zur Diagnose und Therapie von Krebs erkannt worden ist, sind Membranproteine schwierig zu analysieren gewesen. Aufgrund ihrer allgemein schlechten Löslichkeit werden sie gewöhnlich in Standard-2D-Gelelektrophoresemustern unterrepräsentiert, und Versuche, 2D-Elektrophoresebedingungen der Trennung von Membranproteinen anzupassen, trafen auf eingeschränkten Erfolg. Das erfindungsgemäße Verfahren kann die in den traditionellen Techniken inhärenten Einschränkungen überwinden.

[0064] Die Analyse von Membranproteinen ist anspruchsvoll, da es im allgemeinen schwierig ist, sie in Lösung unter Bedingungen aufrechtzuerhalten, die mit hochempfindlichen Analysegeräten, wie Massenspektrometern, kompatibel sind. Die Anwendung der erfindungsgemäßen Verfahren auf die Analyse von Membranproteinen wird unter Verwendung der humanen T-Zellymphom-Zelllinie Jurkat zum Membranproteinmarkieren und Extraktion und die gut charakterisierte humane Prostataepithelzelllinie P69SV40T und zwei P69SV40T-Sublinien veranschaulicht, die sich bei der IGF-1-Rezeptorexpression durch den Faktor 10 unterscheiden, um quantitative Differentialanalyse von Membranproteinen zu veranschaulichen.

[0065] Jurkat-Zellen sind ein geeignetes Modellsystem, da die Zellen leicht in großer Anzahl wachsen, und da die Modulation von Zelloberflächenproteinen als Antwort auf die unterschiedlichen Reize und experimentellen Bedingungen in T-Lymphozyten gut charakterisiert worden ist. Kommerziell erhältliche biotinylierte Reagenzien oder allgemeiner affinitätsmarkierte Reagenzien werden eingesetzt, um Lysinreste und die freien N-Enden zu derivatisieren. Wasserlösliche biotinylierte Reagenzien, wie Sulfo-NHS-(N-Hydroxysuccinimid-)-biotin, und Analoga (Sulfosuccinimidyl-6-(biotinamido)-hexanoat, Pierce, Rockford, IL), die zum Markieren von Zelloberflächenproteinen umfangreich verwendet worden sind, können eingesetzt werden. Die Reaktion von NHS-Estern mit primären Aminen ist am besten bei neutralen pH-Werten und darüber, und ist mit der Gegenwart von organischem Lösungsmittel, wie DMSO oder DMF, kompatibel. Die Biotinylierung von Zelloberflächenproteinen von Jurkat-Zellen wird in PBS-Puffer bei einem pH von 7,2 durchgeführt. Zellen (1×10^7)

werden mit PBS-Puffer gewaschen, um kontaminierendes Serum und andere Proteine aus dem Kulturmedium zu entfernen. Die Zellen werden bei 25×10^6 Zellen/ml resuspendiert und mit 0,5 mg/ml Sulfo-NHS-Biotin (Pierce, Rockford, IL) 30 min bei RT umgesetzt. Die markierten Zellen werden zweimal mit kaltem PBS gewaschen, um nicht-umgesetztes Biotinylierungsreagenz zu entfernen. Biotinylierte Zellen werden bei 5×10^7 Zellen/ml in Lysispuffer, der 1% Triton X-114 enthält, löslich gemacht. Triton X-114 hat die Eigenschaft der Phasenverteilung in eine Detergenzphase und eine wässrige Phase bei 30°C. Nach der Phasenverteilung wird die Detergenzphase aus der wässrigen Phase durch Zentrifugation bei 300 xg entfernt. Phasenverteilung ist zuvor erfolgreich verwendet worden, um die Zellmembran anzureichern. Ebenso wurde festgestellt, daß diese Technik Membranproteine aus Jurkat-Zellysaten anreichert. Die Tritonphase wird 1:5 (v/v) unter Verwendung von 50 mM Ammoniumbicarbonatpuffer, pH 8,5, verdünnt, und hochreines modifiziertes Schweinetrypsin wird zugegeben, um die Proteine bei einer Konzentration von 12,5 ng/ml über Nacht bei 37°C aufzuschließen. Das Trypsin wird durch Zugabe eines Cocktails aus Serinproteaseinhibitoren neutralisiert und Trypsinpeptide werden durch Avidinaffinitätschromatographietechniken isoliert. Eluierte Peptide werden beispielsweise durch μ LC-Verfahren getrennt und durch Forschungspeptidsequenzdatenbanken unter Verwendung von beispielsweise dem Sequestrogramm identifiziert.

[0066] Die humane Prostataepithelzelllinie P69SV40T, die mit SV 40 T-Antigen immortalisiert wurde, ist gut charakterisiert worden. Diese Zelllinie ist unsterblich, aber nicht tumorerzeugend und exprimiert Typ 1 insulin-ähnlichen Wachstumsfaktorrezeptor (IGF-1R) bei 2×10^4 Rezeptoren pro Zelle. Eine Sublinie, M12 genannt, stammte aus P69SV40T durch sequentiellen Durchlauf bei männlichen athymischen Nacktmäusen. Diese Zelllinie ist stark tumorerzeugend und metastatisch und exprimiert $1,1 \times 10^3$ IGF-1R pro Zelle. Der relative Unterschied in der Häufigkeit von IGF-1R in den Zelllinien P69SV40T und M12 kann unter Verwendung erfindungsgemäßer Verfahren, die zur Anwendung auf Membranproteine angepaßt werden, quantitativ bestimmt werden. Da die Anzahl an IGF-1R bereits für diese Zellen bestimmt worden ist, kann dieses gut charakterisierte System eine Referenz bereitstellen, um die Wirksamkeit der erfindungsgemäßen quantitativen Verfahren zu bestätigen.

[0067] P69SV40T-Zellen (1×10^7) werden mit einem isotopenschweren, biotinmarkierten, reaktiven Aminoreagenz biotinyliert und die M12-Zellen (1×10^7) werden mit einem entsprechenden isotopeleichten, Amin-reaktiven, biotinmarkierten, Amino-reaktiven Reagenz biotinyliert. IGF-1R wird dann aus dem kombinierten Lysat von beiden Zelllinien unter Verwendung eines Antikörpers gegen humanes IGF-1R immun ausgefällt, und die Gesamtmasse an immun ausgefallenen Proteinen wird mit Trypsin aufgeschlossen. Trypsin wird dann beispielsweise durch die Zugabe von Inhibitoren neutralisiert und markierte Peptide werden durch Biotin-Avidin-Affinitätschromatographie markiert. Die eluierten Peptide werden durch LC-MS und LC-MSⁿ zur Peptidquantifizierung bzw. -identifizierung analysiert, wie es oben beschrieben worden ist. Die Quantifizierung wird in diesem Experiment durch die Option, selektives Ionenüberwachen in der MS zu verwenden, erleichtert. In dieser Weise brauchen nur die Massen an markierten Peptidionen, bei denen erwartet wird, daß sie aus IGF-1R stammen, überwacht zu werden.

[0068] Die beschriebene Technik kann zum Vergleich der Unterschiede in der relativen Häufigkeit von Zelloberflächenproteinen zwischen der parenteralen Prostatazelllinie (P69SV40T) und M12-Zellen angewendet werden, um diese Zelloberflächenproteine, deren Expressionsniveau in den zwei Zelllinien unterschiedlich ist, und die für die unterschiedlichen Zellzustände charakteristisch sind, nachzuweisen und zu identifizieren. Unter Verwendung der hierin beschriebenen Verfahren kann die globale, relative Quantifizierung der Zelloberflächenproteine in zwei oder mehr Zelllinien analysiert werden, um diese Zelloberflächenproteine, die für die unterschiedlichen Zellzustände charakteristisch sind, nachzuweisen und zu identifizieren. Die Ergebnisse können unter Verwendung der Verfahrensweise, wie 1D- oder 2D-Gele, wenn anwendbar, oder quantitativem Western-Blotting unabhängig voneinander bestätigt werden, um die Quantifizierungsergebnisse zu bestätigen.

[0069] Es wird erwartet, daß die experimentelle Variabilität der Quantifizierung von Zelloberflächenproteinen beträchtlich besser sein wird als die Genauigkeit der Quantifizierung, die durch die derzeit verfügbare cDNA-Anordnungstechnik erreicht wird. Zusätzlich zur relativen Proteinmenge und -identität kann das Verfahren ebenso verwendet werden, um die Orientierung des Proteins in der Membran zu offenbaren, basierend auf der Vermutung, daß intakte, lebende Zellen das Biotinylierungsreagenz ausschließen.

[0070] Alternative Verfahren können angewendet werden, um die Selektivität für markierte Peptide, die aus den Zelloberflächenproteinen stammen, zu verbessern. Beispielsweise können markierte Zelloberflächenproteine, wie erläutert, direkt auf den intakten Zellen trypsinisiert, um markierte Peptide zu erzeugen, gereinigt und analysiert werden. Außerdem können traditionelle Zellmembranpräparate als Ausgangsschritt verwendet werden, um die Zelloberflächenproteine anzureichern. Diese Verfahren können vorsichtige Zytolyse mit einer

Dounce-Homogenisiervorrichtung und Reihen von Dichtegradientenzentrifugationen umfassen, um Membranproteine vor der Proteolyse zu isolieren. Dieses Verfahren kann hoch angereicherte Präparate von Zelloberflächenproteinen bereitstellen. Affinitätsmarkierte Proteine können ebenso durch Affinitätschromatographie vor der Proteolyse sowie nach der Proteolyse isoliert werden. Diese Chromatographie kann in Gegenwart von grenzflächenaktiven Mitteln, wie TX-100, NP-40 oder Tween-20, durchgeführt werden, um die Proteinlöslichkeit aufrechtzuerhalten. Die sequentielle Anwendung von Affinitätschromatographieschritten (einen für das intakte Protein und einer für die markierten Peptidfragmente) stellte einen hohen Selektivitätsgrad bereit. Diese alternativen Verfahren sind für den Nachweis von Membranproteinen mit geringer Häufigkeit leicht skalierbar, und die relative Menge von markierten Peptiden wird durch die selektiven Anreicherungsschritte aufrechterhalten.

[0071] Wenn bei der Anwendung dieser erfindungsgemäßen Verfahren auf Zelloberflächenproteine die markierten Proteine einmal fragmentiert sind, verhalten sich die markierten Peptide nicht anders als die Peptide, die aus löslicheren Proben erzeugt wurden.

Synthese von affinitätsmarkierten reaktiven Proteinreagenzien, die für bestimmte Proteingruppen selektiv sind

[0072] Synthesewege für beispielhafte affinitätsmarkierte Reagenzien, die zur Verwendung bei den erfindungsgemäßen Verfahren geeignet sind, werden in den Schemen 2 bis 3 bereitgestellt, wo allgemein bekannte Synthesetechniken bei der Synthese von nicht-deutierten und deutierten Reagenzien eingesetzt werden.

[0073] Biotinyl-iodoacetylamidyl-4,7,10-trioxatridecandiamin-4 (Schema 3) besteht aus einer Biotingruppe, einem chemisch inerten Spacer, der fähig ist, mit stabilen Isotopen isotonenmarkiert zu werden, bzw. einer Iodoacetylgruppe. Die Biotingruppe wird zur Affinitätsanreicherung von Peptiden, die mit dem Reagenz derivatisiert sind, verwendet, der Ethylenglykollinker wird zur Massenspektrometrie unterschiedlich isotonenmarkiert, und die Iodoacetylgruppe stellt die Spezifität des Reagenz für Sulfhydryl-enthaltende Peptide bereit. Das Reagenz kann in einer reinen Wasserstoffform (isotonenleichte Form) mit 1 bis 20, und vorzugsweise 4 bis 8 Deuteriumatomen in dem Linker (isotonenschwere Formen) synthetisiert werden.

Beispielhafte MS^N-Techniken und Vorrichtungen

[0074] Ein automatisiertes LC-MS/MS-System zur Identifizierung von Proteinen durch deren Aminosäuresequenz ist entwickelt worden. Eine schematische Darstellung wird in [Fig. 7](#) gezeigt. Das System, das aus einem Autosampler, einem Kapillar-HPLC-System, das an eine ESI-Dreifach-Quadrupol-MS/MS-Vorrichtung on-line gebunden ist, und einem Datensystem besteht, wird in der folgenden Weise betrieben: Proteine (typischerweise durch 1D- oder 2D-Gelelektrophorese abgetrennt) werden mit einer speziellen Protease, normalerweise Trypsin, gespalten. Die resultierenden Spaltungsfragmente werden in einen Autosampler eingebracht. Alle 37 Minuten injiziert der Autosampler eine Probe in das HPLC-System und die Peptide werden durch Kapillar-Umkehrphasen-Chromatographie abgetrennt. Da abgetrennte Peptide aus der Chromatographiesäule eluieren, werden sie durch das ESI-Verfahren ionisiert, treten in die MS ein und das Masse zu Ladungsverhältnis (m/z) wird gemessen. Irgendein Peptidion, dessen Intensität einen vorbestimmten Intensitätsschwellenwert übersteigt, wird automatisch durch die Vorrichtung ausgewählt und in der Kollisionszelle mit Inertgas kollidiert. Diese Kollisionen führen zur Peptidfragmentierung, in erster Linie an den Bindungen der Peptidhauptkette (stoßinduzierte Dissoziation, CID). Die Massen der CID-Fragmente werden gemessen und in dem Datensystem aufgezeichnet. Das CID-Spektrum eines Peptids enthält ausreichende Informationen, um das Protein durch Forschungssequenzdatenbanken mit den nicht-ausgewerteten MS/MS-Spektren zu identifizieren. Dies wird mit dem Sequenz-Programm erreicht. Das Programm identifiziert jedes Peptid in einer Sequenzdatenbank, das dieselbe Masse wie das Peptid aufweist, das in der MS für CID ausgewählt wurde, und sagt das MS/MS-Spektrum für jeweils eine der isobaren Peptide voraus. Durch Vergleichen des experimentell bestimmten CID-Spektrums mit den computererzeugten theoretischen CID-Spektren wird das Protein, aus dem das beobachtete Peptid stammt, identifiziert. Das System ist zum Analysieren von Proteinproben in einer vollständig automatisierten Weise bei einer Geschwindigkeit von weniger als 40 min pro Probe fähig. Da jedes Peptid eine unabhängige Proteinidentifizierung darstellt und normalerweise mehrfache Peptide aus einem Protein stammen, ist die Proteinidentifizierung durch dieses Verfahren überflüssig und tolerant zu Proteinen, die in ein Gel komigrieren. Das System ist zum Nachweis und zur Charakterisierung von modifizierten Resten innerhalb der Polypeptidketten gut geeignet. Die LC-MS/MS-Technik und automatisierte Analyse von erzeugten CID-Spektren können für die erfindungsgemäßen Verfahren verwendet werden.

Identifizierung von Proteinen bei sub-femtomol-Empfindlichkeit durch Festphasenextraktion-Kapillarelektrophorese-Tandem-Massenspektrometrie (SPE-CE-MS/MS)

[0075] Die Proteinidentifizierung durch dieses Verfahren basiert auf demselben Prinzip, wie oben beschrieben, außer daß die Peptidtrennung und -ionisierung bei signifikant höherer Empfindlichkeit durchgeführt werden. [Fig. 8](#) zeigt eine schematische Darstellung des Schlüsselkonstruktionselements. Die Konstruktion des Systems und seine Betriebsweise ist veröffentlicht worden. Peptide, die aus Proteinauflösungen stammen, werden durch SPE konzentriert, durch CE getrennt und durch ESI-MS/MS analysiert. Die resultierenden nicht-ausgewerteten CID-Spektren werden verwendet, um Sequenzdatenbanken mit dem Sequest-Software-System zu erforschen. Die SPE-Extraktionsvorrichtung ist eine kleine Umkehrphasen-Chromatographiesäule mit Dimensionen von 0,18 × 1 mm, die direkt in eine Quarzglasstrennungskapillare gepackt wird. Peptide, die in einer Probelösung enthalten sind, werden adsorbiert und auf der SPE-Vorrichtung konzentriert, in geschätzten 100 bis 300 nl organischen Lösungsmittels eluiert und außerdem durch elektrophoretisches Stapeln und/oder Isotachophorese auf ein geschätztes Volumen von 5 bis 30 nl konzentriert. Die Peptide werden dann durch CE in einer Kapillare mit einem Innendurchmesser von 20 µm oder 50 µm getrennt und direkt durch ESI ionisiert, da sie die Kapillare verlassen (siehe Referenz 13 zur Konstruktion der Mikrosprühionisierungsquelle). Mit diesem System können Peptidmassen bei der Empfindlichkeit von 660 Attomol (ungefähr 500 fg für einen Peptidrest von 20) bei einer Konzentrationsgrenze von 33 amol/µl bestimmt werden, und diese Proteine können durch die CID-Spektren von automatisch ausgewählten Peptiden bei weniger als 10 fmol (0,5 ng für ein Protein von 50 kDa) der Probe bei einer Konzentrationsgrenze von weniger als 300 amol/µl identifiziert werden. Diese Technik wird für die Analyse bei sehr hoher Empfindlichkeit der Peptidproben, die durch die Experimente erzeugt wurden, verwendet. Es ist ebenso gezeigt worden, daß die Analysezeit, die für automatisierte CID-Experimente verfügbar ist, durch die datenabhängige Modulation der CE-Spannung signifikant ausgedehnt werden kann. Wenn mehrere Peptidionen gleichzeitig bei der MS nachgewiesen werden, wird die CE-Spannung automatisch verringert. Dies führt zur Reduktion des elektroosmotischen Flusses aus der Kapillare und daher zur Verlängerung des Zeitraums, der zur Auswahl der Peptidionen für CID verfügbar ist. Die Nettowirkung dieser Peak-parkenden Technik ist eine Erweiterung des dynamischen Bereiches der Technik, da die verfügbare erhöhte Zeit zum CID von Ionen mit einem geringen Ionenstrom verwendet wird. Wenn alle Peptidionen analysiert werden, wird die Elektrophorese automatisch durch Erhöhen der CE-Spannung auf den ursprünglichen Wert erneut beschleunigt.

Tabelle 1: Relative überflüssige Quantifizierung von α-Lactalbuminhäufigkeit (nach dem Mischen mit bekannter Menge desselben Proteins mit Cysteinen, modifiziert mit isotopenschwerem Biotinylierungsreagenz)

Peptid #	m/z (leicht)	Ladungszustand	Peptidsequenz	Verhältnis (schwer : leicht)
1	518,4	2+	(K) IWCK	2,70
2	568,4	2+	(K) ALCSEK	2,68
3	570,4	2+	(K) CEVFR	2,90
4	760,5	2+	(K) LDQWLCEK	2,82
5	710,1	3+	(K) FLDDDLTDDIMCVK	2,88
6	954,2	3+	(K) DDQNP HSSNICNISCDK	2,90
7	1286,9	4+	(K) GYGGVSLPEWVCTTFHTSGY DTQAIVQNNDSTEYGLFQINNK	NA ^a

^aIsotopenverhältnis wurde nicht analysiert, da auf einem 4⁺- Peptid die Isotopmuster stark überlappend waren aufgrund der Unterschiede von nur 2 amu zwischen schweren und leichten Ionen.

Tabelle 2: Sequenzidentifizierung und Quantifizierung der Komponenten eines Proteingemisches in einer Einzelanalyse

Genname*	identifizierte Peptid-sequenz	beobachtetes Verhältnis (d0/d8) ⁺	Mittelwert ± SD	erwartetes Verhältnis (d0/d8) ⁻	% Fehler
LCA_BOVIN	ALC#SEK	0,94	0,96±0,06	1,00	4,2
	C#EVFR	1,03			
	FLDDLTDIMC#VK	0,92			
OVAL_CHICK	ADHPFLFC#IK	1,88	1,92±0,06	2,00	4,0
	YPILPEYLQC#VK	1,96			
BGAL_ECOLI	LTAAC#FDR	1,00	0,98±0,07	1,00	2,0
	IGLNC#QLAQVAER	0,91			
	IIFDGVNSAFHLWC#NGR	1,04			
LACB_BOVIN	WENGEC#AQK	3,64	3,55±0,13	4,00	11,3
	LSFNPTQLEEQC#HI	3,45			
G3P_RABIT	VPTPNVSVVDLTC#R	0,54	0,56±0,02	0,50	12,0
	IVSNASC#TTNC#LAPLAK	0,57			
PHS2_RABIT	IC#GGWQMEEADDWLR	0,32	0,32±0,03	0,33	3,1
	TC#AYTNHTVLPEALER	0,35			
	WLVLC#NPGLAEIIAER	0,30			

*Gennamen nach Swiss Prot-Nomenklatur (www.expasy.ch)

⁺Verhältnisse wurden für jedes Peptid berechnet, wie in **Fig. 3** gezeigt.

⁻Erwartete Verhältnisse wurden aus den bekannten Mengen an Proteinen, die in jedem Gemisch vorliegen, berechnet.

ICAT-markierter Cysteinylrest.

Tabelle 3: Proteinprofile aus Hefe, die auf Galaktose oder Ethanol als Kohlenstoffquelle wächst.

Genna- me*	identifizierte Peptidsequenz	beobachtetes Verhältnis ⁺ (Eth:Gal) ⁼	Galaktose- unterdrückt [§]	Glukose- unterdrückt [§]
ACH1	KHNC#LHEPHMLK	> 100 : 1	✓	
ADH1	YSGVC#HTDLHAWHGDWPLPVK	0,57 : 1		
	C#C#SDVFNQVVK	0,48 : 1		
ADH2	YSGVC#HTDLHAWHGDWPLPTK	> 200 : 1	✓	✓
	C#SSDVFNHVVK	> 200 : 1		
ALD4	TFEVINPSTEEEIC#HIYEGR	> 100 : 1	✓	✓
BMH1	SEHQVELIC#SYR	0,95 : 1		
CDC19	YRPNC#PIILVTR	0,49 : 1		
	NC#TPKPTSTTETVAASAVAAVFEQK	0,65 : 1		
	AC#DDK	0,67 : 1		
FBA1	SIAPAYGIPVVLHSDHC#AK	0,60 : 1		
	EQVGC#K	0,63 : 1		
GAL1	LTGAGWGGC#TVHLVPGGPNGNIEK	1 : > 200		✓
GAL10	HHIPFYEV DLC#DR	1 : > 200		✓
	DC#VTLK	1 : > 200		
GCY1	LWC#TQHHPEVALDQSLK	0,34 : 1		✓
GLK1	IC#SVNLHGDHTFSMEQMK	0,65 : 1		
GPD1	IC#SQLK	0,54 : 1		✓
ICL1	GGTQC#SIMR	> 100 : 1		✓
IPP1	NC#FPHHGYIHNYGAFPQTWEDPNVS- HPETK	0,76 : 1		
LPD1	VC#HAHPTLSEAFK	1,30 : 1		✓
PEP4	KGWTGQYTLDC#NTR	2,60 : 1		✓
PSA1	SVVLC#NSTIK	0,56 : 1		
PGM2	C#TGGIILTASHNPGGPENDMGIK	0,58 : 1		✓
	LSIC#GEESFGTGSNHVR	0,62 : 1		
PCK1	C#PLK	1,59 : 1		✓
	IPC#LADSHPK	1,47 : 1		
	C#INLSAEKEPEIFDAIK	1,52 : 1		
	C#AYPIDYIPSAK	1,41 : 1		
	IVEEPTSKDEIWWGPVNKPC#SER	1,85 : 1		
QCR6	ALVHHYEEC#AER	1,30 : 1		✓

RPL1A°	SC#GVDAMSVDDLKK	0,82 : 1	
SAH1	HPEMLEDC#FGLSEETTTGVHHLR	0,62 : 1	
	EC#INIKPQVDR	0,74 : 1	
SOD1	GFHIHEFGDATNGC#VSAGPHFNPFK	0,46 : 1	✓
TEF1	RGNVC#GDAK	0,81 : 1	
	C#GGIDK	0,70 : 1	
	FVPSKPMC#VEAFSEYPPLGR	0,74 : 1	
VMA2	IPIFSASGLPHNEIAAQIC#R	0,70 : 1	
YHB1	HYSLC#SASTK	0,69 : 1	

*Gennamen nach Yeast Proteom-Datenbank (YPD) (19).

Cysteinylnrest ist ICAT-markiert.

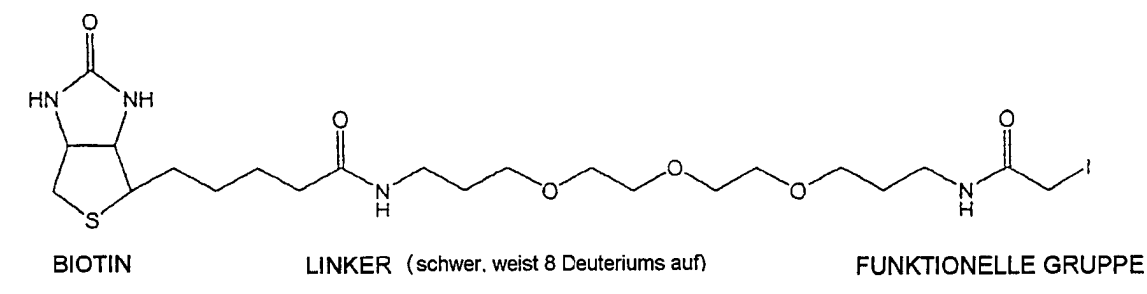
*Proteinexpressionsverhältnisse wurden, wie in **Fig. 3** beschrieben, berechnet.

*Kohlenstoffquelle für Hefewachstum war 2% Ethanol (Eth) oder 2% Galaktose (Gal).

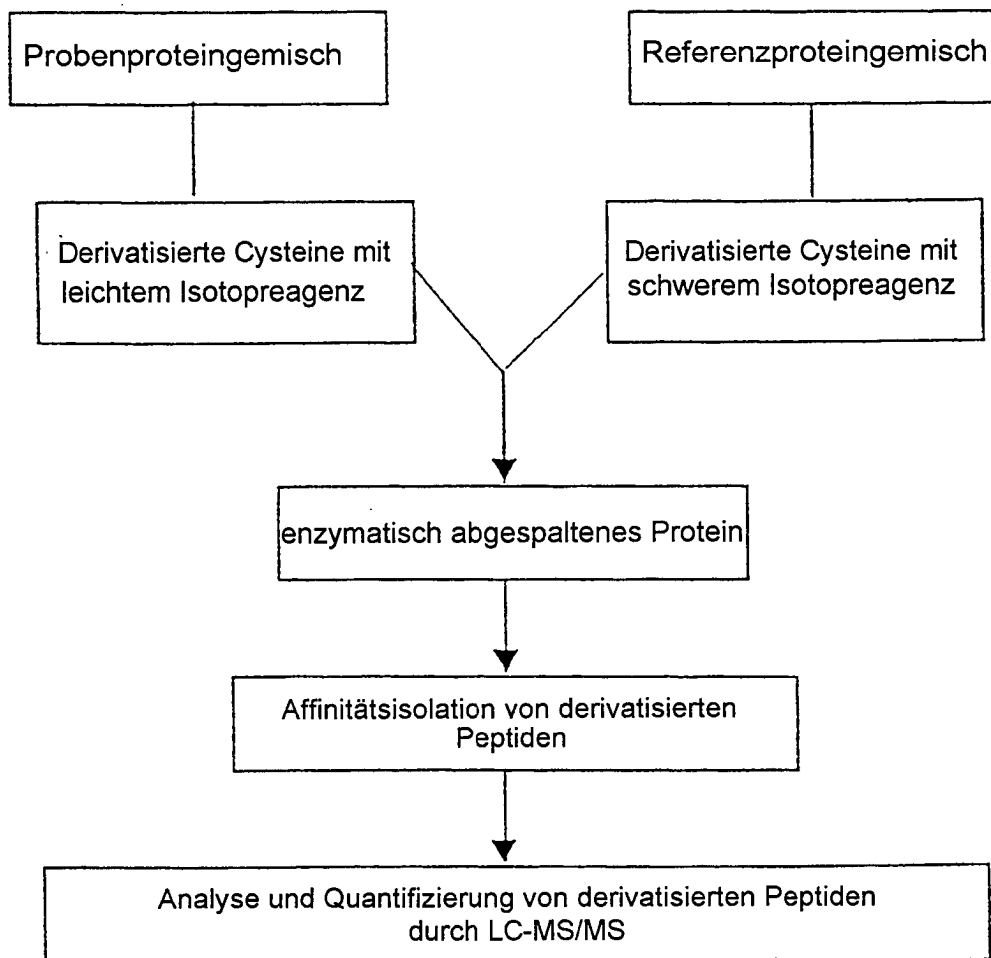
§Gen ist bekannt, Galaktose- oder Glukose-unterdrückt zu sein (19).

°Acht andere ribosomale Proteine wurden bei ähnlichen Genexpressionsniveaus nachgewiesen.

Schema 1

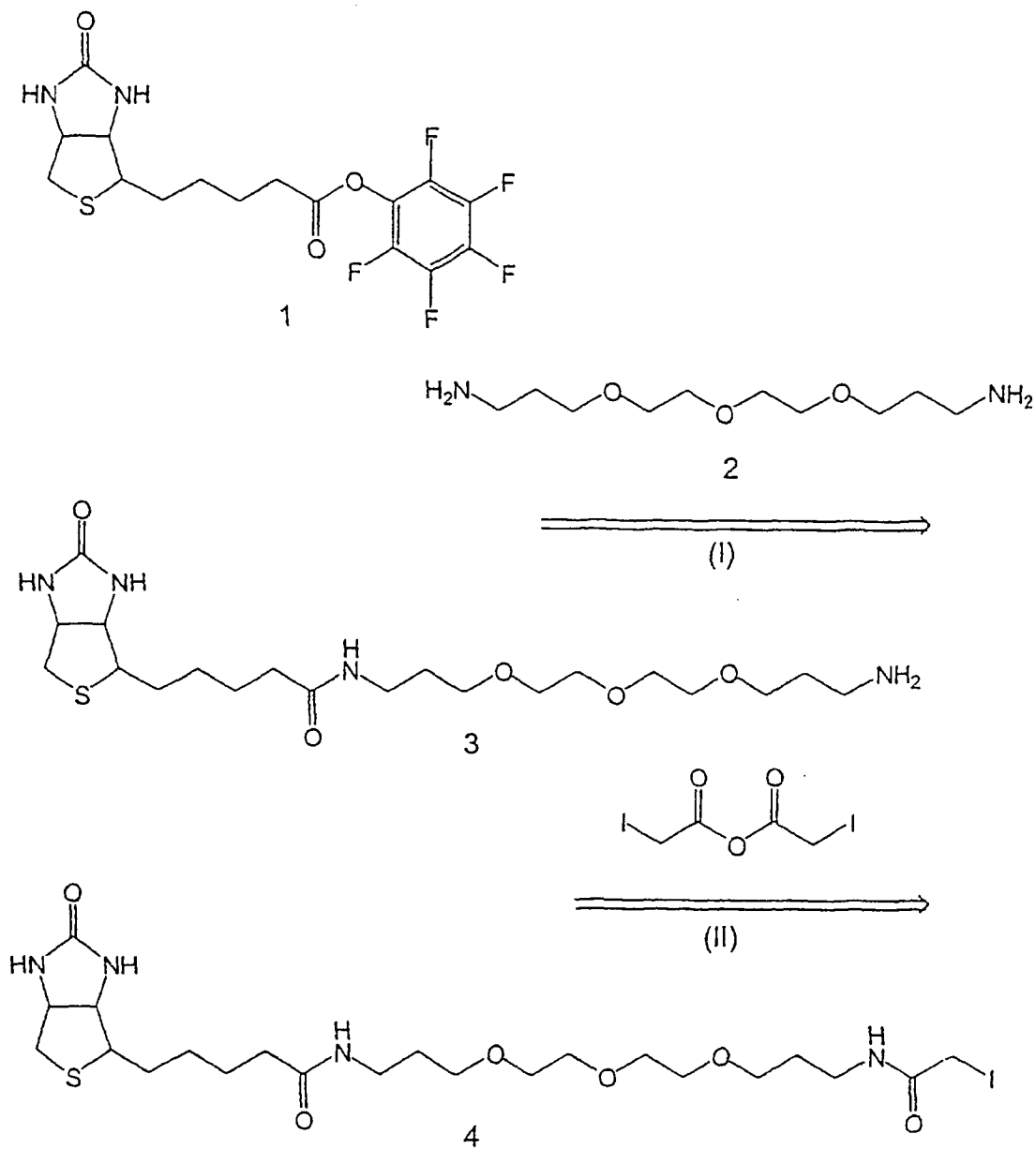


Fließschemadarstellung für die Verfahrensweise für quantitative Proteinprofilmessungen, basierend auf stabilen Isotopenverhältnissen durch Massenspektrometrie



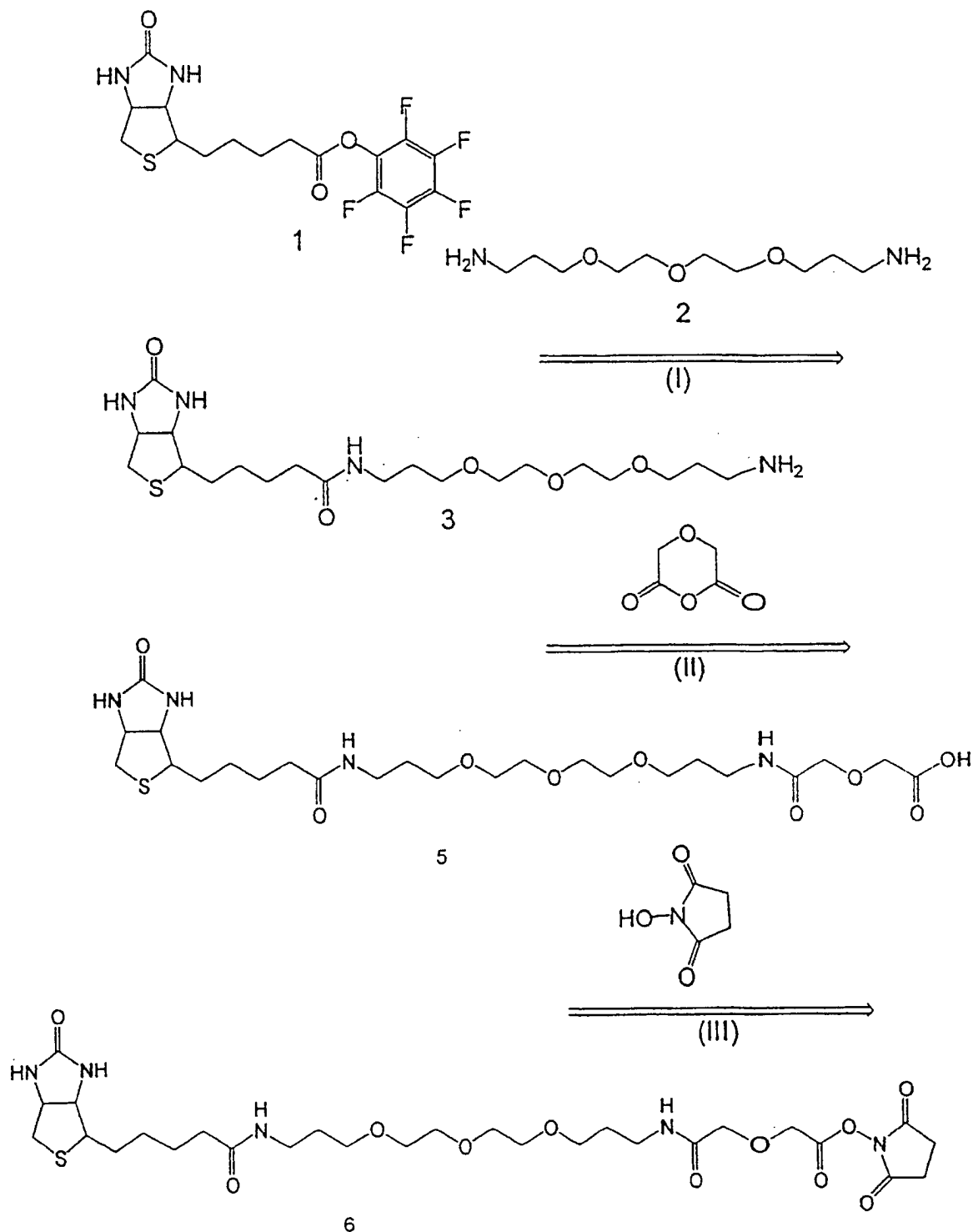
Schema 2

Gewöhnliche chemische Synthese von Sulfhydryl-reaktivem Biotinylierungsreagenz. Der Linker kann entweder isotopenschwer (d8) oder -leicht (d0) hergestellt werden.



Schema 3

Synthese eines Reagens, spezifisch für Biotinylierung von freien Aminogruppen. Das Reagenz wird sowohl in isotopenschwerer (d8) als auch -leichter (d0) Form synthetisiert.



Referenzen

- Ashikaga, K. et al. (1988) Bull. Chem. Soc. Jpn. 61: 2443–2450.
 Bayer, E. und Wilchek, M. (1990) Methods Enzymol. 184: 49–51.
 Bleasby, A. J. et al. (1994), "OWL-a non-redundant composite protein sequence database," Nucl. Acids Res. 22: 3574–3577.
 Boucherie, H. et al. (1996), "Two-dimensional gel protein database of *Saccharomyces cerevisiae*," Electrophoresis 17: 1683–1699.

- Chef, Y.-T. und Burchell, A. (1995), *The Metabolic and Molecular Bases of Inherited Disease*, Scriver, C. R. et al. (eds.) McGraw-Hill, New York, S. 935–966.
- Clauser, K. R. et al. (1995), "Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE," *Proc. Natl. Acad. Sci. USA* 92: 5072–5076.
- Cole, R. B. (1997) *Electrospray Ionization Mass Spectrometry Fundamentals, Instrumentation and Practice*, Wiley, New York.
- De Leenheer, A. P. und Thienpont, L. M. (1992), "Application of isotope dilution-mass spectrometry in clinical chemistry, pharmacokinetics, and toxicology," *Mass Spectrom. Rev.* 11: 249–307.
- DeRisi, J. L. et al. (1997), "Exploring the metabolic and genetic control of gene expression on a genomic scale," *Science* 278: 680–6.
- Dongre, A. R., Eng, J. K., und Yates, J. R., 3rd (1997), "Emerging tandem mass spectrometry techniques for the rapid identification of proteins," *Trends Biotechnol.* 15: 418–425.
- Ducret, A., VanOostveen, I., Eng, J. K., Yates, J. R., und Aebersold, R. (1998), "High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry," *Prot. Sci.* 7: 706–719.
- Eng, J., McCormack, A., und Yates, J. I. (1994), "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database," *J. Am. Soc. Mass Spectrom.* 5: 976–989.
- Figeys, D. et al. (1998), "Electrophoresis combined with mass spectrometry techniques: Powerful tools for the analysis of proteins and proteomes," *Electrophoresis* 19: 1811–1818.
- Figeys, D., und Aebersold, R. (1998), "High sensitivity analysis of proteins and peptides by capillary electrophoresis tandem mass spectrometry: Recent developments in technology and applications," *Electrophoresis* 19: 885–892.
- Figeys, D., Ducret, A., Yates, J. R., und Aebersold, R. (1996), "Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray-tandem mass spectrometry," *Nature Biotech.* 14: 1579–1583.
- Figeys, D., Ning, Y., und Aebersold, R. (1997), "A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry," *Anal. Chem.* 69: 3153–3160.
- Garrels, J. I., McLaughlin, C. S., Warner, J. R., Fitcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita, F.-R., und Payne, W. E. (1997), "Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis*," 18: 1347–1360.
- Gerber, S. A. et al. (1999), "Analysis of rates of multiple enzymes in cell lysates by electrospray ionization mass spectrometry," *J. Am. Chem. Soc.* 121: 1102–1103.
- Gygi, S. P. et al. (1999), "Correlation between protein and mRNA abundance in yeast," *Mol. Cell. Biol.* 19: 1720–1730.
- Gygi, S. P. et al. (1999), "Protein analysis by mass spectrometry and sequence database searching: tools for cancer research in the post-genomic era," *Electrophoresis* 20: 310–319.
- Haynes, P. A., Fripp, N., und Aebersold, R. (1998), "Identification of gel-separated proteins by liquid chromatography electrospray tandem mass spectrometry: Comparison of methods and their limitations," *Electrophoresis* 19: 939–945.
- Hodges, P. E. et al. (1999), "The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data," *Nucl. Acids Res.* 27: 69–73.
- Johnston, M. und Carlson, M. (1992), in *The Molecular and Cellular Biology of the Yeast *Saccharomyces**, Jones, E. W. et al. (eds.), Cold Spring Harbor Press, New York City, S. 193–281.
- Link, A. J., Hays, L. G., Carmack, E. B., und Yates, J. R., 3rd (1997), "Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143," *Electrophoresis* 18: 1314–1334.
- Link, J. et al. (1999), "Direct analysis of large protein complexes using mass spectrometry," *Nat. Biotech.* In press.
- Mann, M., und Wilm, M. (1994), "Error-tolerant identification of peptides in sequence databases by peptide sequence tags," *Anal. Chem.* 66: 4390–4399.
- Morris, A. A. M. und Turnbull, D. M. (1994) *Curr. Opin. Neurol.* 7: 535–541.
- Neufeld, E. und Muenzer, J. (1995), "The mucopolysaccharidoses" In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver, C. R. et al. (eds.) McGraw-Hill, New York, S. 2465–2494.
- Oda, Y. et al. (1999), "Accurate quantitation of protein expression and site-specific phosphorylation," *Proc. Natl. Acad. Sci. USA* 96: 6591–6596.
- Okada, S. und O'Brien, J. S. (1968) *Science* 160: 10002.
- Opitck, G. J. et al. (1997), "Comprehensive on-line LC/LC/MS of proteins," *Anal. Chem.* 69: 1518–1524.
- Pennington, S. R., Wilkins, M. R., Hochstrasser, D. F., und Dunn, M. J. (1997), "Proteome analysis: From protein characterization to biological function," *Trends Cell Bio.* 7: 168–173.
- Qin, J. et al. (1997), "A strategy for rapid, high-confidence protein identification," *Anal. Chem.* 69: 3995–4001.

- Ronne, H. (1995), "Glucose repression in fungi," *Trends Genet.* 11: 12–17.
- Scriber, C. R. et al. (1995), *The Metabolic and Molecular Bases of Inherited Disease*. McGraw-Hill, New York.
- Scriber, C. R. et al. (1995), *The Metabolic and Molecular Bases of Inherited Disease*, Scriber, C. R. et al. (eds.) McGraw-Hill, New York, S. 1015–1076.
- Sechi, S. und Chait, B. T. (1998), "Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification," *Anal. Chem.* 70: 5150–5158.
- Segal, S. und Berry, G. T. (1995), *The Metabolic and Molecular Bases of Inherited Disease*, Scriber, C. R. et al. (eds.), McGraw-Hill, New York, S. 967–1000.
- Romanowska, A. et al. (1994) *Methods Enzymol.* 242: 90–101.
- Roth, F. P. et al. (1998), "Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation," *Nat. Biotechnol.* 16: 939–945
- Shalon, D., Smith, S. J., und Brown, P. O. (1996), "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization," *Genome Res.* 6: 639–645.
- Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., und Mann, M. (1996), "Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels," *Proc. Natl. Acad. Sci. U. S. A.* 93: 14440–14445.
- Shevchenko, A., Wilm, M., Vorm, O., und Mann, M. (1996), "Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels," *Anal. Chem.* 68: 850–858.
- Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Jr., Hieter, P., Vogelstein, B., und Kinzler, K. W. (1997), "Characterization of the yeast transcriptome," *Cell* 88: 243–251.
- Wilbur, D. S. et al. (1997) *Bioconjugate Chem.* 8: 572–584.
- Yates, J. R. d., Eng, J. K., McCormack, A. L., und Schieltz, D. (1995), "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database," *Anal. Chem.* 67: 1426–1436.
- Brockhausen, I.; Hull, E.; Hindsgaul, O.; Schachter, H.; Shah, R. N.; Michnick, S. W.; Carver, J. P. (1989) Control of glycoprotein synthesis. *J. Biol. Chem.* 264, 11211–11221.
- Chapman, A.; Fujimoto, K.; Kornfeld, S. (1980) The primary glycosylation defect in class E Thy-1-negative mutant mouse lymphoma cells is an inability to synthesize dolichol-P-mannose. *J. Biol. Chem.* 255, 4441–4446.
- Freeze, H. H. (1998) Disorders in protein glycosylation and potential therapy. *J. Pediatrics* 133, 593–600.
- Freeze, H. H. (1999) Human glycosylation disorders and sugar supplement therapy. *Biochem. Biophys. Res. Commun.* 255, 189–193.
- Gerber, S. A.; Scott, C. R.; Turecek, F.; Gelb, M. H. (1999) Analysis of rates of multiple enzymes in cell lysates by electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* 121, 1102–1103.
- Glaser, L. (1966) Phosphomannomutase from yeast. In *Meth. Enzymol.* Vol. VIII, Neufeld, E. F.; Ginsburg, V. Eds; Academic Press: New York 1966, S. 183–185.
- Kaur, K. J.; Hingsgaul, O. (1991) A simple synthesis of octyl 3,6-di-O-(α -D-mannopyranosyl)- β -D-manopyranoside and its use as an acceptor for the assay of N-acetylglucosaminyltransferase I activity. *Glycoconjugate J.* 8, 90–94.
- Kaur, K. J.; Alton, G.; Hindsgaul, O. (1991) Use of N-acetylglucosaminyltransferases I and II in the preparative synthesis of oligosaccharides. *Carbohydr. Res.* 210, 145–153.
- Korner, C.; Knauer, R.; Holzbach, U.; Hanefeld, F.; Lehle, L.; von Figura, K. (1998) Carbohydrate-deficient glycoprotein syndrome type V: deficiency of dolichyl-P-Glc:Man9GlcNAc2-PP-dolichyl glucosyltransferase. *Proc Natl Acad Sci U. S. A.* 95, 13200–13205.
- McMurry, J. E.; Kocovsky, P. (1984) A method for the palladium-catalyzed allylic oxidation of olefins. *Tetrahedron Lett.* 25, 4187–4190.
- Paulsen, H.; Meinjohanns, E. (1992) Synthesis of modified oligosaccharides of N-glycoproteins intended for substrate specificity studies of N-acetylglucosaminyltransferases II–V *Tetrahedron Lett.* 33, 7327–7330.
- Paulsen, H.; Meinjohanns, E.; Reck, F.; Brockhausen, I. (1993) Synthese von modifizierten Oligosacchariden der N-Glycoproteine zur Untersuchung der Spezifität der N-Acetylglucosaminyltransferase II. *Liebigs Ann. Chem.* 721–735.
- Preiss, J. (1966) GDP-mannose pyrophosphorylase from *Arthrobacter*. In *Meth. Enzymol.* Vol. VIII, Neufeld, E. F.; Ginsburg, V. Eds; Academic Press: New York 1966, S. 271–275.
- Ronin, C.; Caseti, C.; Bouchilloux, C. (1981) Transfer of glucose in the biosynthesis of thyroid glycoproteins. I. Inhibition of glucose transfer to oligosaccharide lipids by GDP-mannose. *Biochim. Biophys. Acta* 674, 48–57.
- Ronin, C.; Granier, C.; Caseti, C.; Bouchilloux, S.; Van Rietschoten, J. (1981a) Synthetic substrates for thyroid oligosaccharide transferase. Effects of peptide chain length and modifications in the -Asn-Xaa-Thr-region. *Eur. J. Biochem.* 118, 159–164.
- Rush, J. S.; Wachter, C. J. (1995) Transmembrane movement of a water-soluble analogue of mannosylphosphoryldolichol is mediated by an endoplasmic reticulum protein. *J. Cell.* 8101. 130, 529–536.

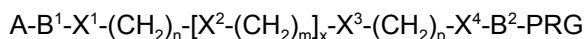
Schachter, H. (1986) Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochem. Cell Biol.* 64, 163–181.

Tan, J.; Dunn, J.; Jaeken, J.; Schachter, H. (1996) Mutations in the MGAT2 gene controlling complex glycan synthesis cause carbohydrate deficient glycoprotein syndrome type II, an autosomal recessive disease with defective brain development. *Am. J. Hum. Genet.* 59, 810–817.

Patentansprüche

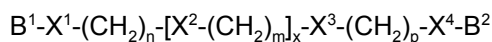
1. Verfahren zur Untersuchung relativer Expressionsniveaus von Proteinen in zwei oder mehr Proben, die Proteine enthalten, welches die Schritte umfaßt

(a) das Bereitstellen eines affinitätsmarkierten, im wesentlichen chemisch identischen und mit Isotopen unterschiedlich markierten, reaktiven Proteinreagenz für jede Probe, wobei das Reagenz die allgemeine Formel aufweist



worin A eine Affinitätsmarkierung ist, welche selektiv an ein Abfangreagenz bindet,

PRG eine reaktive Proteingruppe ist, welche selektiv mit bestimmten funktionellen Proteingruppen reagiert, und



eine Linkergruppe ist, welche mit stabilen Isotopen unterschiedlich markiert sein kann, und wobei

X^1 , X^2 , X^3 und X^4 unabhängig voneinander und X^2 unabhängig von anderem X^2 aus O, S, NH, NR, NRR'^+ , CO, COO, COS, S-S, SO, SO_2 , CO-NR', CS-NR', Si-O, Aryl- oder Diarylgruppen ausgewählt sein können, oder X^1 bis X^4 nicht vorhanden sein können,

B^1 und B^2 unabhängig voneinander optionale Gruppen sind, ausgewählt aus COO, CO, CO-NR', CS-NR', $(CH_2)_q$ -CONR', $(CH_2)_q$ -CS-NR', oder $(CH_2)_q$,

n, m, p, q und x ganze Zahlen sind, welche Werte von 0 bis 100 annehmen können, wobei die Summe von n + xm + p + q weniger als 100 ist,

R eine Alkyl-, Alkenyl-, Alkiny-, Alkoxy- oder eine Arylgruppe ist, welche gegebenenfalls mit einer oder mehreren Alkyl-, Alkenyl-, Alkiny- oder Alkoxygruppen substituiert ist, und

R' ein Wasserstoffatom, eine Alkyl-, Alkenyl-, Alkiny-, Alkoxy- oder eine Arylgruppe ist, welche gegebenenfalls mit einer oder mehreren Alkyl-, Alkenyl-, Alkiny- oder Alkoxygruppen substituiert ist,

wobei eine oder mehrere der CH_2 -Gruppen im Linker gegebenenfalls mit Alkyl-, Alkenyl-, Alkoxygruppen, einer Arylgruppe, welche gegebenenfalls mit einer oder mehreren Alkyl-, Alkenyl-, Alkiny- oder Alkoxygruppen, einer sauren Gruppe, einer basischen Gruppe oder einer Gruppe, welche eine bleibende positive oder negative Ladung trägt, substituiert ist, substituiert sein können, wobei eine oder mehrere Einfachbindungen, welche nicht benachbarte CH_2 -Gruppen im Linker verbinden, durch eine Doppel- oder Dreifachbindung ersetzt werden können, und wobei eines oder mehrere der Atome im Linker mit einem stabilen Isotop substituiert sein können, und

(b) das Reagieren jeder Probe mit einem der reaktiven Proteinreagenzien, um affinitätsmarkierte Proteine in der Probe bereitzustellen, wobei affinitätsmarkierte Proteine in verschiedenen Proben dadurch mit stabilen Isotopen unterschiedlich markiert sind,

(c) das Kombinieren der unterschiedlich markierten Proben und das Behandeln der kombinierten Probe, um die Proteine darin zu spalten und Peptide zu erzeugen,

(d) das Abfangen affinitätsmarkierter, unterschiedlich markierter Peptide der kombinierten Probe unter Verwendung des Abfangreagenz, welches selektiv A bindet,

(e) das Freisetzen abgefangener, affinitätsmarkierter, unterschiedlich markierter Peptide vom Abfangreagenz durch Unterbrechen der Wechselwirkung zwischen den affinitätsmarkierten Peptiden und dem Abfangreagenz, und

(f) das Nachweisen und Identifizieren der freigesetzten, affinitätsmarkierten unterschiedlich markierten Peptide durch Massenspektrometrie, und

(g) das Messen der relativen Häufigkeit der isotopisch verschiedenen Ionenpeaks, welche durch ein affinitätsmarkiertes, unterschiedlich markiertes Peptid erzeugt werden, um relative Expressionsniveaus des Proteins, von welchem das affinitätsmarkierte, unterschiedlich markierte Peptid stammt, zu untersuchen.

2. Verfahren nach Anspruch 1, wobei die affinitätsmarkierten Proteine in den Proben vor oder nach ihrem Abfangen enzymatisch oder chemisch bearbeitet werden, um sie in affinitätsmarkierte Peptide umzuwandeln.

3. Verfahren nach Anspruch 1, welches weiter den Schritt des Einführens einer bekannten Menge eines oder mehrerer interner Standards in eine Probe für jedes der zu quantifizierenden Proteine umfaßt, wobei der interne Standard ein Affinitätsmarkiertes Peptid ist, welches in der chemischen Struktur identisch zum Affinitätsmarkierten, zu quantifizierenden Peptid ist, mit der Ausnahme, daß der interne Standard entweder im Peptid- oder im Affinitätsmarkierungsanteil mit Isotopen unterschiedlich markiert ist, um es vom Affinitätsmarkierten, zu quantifizierenden Peptid zu unterscheiden.
4. Verfahren nach Anspruch 1, welches die relativen Mengen von Membranproteinen in einer oder mehreren unterschiedlichen Proben bestimmt.
5. Verfahren nach Anspruch 1, wobei unterschiedliche Proben Proteine, welche aus unterschiedlichen Organellen oder unterschiedlichen subzellulären Fraktionen stammen, enthalten.
6. Verfahren nach Anspruch 1, wobei die Affinitätsmarkierung Biotin oder ein modifiziertes Biotin ist.
7. Verfahren nach Anspruch 1, wobei PRG eine Sulfhydryl-reaktive Gruppe oder eine Amin-reaktive Gruppe ist.
8. Verfahren nach Anspruch 1, wobei PRG eine Sulfhydryl-reaktive Gruppe ist.
9. Verfahren nach Anspruch 1, wobei PRG eine Amin-reaktive Gruppe, eine Gruppe, welche mit einem Homoserinlaktat reagiert oder eine Gruppe, welche mit einer Carbonsäuregruppe reagiert, ist.
10. Verfahren nach Anspruch 1, wobei eines aus n, m, p oder q nicht 0 ist.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

Fig. 1A

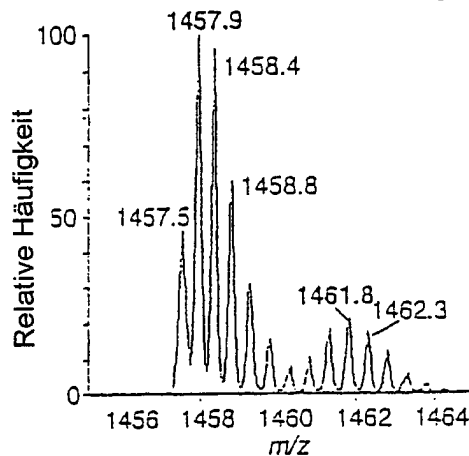
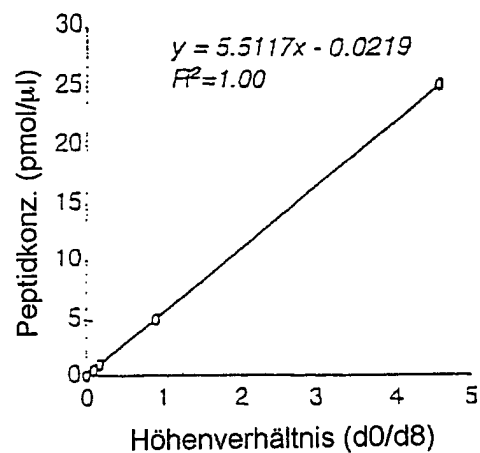
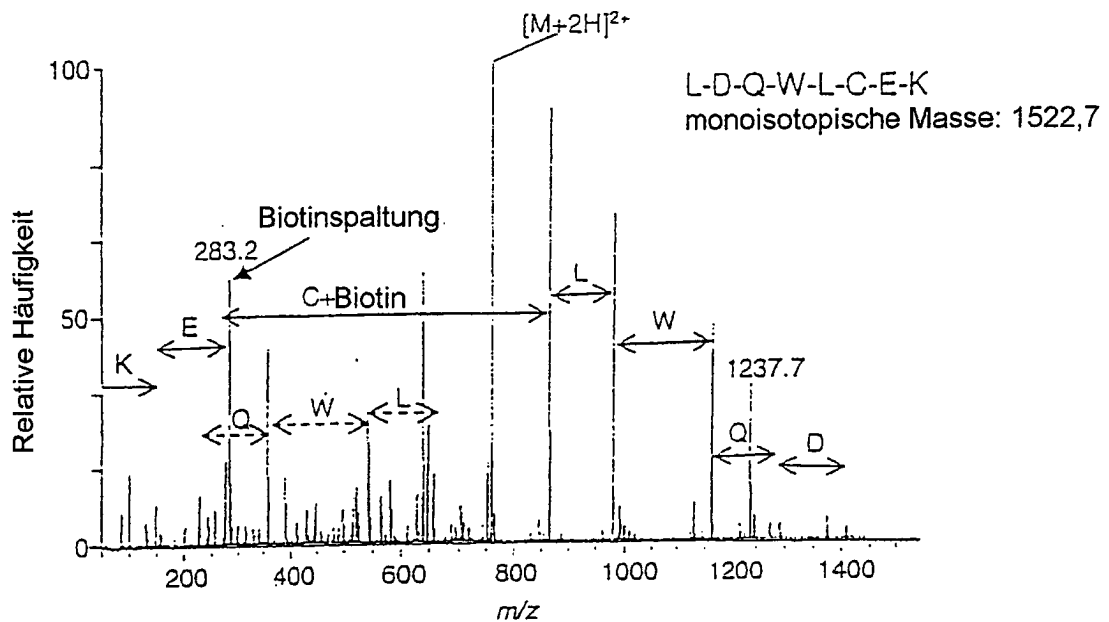


Fig. 1B



Standardkurve, erzeugt mit einem Cystein-biotinylierten Peptid, und Quantifizierung durch stabile Isotopenverdünnung. A) Zoom-Scann aus einem Ion-Trap-Massenspektrometer, das eine 4 amu Isotopverteilung für die $[M+2H]^{2+}$ -Ionen des Peptids zeigt, das mit den isotopeleichten (1457,9 u) und -schweren (1461,8) Biotinylierungsreagenzien modifiziert wurde. Das Verhältnis (d_0/d_8) betrug 4,54. B) Kurve, erzeugt aus der Analyse von Isotopenverhältnissen aus den Zoom-Scans von 5 unterschiedlichen Konzentrationen des d_0 -markierten Peptids, gemessen in Gegenwart einer bekannten Menge an Peptid, das mit dem isotopenschweren Reagenz markiert wurde.



Tandem-Massenspektrum eines Cystein-modifizierten Peptids aus α -Lactalbumin. Die Modifikation des Cysteinrestes mit dem gewöhnlich synthetisierten Biotinylierungsreagenz beeinflusste nicht die Fähigkeit des Sequest-Computerprogramms, dieses Peptid mit einer Datenbanksequenz genau zu vergleichen.

Fig. 2

Fig. 3A

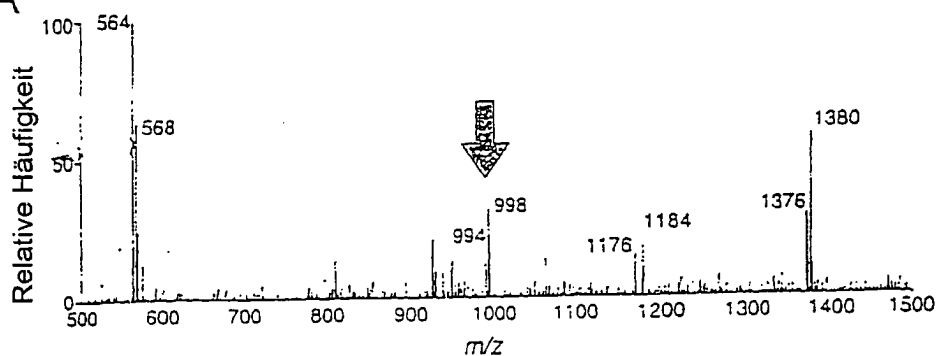


Fig. 3B

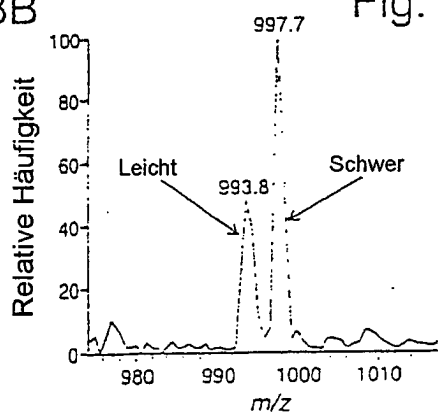
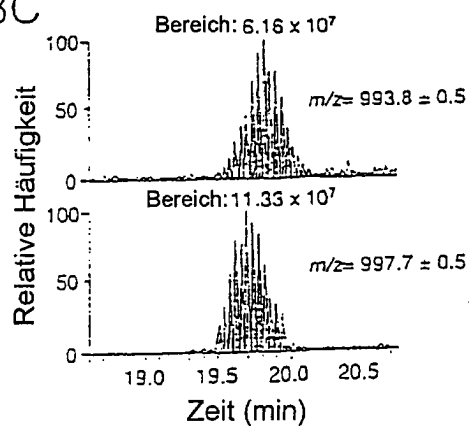


Fig. 3C



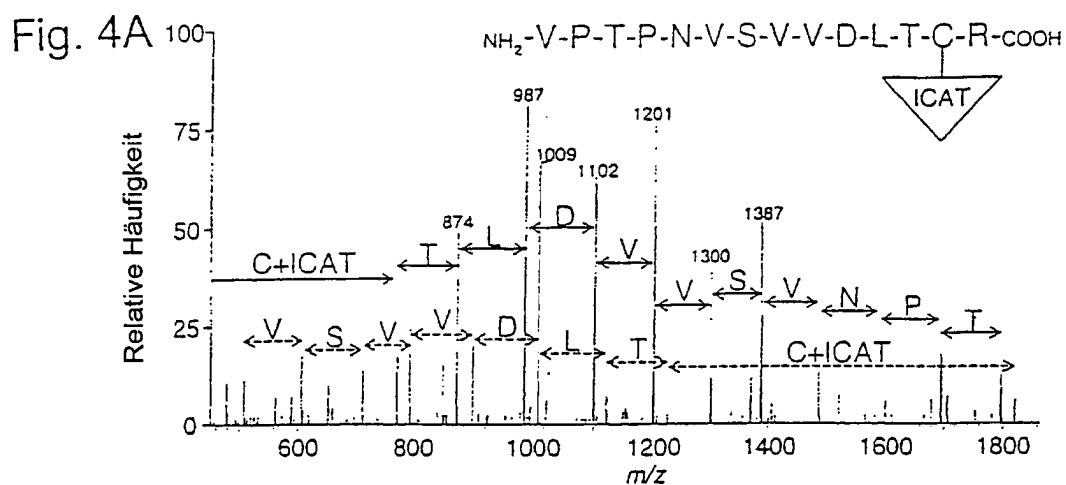
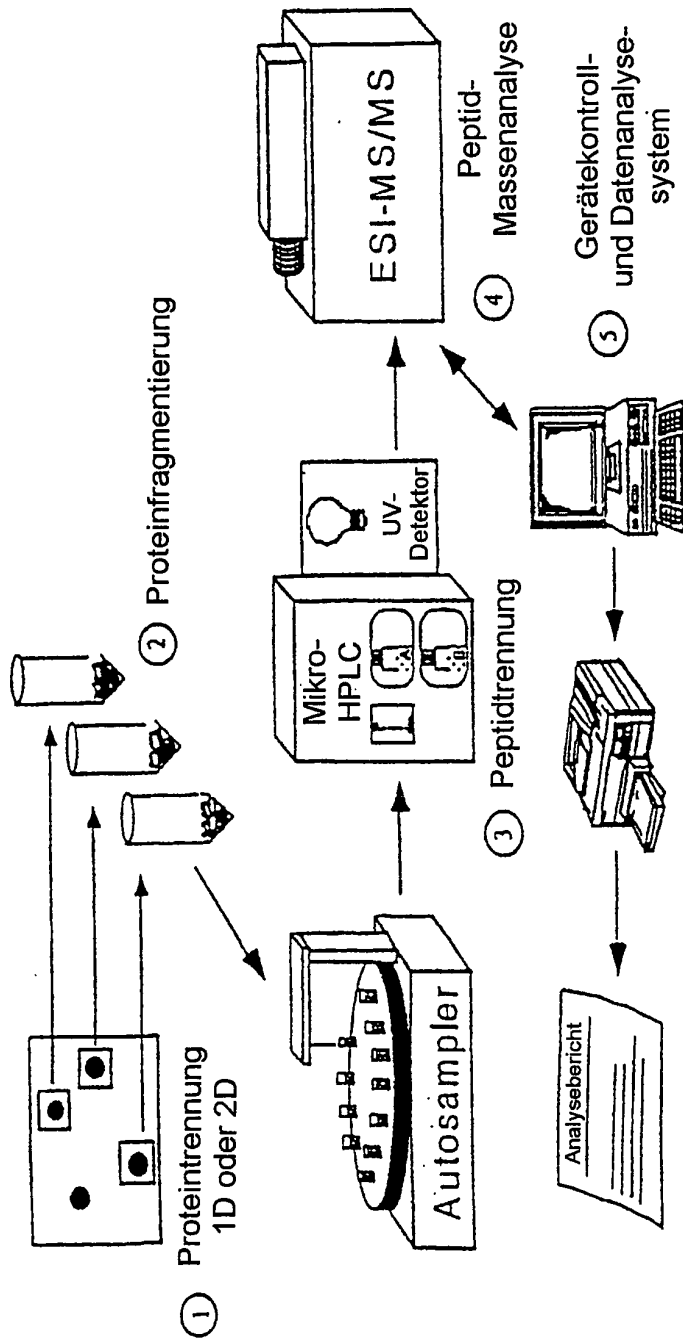


Fig. 4B

s0319_hlnavcid.0364.0364.2.out
 # amino acids = 93009033. # proteins = 190043. # matched peptides = 1973750
 C:\LCQ\database\owl.v11.3, (C# +494.50)

#	Rank/Sp	(M+H) ⁺ C*10 ⁻⁴	Ions	Reference	Peptide
1.	1 / 1	1994.3 4.4675	17.25	G3P_RABIT	(R) VPTPNVSVVDLTC#R
2.	3 / 403	1993.1 3.7366	13.34	SLC7RGL	(E) LGKEVLTANQVTEWGLR
3.	1 / 3	1995.0 2.6591	16.36	FLP_LACCA	(N) LANPNVYTTTLTAATVCTE
4.	4 / 209	1995.0 2.6235	14.35	A42912	(Y) LALLPSDREGPHGQFVTDK
5.	5 / 361	1993.1 2.4634	13.33	W59373	(E) ALLVLVAPMAAGNGEDLRN



Figur 7. Schematische Darstellung des automatisierten LC-MS/MS-Systems. Proteine werden typischerweise durch 1D- oder 2D-SDS-PAG

(1) getrennt. Proteinflecke oder -streifen werden ausgewählt, abgetrennt und proteolytisch mit Trypsin gespalten (2). Aufschlüsse werden in einen Autosampler geladen, der sie nacheinander zu dem Injektionsmechanismus eines englumigen HPLC-Systems (Michrom) führt. Der Säulengradient wird automatisch angewendet, um einzelne Peptide abzutrennen (3). Das Säuleneluat wird direkt in ein Massenspektrometer gesprüht, wo Sequenzinformationen aus den Peptiden gesammelt werden (4). Aufgezeichnete Peptidmassen und CID-Spektren werden zu einer Datenstation zur Sequest-Analyse übertragen, und eine Endzusammenfassung aller Identifikationen, die für alle ursprünglich geladenen Proben durchgeführt wurden, werden an einen Drucker (5) geschickt.

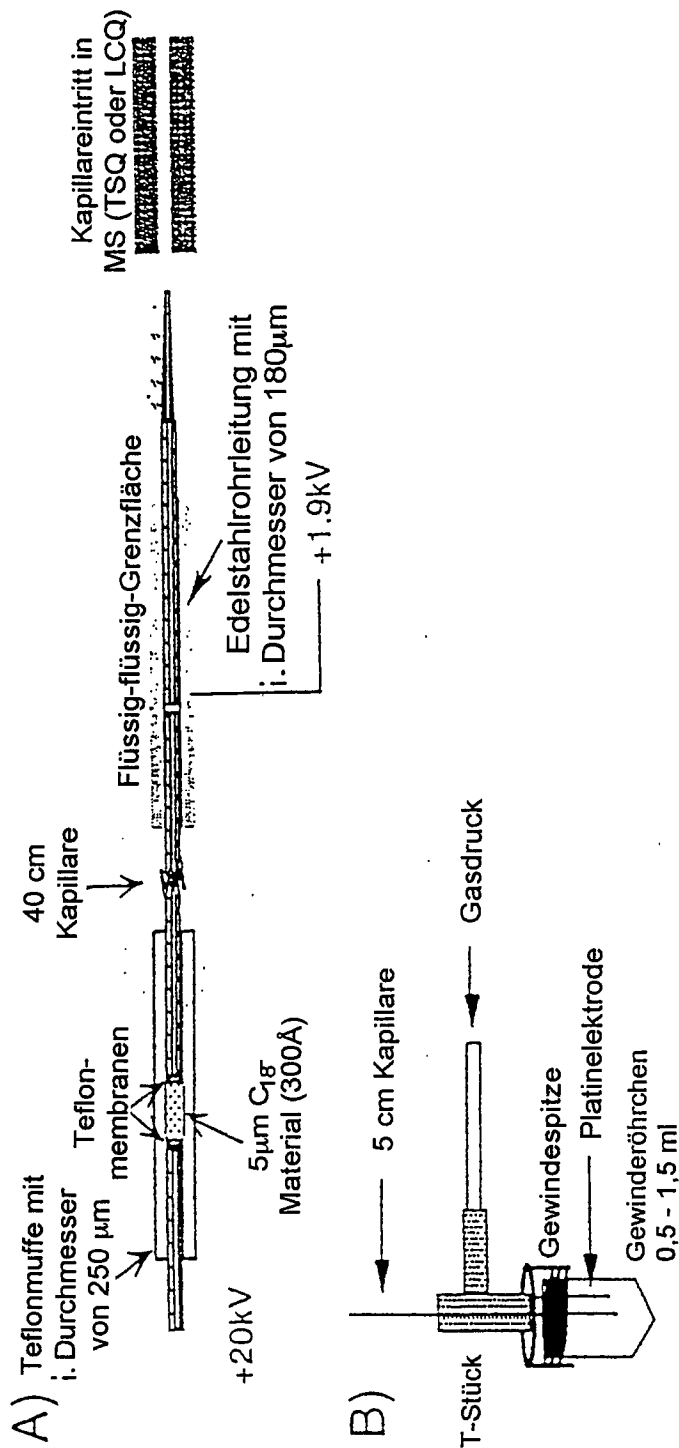


Fig.8. Schema des SPE-CE-MS/MS-Systems. A) Eine Quarzglas kapillare, typischerweise mit einem Innendurchmesser von 50 µm, wird an dem Elektrospühende mit einer flüssig-flüssig-Grenzfläche modifiziert, um elektrischen Kontakt mit den Analyten im Inneren der Kapillare herzustellen. Ungefähr 5 cm von dem Ende der Kapillare wird die SPE-Vorrichtung eingebracht. Diese besteht aus C₁₈-derivatisierten, großporigen Siliciumdioxidkügelchen, die im Inneren einer Teflonrohrleitung mit einem Innendurchmesser von 250 µm und mit Teflonmembranen an jedem Ende gepackt sind, um die Kügelchen an Ort und Stelle zwischen den zwei Quarzglas kapillaren zu halten. B) Das Injektionsende der Kapillare wird in einen versiegelten Behälter eingeführt, der bei einem konstanten, leicht hyperbarischen Druck gehalten wird, um konstanten Fluß zu gewährleisten. Eine Platinelektrode wird durch den Deckel in den Behälter eingeführt, um den elektrischen Kontakt zu gestatten.