



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) DE 699 27 343 T2 2006.06.22

(12)

Übersetzung der europäischen Patentschrift

(97) EP 1 015 643 B1

(51) Int Cl.⁸: C12Q 1/68 (2006.01)

(21) Deutsches Aktenzeichen: 699 27 343.9

(86) PCT-Aktenzeichen: PCT/US99/15705

(96) Europäisches Aktenzeichen: 99 935 503.5

(87) PCT-Veröffentlichungs-Nr.: WO 00/03038

(86) PCT-Anmeldetag: 09.07.1999

(87) Veröffentlichungstag
der PCT-Anmeldung: 20.01.2000

(97) Erstveröffentlichung durch das EPA: 05.07.2000

(97) Veröffentlichungstag
der Patenterteilung beim EPA: 21.09.2005

(47) Veröffentlichungstag im Patentblatt: 22.06.2006

(30) Unionspriorität:

112437 09.07.1998 US

(84) Benannte Vertragsstaaten:

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

(73) Patentinhaber:

Agilent Technologies Inc., A Delaware Corp., Palo
Alto, Calif., US

(72) Erfinder:

SAMPSON, R., Jeffrey, Burlingame, US; YAKHINI,
H., Zohar, 30900 Zikhron Ya'acov, IL; SAMPAS, M.,
Nicholas, San Jose, US; TSALENKO, Anna M.,
Chicago, US; MYERSON, Joel, Berkeley, US;
WEBB, G., Peter, Menlo Park, US

(74) Vertreter:

Schoppe, Zimmermann, Stöckeler & Zinkler, 82049
Pullach

(54) Bezeichnung: VERFAHREN UND MITTEL ZUR ANALYSE DER NUKLEOTIDSEQUENZ VON NUKLEINSÄUREN

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelebt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

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**Gebiet der Erfindung**

[0001] Diese Erfindung bezieht sich auf Verfahren und Reagenzien zum Analysieren von Nukleotidsequenzen von Nukleinsäuren mittels Massenspektrometrie und insbesondere auf Verfahren zum Analysieren von Nukleotidsequenzen unter Verwendung von Reagenzien, die Gemische von Oligonukleotid-Vorläufern sind, die ein hohes Maß an Massenzahlkomplexität und Sequenzbedeckungskomplexität aufweisen.

Hintergrund der Erfindung

[0002] Ein Bestimmen der Nukleotidsequenz von Nukleinsäuren (DNA und RNA) ist kritisch für das Verständnis der Funktion und Steuerung von Genen und ihrer Beziehung, beispielsweise in Bezug auf Krankheitserkennung und Krankheitsbewältigung. Die Analyse genetischer Informationen spielt bei biologischen Experimenten eine wesentliche Rolle. Dies gilt mittlerweile besonders in Bezug auf Studien, die auf ein Verständnis der grundlegenden genetischen und Umweltfaktoren abzielen, die mit Krankheit und den Auswirkungen potentieller therapeutischer Wirkstoffe auf die Zelle zusammenhängen. Diese Paradigmaverschiebung führte in der Branche der Biowissenschaften zu einem erhöhten Bedarf an sensibleren, präziseren und einen höheren Durchsatz ermöglichen Technologien zum Durchführen von Analysen an genetischem Material, das von einer Vielzahl biologischer Quellen gewonnen wird.

[0003] Da ein Sequenzieren der enorm hohen Anzahl von Nukleinsäuren in jeder menschlichen Zelle notwendigerweise ein zeitaufwendiger Vorgang ist, besteht immer ein dringender Bedarf an schnelleren Analysen mit einem höheren Durchsatz, die nicht zu Lasten der Empfindlichkeit und Genauigkeit gehen. Es wurden bereits eine Reihe von Techniken entwickelt, einschließlich, unter anderem, Elektrophorese, enzymatische und chemische Analyse, Arraytechnologie und Massenspektrometrie, um die Nukleotidsequenz von Nukleinsäuren zu bestimmen.

Elektrophoretische Techniken

[0004] Platten- oder Kapillarpolyacrylamidgel-Elektrophoresetechnologien wie z. B. diejenigen, die bei automatisierten DNA-Sequenzierern eingesetzt werden, liefern äußerst genaue De-Novo-Sequenzinformationen für relativ lange (500 – 700 Reste oder Basen) DNA-Segmente. Obwohl auf Elektrophorese beruhende Techniken viele Informationen pro Probe liefern, erfordern sie lange Probenpräparations- und aufbauzeiten und begrenzen somit den Durchsatz.

Enzymatische und chemische Analyse

[0005] Es gibt eine Reihe von enzymatischen und chemischen Techniken, um die De-Novo-Nukleotidsequenz von Nukleinsäuren zu bestimmen. Jedoch weist jede Technik inhärente Beschränkungen auf. Zum Beispiel offenbaren Maxam und Gilbert [Proc. Natl. Acad. Sci. USA 74: 5.460 (1977)] einen Lösungsansatz einer chemischen Qualitätsverringerung, und Sanger u. a. [Proc. Natl. Acad. Sci. USA 74: 5.463 (1977)] offenbaren ein Kettenabbruchverfahren, das eine Komplementärstrang-Primer-Verlängerung bzw. -Extension bzw. -Ausdehnung verwendet. Jede dieser Techniken verwendet vier getrennte Reaktionsgemische, um einen verschachtelten Satz von Fragmenten zu erzeugen, die sich in ihrer Länge durch ein einziges Nukleotid unterscheiden, wodurch somit eine vollständige Nukleotidsequenz darstellt wird. Eine Auflösung der Fragmente auf der Grundlage ihrer Größe und ihres entstehenden Nukleotids wird durchgeführt, um die Reihenfolge der Fragmente und somit die Nukleotidsequenz zu bestimmen.

[0006] Die Einsträngige-Konformation-Polymorphie-Analyse (SSCP-Analyse, SSCP = single-stranded conformation polymorphism) ist eine nützliche Technik zum Erfassen relativ geringer Unterschiede zwischen ähnlichen Sequenzen. Die Technik ist einfach zu implementieren und kann, wenn sie mit Multifarbstofferfassungs- oder Massenmarkierungsmethodologien kombiniert wird, multiplexiert werden und dadurch den Durchsatz verbessern. Wie auch Techniken, die sich auf ein Erfassen von Heteroduplexen stützen, z. B. Denaturierungs-Gradientengel-Elektrophorese (DGGE – denaturing gradient gel electrophoresis), chemische Spaltung (CCM – chemical cleavage), enzymatische Spaltung (unter Verwendung von Cleavase) von Nichtübereinstimmungen sowie Denaturierungs-Hochleistungs-Flüssigchromatographie (DHPLC – denaturing high performance liquid chromatography), ist die Technik nur qualitativ, d. h. die Technik offenbart lediglich, ob in der anvisierten bzw. Targetnukleinsäure eine Mutation vorliegt, liefert jedoch minimale Informationen über die Identität und Lage der Mutation.

[0007] Andere Techniken, die Ligase- und Polymeraseverlängerungsuntersuchungen verwenden, sind nützlich darin, zu bestimmen, ob an einer definierten Stelle in einer ansonsten bekannten anvisierten bzw. Targetnukleinsäuresequenz eine Mutation vorliegt. Beispielsweise offenbart die U.S.-Patentschrift Nr. 4,988,617 ein Verfahren zum Bestimmen, ob an einer definierten Stelle in einer ansonsten bekannten anvisierten bzw. Targetnukleinsäuresequenz eine Mutation vorliegt, indem eine Untersuchung bezüglich der Ligation zweier natürlicher Oligonukleotide durchgeführt wird, die dazu entworfen sind, nebeneinander entlang der anvisierten bzw. Targetsequenz zu hybridisieren. Die U.S.-Patentschrift Nr. 5,494,810 offenbart ein Verfahren, das eine thermostabile Ligase und die Ligasekettenreaktion (LCR) verwendet, um spezifische Nukleotidsubstitutionen, -deletionen, -insertionen und -translokationen in einer ansonsten bekannten Nukleinsäuresequenz unter Verwendung lediglich natürlicher Nukleinsäuren zu erfassen. Die U.S.-Patentschrift Nr. 5,403,709 offenbart ein Verfahren zum Bestimmen der Nukleotidsequenz durch Verwendung eines anderen Oligonukleotids als Verlängerung und eines dritten, Brücken bildenden Oligonukleotids, um die ersten zwei zum Zweck einer Ligation zusammenzuhalten, und die WO 97/35033 offenbart Verfahren zum Bestimmen der Identität eines Nukleotid-3' zu einem definierten Primer unter Verwendung einer Polymeraseverlängerungsuntersuchung. Obwohl die Untersuchungen mit einem relativ hohen Durchsatz durchgeführt werden können, sind sie sequenzspezifisch und erfordern somit für jedes zu analysierende Target einen unterschiedlichen Satz von Reagenzien.

[0008] Die U.S.-Patentschriften Nrn. 5,521,065, 4,883,750 und 5,242,794 (Whiteley u. a.) offenbaren Verfahren zum Testen des Vorliegens oder des Nichtvorliegens einer Targetsequenz in einem Gemisch aus Einsträngige-Nukleinsäure-Fragmenten. Das Verfahren beinhaltet, dass ein Gemisch aus Einsträngige-Nukleinsäure-Fragmenten mit einer ersten Sonde, die zu einer ersten Region der Targetsequenz komplementär ist, und mit einer zweiten Sonde, die zu einer zweiten Region der Targetsequenz komplementär ist, zur Reaktion gebracht wird. Die erste und die zweite Targetregion grenzen aneinander an. Es werden Hybridisierungsbedingungen verwendet, bei denen die zwei Sonden auf stabile Weise zu ihren zugeordneten Targetregionen hybridisiert werden. Im Anschluss an die Hybridisierung wird bzw. werden die erste und/oder die zweite Sonde, die zu angrenzenden ersten und zweiten Targetregionen hybridisiert werden, ligiert, und anschließend wird die Probe auf das Vorliegen eines erwarteten Sondenligationsprodukts getestet.

Arraytechnologie

[0009] Techniken, die eine Hybridisierung zu oberflächengebundenen DNA-Sondenarrays verwenden, sind nützlich beim Analysieren der Nukleotidsequenz von Targetnukleinsäuren. Diese Techniken stützen sich auf die inhärente Fähigkeit von Nukleinsäuren, über eine Bildung von Wasserstoffbrücken gemäß Basenpaarungsregeln von Watson-Crick Duplexe zu bilden. In der Theorie, und in gewissem Umfang auch in der Praxis, kann eine Hybridisierung zu oberflächengebundenen DNA-Sondenarrays in einem einzigen Experiment relativ viele Informationen liefern. Beispielsweise hat die Arraytechnologie einzelne Nukleotidpolymorphien innerhalb relativ langer (aus 1.000 Resten oder Basen bestehender) Sequenzen identifiziert (Kozal, M., u. a., *Nature Med.* 7: 753 – 759, Juli 1996). Ferner ist die Arraytechnologie für manche Arten der Genexpressionsanalyse nützlich, wobei man sich auf eine komparative Analyse von komplexen Gemischen von mRNA-Targetsequenzen stützt (Lockart, D., u. a., (1996), *Nat. Biotech.* 14, 1.675 – 1.680). Obwohl Arraytechnologien den Vorteil bieten, dass sie einigermaßen empfindlich und genau sind, wenn sie für spezifische Anwendungen und für spezifische Sätze von Targetsequenzen entwickelt werden, lassen sie eine generische Implementierung vermissen, die gleichzeitig auf mehrere bzw. unterschiedliche Anwendungen und Targets angewendet werden kann. Dies ist zum großen Teil auf das Erfordernis relativ langer Sondensequenzen zurückzuführen, die die Sonden-/Targetduplexe bilden und anschließend erfassen müssen. Überdies macht es diese Verwendung relativ langer Sonden aufgrund des inhärent geringen thermodynamischen Unterschieds zwischen dem perfekten Komplement und der einzigen Nichtübereinstimmung in dem Sonden-/Targetduplex schwierig, einzelne Nukleotidunterschiede abzufragen. Außerdem hängt eine Erfassung von Lösungsdiffusionseigenschaften und einer Wasserstoffbrückenbindung zwischen komplementären Target- und Sondensequenzen ab.

Massenspektrometrie-Techniken

[0010] Die Massenspektrometrie (MS) ist ein leistungsstarkes Hilfsmittel zum Analysieren von komplexen Gemischen aus Verbindungen, einschließlich Nukleinsäuren. Zusätzlich dazu, dass eine intakte Masse präzise bestimmt wird, können mittels mehrerer unterschiedlicher MS-Strategien Primärstrukturinformationen erhalten werden. Die Verwendung von MS zur DNA-Analyse ist potentiell auf die Erfassung von DNA-Modifikationen, auf eine DNA-Fragmentmassebestimmung und auf eine DNA-Sequenzierung anwendbar (siehe z. B. Fields, G.B., *Clinical Chemistry* 43, 1.108 (1997)). Zur Identifizierung von DNA-Modifikationsstellen wurden bereits sowohl eine auf einer Bombardierung mit schnellen Atomen (FAB – fast atom bombardment) als auch eine auf Elektrosprayionisierung (ESI – electrospray ionization) beruhende Stoßinduzierte-Dissoziations-/Tandem-MS

angewendet.

[0011] Obwohl die MS ein leistungsstarkes Hilfsmittel zum Analysieren von komplexen Gemischen verwandter Verbindungen, einschließlich Nukleinsäuren, ist, ist ihre Nützlichkeit beim Analysieren der Sequenz von Nukleinsäuren durch verfügbare Ionisierungs- und Erfassungsverfahren begrenzt. Beispielsweise erzeugt die ESI-Spektrometrie eine Verteilung von stark geladenen Ionen, deren Masse-zu-Ladung-Verhältnis im Bereich von im Handel erhältlichen Quadrupolmassenanalysatoren liegt. Obwohl ESI empfindlich ist und lediglich Femtomol-Mengen einer Probe benötigt, stützt sie sich auf mehrere Ladungen, um eine effiziente Ionisierung zu erzielen, und erzeugt sogar für einfache Nukleinsäuren komplexe und schwierig zu interpretierende mehrfach geladene Spektren.

[0012] Eine matrixunterstützte Laser-Desorption-Ionisierung (MALDI – matrix-assisted laser desorption ionization), die in Verbindung mit einem Flugzeit-Massenanalysator (TOF-Massenanalysator, TOF = time of flight) verwendet wird, weist aufgrund ihrer relativ großen Massenbandbreite, hohen Auflösung ($m/\Delta m \leq 1,0$ bei Masse 5.000) und Abtastrate (bis zu 1 Abtastwert/Sekunde) ein großes Potential zum Sequenzieren von Nukleinsäuren auf. Bei einem Aspekt bietet die MALDI insofern einen potentiellen Vorteil gegenüber ESI und FAB, als Biomoleküle einer großen Masse ohne weiteres ionisiert und analysiert werden können. Ferner erzeugt die MALDI im Gegensatz zur ESI vorwiegend einfach geladene Spezies.

[0013] Jedoch kann eine MALDI-Analyse einer DNA allgemein eine mangelnde Auflösung von DNA-Fragmenten von hoher relativer Molekülmasse, eine DNA-Instabilität und eine Interferenz seitens Probenherstellungsreagenzien aufweisen. Längere Oligonukleotide können breitere, weniger intensive Signale liefern, da die MALDI Ionen von höherer relativer Molekülmasse eine größere kinetische Energie verleiht. Obwohl sie zum Analysieren von Nukleinsäuren von hoher relativer Molekülmasse verwendet werden kann, bewirkt MALDI-TOF eine Spaltung der Nukleinsäurehauptkette, was das resultierende Spektrum weiter verkompliziert. Folglich ist die Länge von Nukleinsäuresequenzen, die derzeit mittels MALDI-TOF analysiert werden können, auf etwa 100 Basen oder Reste beschränkt. Wang u. a. (WO 98/03684) nutzten eine „Fragmentierung in der Quelle“ („in source fragmentation) und koppelten sie mit Verfahren einer verzögerten gepulsten Ionenextraktion zum Bestimmen der Sequenz von Nukleinsäureanalyten.

[0014] Es wurden bereits eine Reihe von Verfahren offenbart, die standardmäßige Sequenzierungsverfahren nutzen, um Targetfragmente zur Analyse mittels Massenspektroskopie zu erzeugen. Beispielsweise offenbaren die U.S.-Patentschrift Nr. 5,288,644 (Beavis u. a.); die U.S.-Patentschrift Nr. 5,547,835 (Koster) und die U.S.-Patentschrift Nr. 5,622,824 (Koster) Verfahren zum Bestimmen der Sequenz einer Targetnukleinsäure unter Verwendung von MALDI-TOF von Leitern des Targets, entweder mittels Exonukleaseaufschluss- oder mittels standardmäßiger Sanger-Sequenzierungsverfahren erzeugt. Beavis erörtert ein Verfahren zur DNA-Sequenzierung unter Verwendung verschiedener basenspezifischer Reaktionen, um verschiedene Sätze von DNA-Fragmenten zu verwenden, um ein Stück DNA einer unbekannten Sequenz zu bilden. Jeder der verschiedenen Sätze von DNA-Fragmenten weist einen gemeinsamen Ursprung auf und endet an einer bestimmten Base entlang der unbekannten Sequenz. Die relativen Molekülmassen der DNA-Fragmente in jedem der verschiedenen Sätze werden durch ein MALDI-Massenspektrometer bestimmt, das anschließend verwendet wird, um die Nukleotidsequenz der DNA abzuleiten.

[0015] Koster verwendet die Sanger-Sequenzierungsstrategie und setzt die Sequenzinformationen mittels einer Analyse der verschachtelten Fragmente, die durch einen basenspezifischen Kettenabbruch erhalten werden, mittels ihrer verschiedenen Molekülmassen unter Verwendung von Massenspektrometrie wie z. B. MALDI- oder ESI-Massenspektrometrie zusammen. Dieses Verfahren wurde bereits mit einem Festphasensequenzierungs-Lösungsansatz gekoppelt, bei dem die Matrize mit Biotin markiert und an mit Streptavidin beschichtete magnetische Kugelchen gebunden wird. Unter Verwendung dieses Verfahrens war es möglich, Exons 5 und 8 des p53-Gens unter Verwendung von 21 definierten Primern zu sequenzieren (Fu u. a., Nat. Biotechnol. 16, 381 (1998)). Der Durchsatz kann erhöht werden, indem Massenmodifikationen in den Oligonukleotid-Primer, in Kettenabbruch-Nukleosid-Triphosphate und/oder in die kettenverlängernden Nukleosid-Triphosphate eingebracht werden, sowie dadurch, dass integrierte Markierungssequenzen verwendet werden, die ein Multiplexieren, mittels Hybridisierung, von markierungsspezifischen Sonden mit massendifferenzierten relativen Molekülmassen ermöglichen (U.S.-Patentschrift Nr. 5,547,835). Es ist jedoch wichtig zu beachten, dass all diese Sequenzierungsverfahren entweder eine gewisse vorherige Kenntnis der Targetsequenz oder eine Einführung einer bekannten Sequenz, die als den Primer bindende Stelle dienen soll, erfordern.

[0016] Es wurden bereits Anstrengungen unternommen, Massenspektrometrie mit enzymatischen Untersuchungen zu verwenden, um das Vorliegen, die Position und die Identität von Mutationen in ansonsten bekannt-

ten Sequenzen zu bestimmen, wobei zumindest manche Informationen über das Vorliegen, die Position und/oder die Identität der Mutation vorab bekannt sind. Beispielsweise offenbart die U.S.-Patentschrift Nr. 5,605,798 ein Verfahren, bei dem ein DNA-Primer, der zu einem bekannten Targetmolekül in einer zu der bekannten interessierenden Region benachbarten Region komplementär ist, in der Gegenwart von massenmarkierten Dideoxynukleotiden mit einer DNA-Polymerase verlängert wird. Die Identität der Mutation wird anschließend bestimmt, indem die Masse des Dideoxy-verlängerten DNA-Primers analysiert wird. Es wird offenbart, dass das Multiplexierungsverfahren dabei nützlich ist, alle möglichen Mutanten/Varianten an einer definierten Stelle gleichzeitig zu erfassen, indem mit einem Dideoxynukleotid verlängert wird und bestimmt wird, welches spezifische Dideoxynukleotid integriert wurde.

[0017] Es wurden Anstrengungen unternommen, einige der zuvor erwähnten Unzulänglichkeiten bei massenspektroskopischen Analysen von Nukleinsäuren anzugehen. Beispielsweise offenbart Gut (WO 96/27681) Verfahren zum Verändern der Ladungseigenschaften der Phosphodiester-Hauptkette von Nukleinsäuren auf eine Art und Weise, die sie für MS-Analysen geeigneter werden lässt. Verfahren zum Einführen modifizierter Nukleotide, die die Nukleinsäure bezüglich einer Fragmentierung stabilisieren, wurden ebenfalls beschrieben (Schneider und Chait, Nucleic Acids Res, 23, 1.570 (1995), Tang u. a., J Am Soc Mass Spectrom, 8, 218 – 224, 1997).

[0018] Die Verwendung von nicht-spaltbaren Massemarkierungen wurde ebenfalls genutzt, um manche der zuvor erwähnten Unzulänglichkeiten anzugehen. Beispielsweise offenbart die japanische Patentschrift Nr. 59-131909 einen Massenspektrometerentwurf, der Nukleinsäurefragmente, die durch Elektrophorese, Flüssig-chromatographie oder Hochgeschwindigkeits-Gelfiltration getrennt werden, wobei Atome in die Nukleinsäuren integriert wurden, erfasst. Die Atome, die normalerweise nicht in der DNA auftreten, sind Schwefel, Brom, Jod, Silber, Gold, Platin und Quecksilber.

[0019] Spaltbare Massemarkierungen werden dazu genutzt, manche der Probleme, die mit einer MS-Analyse von Nukleinsäuren verbunden sind, zu umschiffen. Beispielsweise offenbart die PCT-Anmeldung WO 95/04160 (Southern u. a.) ein indirektes Verfahren zum Analysieren der Sequenz von Targetnukleinsäuren unter Verwendung einer durch ein Target bewirkten Ligation zwischen einer oberflächengebundenen DNA-Sonde und spaltbaren, massemarkierten Oligonukleotiden, die Reportergruppen enthalten, unter Verwendung von massenspektrometrischen Techniken. Die zu bestimmende Sequenz wird zuerst zu einem Oligonukleotid, das an einen Feststoffträger angelagert ist, hybridisiert. Der Feststoffträger, der die oben genannten Hybride trägt, wird mit einer Lösung aus codierten Oligonukleotidreagenzien, die eine Bibliothek bilden, die alle Sequenzen einer gegebenen Länge umfasst, inkubiert. Es wird eine Ligase eingebracht, so dass das Oligonukleotid auf dem Träger zu dem Angehörigen der Bibliothek ligiert wird, der zu dem Target neben dem Oligonukleotid hybridisiert wird. Nicht-ligierte Reagenzien werden mittels Waschen beseitigt. Ein Linker, der Teil des Angehörigen der Bibliothek ist, der zu dem Oligonukleotid ligiert ist, wird gebrochen, um eine Markierung abzulösen, die wieder gewonnen und mittels Massenspektrometrie analysiert wird.

[0020] Ein gemeinsamer Fokus der obigen Technologien besteht darin, Verfahren zum Erhöhen der Anzahl von Targetstellen (entweder innerhalb von oder zwischen Targets) zu liefern, die bei einer einzelnen Bestimmung abgefragt werden können, wobei ein gewisser Abschnitt der Targetsequenz bekannt ist. Dieses Multiplexierungsthema wird in den Lehren der obigen Patentanmeldungen entweder direkt oder implizit angegeben. Die Verwendung von mehr als einem Oligonukleotid entweder als Hybridisierungssonde oder als Primer zur Erweiterung bzw. Verlängerung bzw. Extension oder Ligation wird durch die Sequenz, die die interessierende Stelle umgibt, und somit durch die spezifische Anwendung definiert. Mit Ausnahme der von Southern offenbarten Massemarkierungstechnologie sind die oben beschriebenen Oligonukleotidreagenzien somit bezüglich der Targetsequenz nicht generisch, sondern müssen für jede definierte Anwendung erzeugt werden. Als solches ist die Anzahl von gesonderten Oligonukleotiden, die bei einer multiplexierten Abfragung verwendet werden, allgemein lediglich eine geringe Teilmenge des theoretischen Sequenz-Vollständig-Satzes. Dieses Verhältnis zwischen der tatsächlichen Sequenzbedeckung, die durch ein bestimmtes Oligonukleotidgemisch geliefert wird, und der theoretischen Bedeckung, die durch den Sequenz-Vollständig-Satz geliefert wird, ist als die Gemischbedeckungskomplexität definiert (siehe nachstehende Erörterung). Beispielsweise variieren bei vielen der beschriebenen Verfahren (d. h. U.S.-Patentschrift Nr. 5,605,798, WO 92/15712 und WO 97/35033) die Sondenlängen je nach der spezifischen Anwendung und dem Erfassungsverfahren zwischen etwa 8 und 20 Nukleotiden. Die Anzahl von Sonden in einem Sequenz-Vollständig-Satz kann durch die Gleichung 4^L beschrieben werden, wobei L gleich der Länge der Sonden ist. Für 8-mer-Sonden weist der Sequenz-Vollständig-Satz somit 4^8 oder 65.536 Angehörige auf. Wenn die Anzahl von Abfragungsstellen bei der multiplexierten Bestimmung etwa 500 beträgt, was für die oben beschriebenen Technologiearten eine vernünftige Obergrenze für die Anzahl von Oligonukleotidsonden bei einer einzigen Bestimmung ist, so wäre die Gemischbedeckungs-

komplexität (siehe nachfolgende Erörterung) des abfragenden 8-mer-Sonde-Gemischs gleich 500/65.536 oder etwa 1/130. In den meisten Fällen weisen die Sonden jedoch eine Länge von 15 – 20 Nukleotiden auf. Obwohl diese erhöhte Länge eine Spezifität der Sonde für eine definierte Targetsequenz gewährleistet, verringert sie die Gemischbedeckungskomplexität des Sondengemischs beträchtlich. Somit ist klar, dass die abfragenden Oligonukleotidgemische für die oben beschriebenen Arten von Multiplexierungsverfahren und -anwendungen nicht dazu entworfen sind, bezüglich der Targetsequenzbedeckung sequenzvollständig zu sein, und somit nicht als generische Reagenzien betrachtet werden könnten.

[0021] Das Ziel vieler arraybasierter Sequenzierungstechniken besteht darin, den „Kurzwort“-Inhalt, d. h. alle vorliegenden Oligonukleotidteilsequenzen, in der anvisierten Nukleinsäuresequenz zu bestimmen. Beispielsweise ist bei Techniken, die eine Hybridisierung zu oberflächengebundenen DNA-Sondenarrays verwenden, ein Satz von Oligonukleotiden einer bestimmten Länge an räumlich gesonderten Positionen auf einem Substrat angeordnet, um ein Array zu bilden, und die Targetsequenz darf zu dem Array hybridisieren (siehe z. B. U.S.-Patentschrift Nr. 5,202,231, U.S.-Patentschrift Nr. 5,492,806 und U.S.-Patentschrift Nr. 5,695,940). Die Targetsequenz wird sich an Positionen binden, die ein Kurzwort enthalten, das komplementär zu einem der Kurzwörter in ihrer Sequenz ist. Andere haben Verfahren zum Sondieren von oberflächengebundenen Targets mit einem sequentiellen Satz von Oligonukleotidsonden offenbart (siehe z. B. U.S.-Patentschrift Nr. 5202231, U.S.-Patentschrift Nr. 5492806 und U.S.-Patentschrift Nr. 5695940). Dadurch, dass mittels einer Fluoreszenzmessung oder dergleichen die Hybridisierungspositionen identifiziert werden oder die Identität des sondierenden Oligonukleotids bekannt ist, kann theoretisch der genaue Kurzwortinhalt der Targetnukleinsäuresequenz bestimmt werden. Diese Informationen können anschließend dazu verwendet werden, die Sequenz der Targetnukleinsäure zu rekonstruieren (siehe z. B. Pevzner, P.A., J. Biomolecular Structure Dynamics 7, 63 (1989), Pevzner P.A., u.a., J. Biomolecular Structure Dynamics 9, 399 (1991), Ukkonen, E., Theoretical Computer Science 92, 191 (1992)). Jedoch ist es wichtig zu betonen, dass relativ sequenzvollständige Sätze von Oligonukleotidsonden benötigt werden, um den Kurzwortinhalt eines unbekannten Targets generisch zu bestimmen.

[0022] Techniken, die den Kurzwortinhalt der Targetnukleinsäuresequenz identifizieren, eignen sich für Anwendungen wie z. B. De-Novo-Sequenzieren, Neu-Sequenzieren, Mutationserfassung und Mutationsänderungserfassung. Mit zunehmender Länge der Targetsequenz verringert sich die Erfolgsrate oder die Erfolgsrate, mit der die Analyse durchgeführt werden kann. Da manche der Anwendungen, z. B. Mutationserfassung, lediglich qualitative Informationen benötigen, kann die Erfolgsrate in der Regel höher liegen als die Erfolgsrate für eine Anwendung, die quantitative Informationen erfordert, z. B. De-Novo-Sequenzieren. Beispielsweise würde das Vorliegen einiger weniger Kurzwortwiederholungen die Erfolgsrate für ein De-Novo-Sequenzieren beträchtlich verringern, hätte jedoch eine geringere Auswirkung auf die Erfolgsrate bezüglich einer Mutationserfassung. Bei anderen Anwendungen ist ein beträchtliches Maß an Vorab-Informationen verfügbar, um die Interpretation des Kurzwortinhalts zu unterstützen, wodurch die Erfolgsrate der Ergebnisse erhöht wird.

[0023] Der Zweck der vorliegenden Erfindung besteht darin, den Kurzwortinhalt einer Targetnukleinsäuresequenz unter Verwendung von Massenspektroskopie zu bestimmen. Jedoch erwartet man, dass die Erfolgsrate einer derartigen Analyse relativ gering ist, da das Vorliegen einer bestimmten Masse in dem Massenspektrum lediglich zeigt, dass eine von vielen möglichen Nukleinsäuresequenzen vorliegt. Beispielsweise ist unter Verwendung von lediglich natürlichen Nukleotiden die Sequenz GGCTTTA anhand der Masse von der Sequenz GCTTTAG unterscheidbar, und das Vorliegen einer Massenspitze bei 2.142 Atommasseeinheiten zeigt lediglich, dass in dem Gemisch zumindest eine Nukleinsäuresequenz mit 3Ts, 2Gs und 1A und 1C vorliegt. Durch Massekoordinaten wird die Mehrdeutigkeit noch verwirrender. Beispielsweise kann die Massenspitze bei 2.193 Beiträge von Nukleinsäuresequenzen enthalten, die 6As und 1T oder 1A, 2Cs, 3Gs und 1T enthalten. Der Zweck der vorliegenden Erfindung besteht darin, diese Arten von Mehrdeutigkeiten in dem Kurzwortinhalt einer Targetnukleinsäuresequenz zu verringern.

Zusammenfassung der Erfindung

[0024] Die Erfindung ist auf Reagenzien und Verfahren zum Rekapitulieren einer Targetnukleinsäure in der Kurzwortform gerichtet, die mittels Hochauflösungs-Massenspektrometietechniken analysiert werden kann. Die Verfahren und Reagenzien verwenden generische Oligonukleotid-Vorläufer-Gemische (X-mer-Vorläufer-Gemische) und enzymatische Prozesse, um die Länge und gleichzeitig die Masse lediglich derjenigen X-mer-Vorläufer innerhalb eines definierten Gemischs zu verändern, die zu der Targetnukleinsäure komplementär sind.

[0025] Ein Aspekt der vorliegenden Erfindung ist ein Gemisch zur direkten Massenspektralanalyse von Nukleinsäuren. Das Gemisch umfasst natürliche und massenmodifizierte X-mer-Vorläufer, die eine minimale Län-

ge von 3 Nukleotiden aufweisen. Die minimale Gemischbedeckungskomplexität (CC_M) beträgt 56 geteilt durch die Anzahl von gesonderten X-meren in dem Gemisch. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Die Massenzahlkomplexität (MNC) des Gemischs ist größer als die Massenzahlkomplexität jedes natürlichen Äquivalents des Gemisches. Jeder der X-mer-Vorläufer in dem Gemisch wird durch eine einzelne chemische Spezies dargestellt, um eine eindeutige Masse zu besitzen.

[0026] Ein weiterer Aspekt der vorliegenden Erfindung ist ein Verfahren zum Analysieren einer Targetnukleinsäuresequenz. Ein Gemisch von X-mer-Vorläufern wird zu den Targetnukleinsäuresequenzen hybridisiert. Das Gemisch umfasst natürliche und massenmodifizierte X-mer-Vorläufer, die eine minimale Länge von 3 Nukleotiden und eine minimale Gemischbedeckungskomplexität von 56 geteilt durch die Anzahl von gesonderten X-meren in dem Gemisch aufweisen. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Jeder der X-mer-Vorläufer in dem Gemisch wird durch eine einzelne chemische Spezies dargestellt, um eine eindeutige Masse zu besitzen. Die Hybride werden verarbeitet, um die Masse der X-mer-Vorläufer-Abschnitte der Hybride in einer mittels einer Targetsequenz bewirkten Reaktion zu verändern. Die Produkte des vorherigen Schrittes werden mittels Massenspektrometrie analysiert.

[0027] Derartige X-mer-Vorläufer können eine definierte isotope Zusammensetzung aufweisen.

[0028] Die Massenzahlkomplexität (MNC) des Gemischs ist vorzugsweise größer als die Massenzahlkomplexität jedes natürlichen Äquivalents des Gemisches.

[0029] Die X-mer-Vorläufer können eine Ionisierungsmarke aufweisen. Vorzugsweise weist das Gemisch oder der Satz von Teilgemischen eine Gemischbedeckungskomplexität von zumindest etwa 1/2 auf, wenn das Gemisch zumindest 128 gesonderte X-mere enthält, eine Gemischbedeckungskomplexität von zumindest etwa 1/4, wenn das Gemisch zumindest 256 gesonderte X-mere enthält, und eine Gemischbedeckungskomplexität von zumindest etwa 1/8, wenn das Gemisch zumindest 512 gesonderte X-mere enthält.

[0030] Das Gemisch kann in zumindest zwei Reaktionsgemischen vorgesehen sein.

[0031] Zumindest manche der massenmodifizierten X-mer-Vorläufer können zumindest eine Massenmarke oder zumindest eine chemische Modifizierung einer Internukleosidbindung, einer Zuckerhauptkette oder einer Nukleosidbase aufweisen.

[0032] Der Verarbeitungsschritt umfasst vorzugsweise ein Verändern der Masse der X-mer-Vorläufer-Abschnitte der Hybride um einen Betrag, der größer ist als derjenige, der durch die Massendifferenz zwischen dem leichtesten und dem schwersten X-mer in dem Gemisch definiert wird.

[0033] Das Verfahren kann ferner den Schritt des Reinigens der Produkte aus Schritt (2) vor einer Analyse mittels Massenspektrometrie umfassen.

[0034] Das Verfahren kann ferner den Schritt des Trennens der Produkte aus Schritt (2) vor einer Analyse mittels Massenspektrometrie umfassen.

[0035] Schritte (1) – (2) können in Lösung durchgeführt werden, oder sie können mit einem oberflächengebundenen Gemisch durchgeführt werden.

[0036] Die Produkte können mittels MALDI-TOF-Massenspektrometrie analysiert werden.

[0037] Der Verarbeitungsschritt kann eine mittels einer Targetsequenz bewirkte enzymatische Untersuchung umfassen, die eine Polymeraseverlängerungsuntersuchung oder eine Ligaseuntersuchung sein kann.

[0038] Der Verarbeitungsschritt umfasst vorzugsweise ein Verlängern der hybridisierten X-mer-Vorläufer durch ein Polymerisieren zumindest eines, vorzugsweise eines einzelnen, Nukleotids an dem 3'-Ende der hybridisierten X-mer-Vorläufer. Das Nukleotid ist vorzugsweise ein Kettenabbruch-Nukleotid-Triphosphat, das eine Ionisierungsmarke aufweisen kann.

[0039] Das Kettenabbruch-Nukleotid-Triphosphat kann ein natürliches Dideoxynukleotid-Triphosphat oder ein massenmodifiziertes Dideoxynukleotid-Triphosphat sein.

[0040] Die Masse des massenmodifizierten Dideoxynukleotid-Triphosphats ist vorzugsweise größer als die-

jenige, die durch die Massendifferenz zwischen dem leichtesten und dem schwersten X-mer in dem Gemisch definiert wird.

[0041] Hybridisierte X-mer-Vorläufer können unter Verwendung eines Enzyms, das eine Nukleotidpolymeraseaktivität aufweist, verlängert werden.

[0042] Das Verfahren kann ferner den Schritt des Aufschließens der Produkte aus Schritt (2) vor einer Analyse mittels Massenspektrometrie umfassen.

[0043] Bei einem alternativen Ausführungsbeispiel kann das Nukleotid ein Verlängerungsnukleotid-Triphosphat, ein natürliches Dideoxynukleotid-Triphosphat oder ein massenmodifiziertes Dideoxynukleotid-Triphosphat sein.

[0044] Das Verlängerungsnukleotid-Triphosphat ist vorzugsweise ein Desoxynukleotid, ein 5'-(α)-Phosphothioat-Analog, ein 5'-N- α -Phosphoramidat-Analog oder ein Ribonukleotid.

[0045] Der Aufschlusschritt kann mit einer Nuklease, z. B. 5'-Exonuklease, durchgeführt werden. Die 5'-Exonuklease ist vorzugsweise eine DNA-Polymerase oder ein T7-Gen 6.

[0046] Der Aufschlusschritt kann mittels einer chemischen Reaktion durchgeführt werden.

[0047] Bei einem bevorzugten Ausführungsbeispiel umfasst der Verarbeitungsschritt der vorliegenden Erfindung ein Legieren benachbarter X-mer-Vorläufer unter Verwendung einer DNA-Ligase oder eines Kondensationsmittels. Das Kondensationsmittel ist vorzugsweise ein Carbodiimid oder ein Cyanbromidderivat.

[0048] Der Verarbeitungsschritt kann eine chemische Untersuchung umfassen.

[0049] Ein weiterer Aspekt der vorliegenden Erfindung ist auf ein Verfahren zum Analysieren einer Targetnukleinsäuresequenz gerichtet, die ein 3'-Ende und ein 5'-Ende aufweist. Die Targetnukleinsäuresequenz wird zu einer Vielzahl von Nukleinsäuresonden in einem Array hybridisiert, das eine Oberfläche und eine Vielzahl von Nukleinsäuresequenzsonden aufweist. Jede der Sonden umfasst einen spaltbaren Linker, der an die Oberfläche angelagert ist, und eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist. Ein Gemisch von X-mer-Vorläufern wird zu der Targetnukleinsäuresequenz hybridisiert. Das Gemisch weist natürliche und massenmodifizierte X-mer-Vorläufer auf, die eine minimale Länge von 3 Nukleotiden und eine minimale Gemischbedeckungskomplexität von 56 geteilt durch die Anzahl von gesonderten X-meren in dem Gemisch aufweisen. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Jeder der X-mer-Vorläufer in dem Gemisch wird durch eine einzelne chemische Spezies dargestellt. Die hybridisierten X-mer-Vorläufer, die benachbart zu dem endständigen 5'-Phosphat angeordnet sind, werden mit der oberflächengebundenen Sonde ligiert, um mit der daran angelagerten Targetnukleinsäuresequenz einen Komplex aus hybridisiertem Vorläufer/Sonde zu bilden. Der Komplex wird an dem spaltbaren Linker gespalten und mittels Massenspektrometrie analysiert.

[0050] Die Massenzahlkomplexität (MNC) des Gemisches ist vorzugsweise größer als die Massenzahlkomplexität jedes natürlichen Äquivalents des Gemisches.

[0051] Vorzugsweise weist das Gemisch oder der Satz von Teilgemischen eine Gemischbedeckungskomplexität von zumindest etwa 1/2 auf, wenn das Gemisch zumindest 128 gesonderte X-mera enthält, eine Gemischbedeckungskomplexität von zumindest etwa 1/4, wenn das Gemisch zumindest 256 gesonderte X-mera enthält, und eine Gemischbedeckungskomplexität von zumindest etwa 1/8, wenn das Gemisch zumindest 512 gesonderte X-mera enthält.

[0052] Die Zusammensetzung des Gemisches kann bekannt sein.

[0053] Das Gemisch kann in zumindest zwei Reaktionsgemischen vorgesehen sein.

[0054] Zumindest manche der massenmodifizierten X-mer-Vorläufer können zumindest eine Massenmarke oder zumindest eine chemische Modifizierung einer Internukleosidbindung, einer Zuckerhauptkette oder einer Nukleosidbase aufweisen.

[0055] Der hybridisierte X-mer-Vorläufer kann unter Verwendung einer DNA-Ligase oder eines Kondensationsmittels mit der Sonde ligiert werden. Das Kondensationsmittel ist vorzugsweise aus der Gruppe ausgewählt, die aus Carbodiimiden und Cyanbromidderivaten besteht.

[0056] Der spaltbare Linker ist vorzugsweise ein photospaltbarer Linker oder ein chemischer spaltbarer Linker.

[0057] Die Komplexe werden vorzugsweise mittels MALDI-TOF-Massenspektrometrie analysiert.

[0058] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz gemäß Anspruch 9 zum Durchführen des obigen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, ein Enzym, das eine DNA-Polymeraseaktivität aufweist, und eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht.

[0059] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz gemäß Anspruch 9 zum Durchführen des obigen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, ein Enzym, das eine DNA-Polymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen und massenmodifizierten Kettenabbruch-Triphosphaten und einer Vielzahl von Verlängerungsnukleotid-Triphosphaten besteht.

[0060] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz gemäß Anspruch 9 zum Durchführen des obigen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, ein Enzym, das eine DNA-Polymeraseaktivität aufweist, und eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht.

[0061] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz gemäß Anspruch 9 zum Durchführen des obigen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, ein Enzym, das eine DNA-Polymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen und massenmodifizierten Kettenabbruch-Triphosphaten, einer Vielzahl von Verlängerungsnukleotid-Triphosphaten und einer Nuklease besteht.

[0062] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz gemäß Anspruch 9 zum Durchführen eines oben beschriebenen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, eine DNA-Polymerase, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen und massenmodifizierten Kettenabbruch-Triphosphaten, einer 5'-Exonuklease und einer Vielzahl von Verlängerungsnukleotid-Triphosphaten und 5'-(α)-Phosphothioat-Analoga besteht.

[0063] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz zum Durchführen eines Verfahrens, wie es oben beschrieben wurde. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, und eine DNA-Ligase.

[0064] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz zum Durchführen eines Verfahrens, wie es oben beschrieben wurde. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, und ein Kondensationsmittel.

[0065] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz zum Durchführen eines oben beschriebenen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, eine DNA-Ligase und ein Array, das eine Oberfläche und eine Vielzahl von Nukleinsäuresequenzsonden aufweist, die einen spaltbaren Linker, der an der Oberfläche angelagert ist, und eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist, umfassen.

[0066] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz zum Durchführen eines oben beschriebenen Verfahrens. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, ein Kondensationsmittel, ein Array, das eine Oberfläche und eine Vielzahl von Nukleinsäuresequenzsonden aufweist, die einen spaltbaren Linker, der an der Oberfläche angelagert ist, und eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist, umfassen.

Kurze Beschreibung der Zeichnungen

[0067] [Fig. 1](#) ist eine Rekapitulation einer Targetsequenz mittels unterschiedlicher Arten von Sätzen von kurzen X-meren.

[0068] [Fig. 2](#) ist ein Diagramm, das die Schritte der Polymeraseverlängerungsuntersuchung (PEA – polymerase extension assay) umreißt.

[0069] [Fig. 3](#) ist ein Diagramm, das die Schritte der Polymeraseverlängerungs- und -spaltungsuntersuchung (PEA-CA – polymerase extension and cleavage assay) umreißt.

[0070] [Fig. 4](#) ist ein Diagramm, das die Schritte der Polymeraseverlängerungs- und -spaltungsuntersuchung II umreißt.

[0071] [Fig. 5](#) ist ein Diagramm, das die Schritte der X-mer-Ligation-Untersuchung (XLA – X-mer ligation assay) umreißt.

[0072] [Fig. 6](#) ist ein Diagramm, das die Schritte der arraybasierten X-mer-Ligation-Untersuchung (AXLA – array-based X-mer ligation assay) umreißt.

[0073] [Fig. 7](#) ist ein Histogramm aller 16.384 7-mer-PEA-Produkte, die ausgehend von allen natürlichen und einzeln optimierten massenmodifizierten 6-mer-Vorläufern erzeugt werden.

[0074] [Fig. 8](#) ein Flussdiagramm zur Optimierung von massenmodifizierten X-mer-Vorläufern.

[0075] [Fig. 9](#) ist ein Histogramm für alle 16.384 7-mer-PEA-Produkte, die ausgehend von massemarkierten ddNTPs erzeugt werden.

[0076] [Fig. 10](#) zeigt die Auswirkung einer prozentualen Sequenzbedeckung eines 6-mer-Vorläufer-Gemischs auf die Auflösungsleistung von PEA bei definierten Erfolgsraten.

[0077] [Fig. 11](#) ist eine Sequenz einer Region des menschlichen p53-Gens, das bekannte Mutationsstellen aufweist.

[0078] [Fig. 12](#) zeigt einen verschachtelten Satz von überlappenden 7-mer-PEA-Produkten, der dem 62-Nukleotid-Fragment der Wildtyp-p53-Sequenz entspricht.

[0079] [Fig. 13](#) zeigt Massenspektren für eine PEA-Analyse des Wildtyp- und des G2451C-p53-Mutanten in dem 62-Nukleotid-Fragment unter Verwendung natürlicher X-mer-Vorläufer.

[0080] [Fig. 14](#) zeigt binäre transformierte und Binär-Differenz-Spektren für eine PEA-Analyse des Wildtyp- und des G2451C-p53-Mutanten unter Verwendung natürlicher 6-mer-Vorläufer.

[0081] [Fig. 15](#) zeigt Massenspektren für eine PEA-Analyse von sechs p53-Mutationen unter Verwendung natürlicher 6-mer-Vorläufer.

[0082] [Fig. 16](#) zeigt Binär-Differenz-Spektra für eine PEA-Analyse von sechs p53-Mutationen unter Verwendung natürlicher 6-mer-Vorläufer.

[0083] [Fig. 17](#) zeigt Massenspektren für eine PEA-Analyse des Wildtyp- und des G2451C-p53-Mutanten in dem 378-Nukleotid-Fragment unter Verwendung natürlicher 6-mer-Vorläufer.

[0084] [Fig. 18](#) zeigt Massenspektren für eine PEA-Analyse des Wildtyp- und des G2451C-p53-Mutanten in dem 378-Nukleotid-Fragment unter Verwendung der optimierten massenmodifizierten 6-mer-Vorläufer.

[0085] [Fig. 19](#) zeigt einen Satz von halb-überlappenden 7-mer-PEACA-Produkten, die dem 62-Nukleotid-Fragment der Wildtyp-p53-Sequenz entsprechen.

[0086] [Fig. 20](#) zeigt Massenspektren für eine PEACA-Analyse des Wildtyp- und des G2481A-p53-Mutanten in dem 62-Nukleotid-Fragment unter Verwendung natürlicher 6-mer-Vorläufer.

[0087] [Fig. 21](#) zeigt Massenspektren für eine PEACA-Analyse von drei p53-Mutationen in dem 378-Nukleotid-Fragment.

[0088] [Fig. 22](#) zeigt Massenspektren für eine XLA-Analyse des Wildtyp- und des G2481A-p53-Mutanten in dem 62-Nukleotid-Fragment unter Verwendung natürlicher 6-mer-Vorläufer.

[0089] [Fig. 23](#) zeigt Massenspektren für eine XLA-Analyse des Wildtyp- und des G2481A-p53-Mutanten in dem 378-Nukleotid-Fragment unter Verwendung natürlicher 6-mer-Vorläufer.

Ausführliche Beschreibung der Erfindung

Definitionen

[0090] In dieser Spezifikation und in den folgenden Patentansprüchen wird auf eine Anzahl von Begriffen Bezug genommen, die definitionsgemäß die folgende Bedeutung aufweisen sollen:

Der Begriff „Polynukleotid“ oder „Nukleinsäure“ bezieht sich auf eine Verbindung oder Zusammensetzung, die ein polymeres Nukleotid oder ein Nukleinsäurepolymer ist. Das Polynukleotid kann eine natürliche Verbindung oder eine synthetische Verbindung sein. Das Polynukleotid kann von etwa 20 bis 5.000.000 oder mehr Nukleotide aufweisen. Die größeren Polynukleotide liegen allgemein im natürlichen Zustand vor. In einem isolierten Zustand kann das Polynukleotid etwa 30 bis 50.000 oder mehr Nukleotide, üblicherweise etwa 100 bis 20.000 Nukleotide, häufiger 500 bis 10.000 Nukleotide, aufweisen. Somit ist es offensichtlich, dass eine Isolierung eines Polynukleotids aus dem natürlichen Zustand oft zu einer Fragmentierung führt. Es kann nützlich sein, längere Targetnukleinsäuresequenzen, insbesondere RNR, vor einer Hybridisierung zu fragmentieren, um konkurrierende intramolekulare Strukturen zu verringern.

[0091] Die Polynukleotide umfassen Nukleinsäuren, und Fragmente derselben, von jeglicher Quelle in gereinigter oder ungereinigter Form, einschließlich DNA (dsDNA und ssDNA) und RNA, einschließlich tRNA, mRNA, rRNA, Mitochondrien-DNA und -RNA, Chloroplast-DNA und -RNA, DNA/RNA-Hybride oder Gemische derselben, Gene, Chromosomen, Plasmiden, Cosmiden, der Genome von biologischem Material wie z. B. Mikroorganismen, z. B. Bakterien, Hefen, Phagen, Chromosomen, Viren, Viroiden, Schimmelpilzen, Pilzen, Pflanzen, Tieren, Menschen und dergleichen. Das Polynukleotid kann auch lediglich eine untergeordnete Fraktion eines komplexen Gemisches wie z. B. einer biologischen Probe sein. Ebenfalls enthalten sind Gene, z. B. Hämoglobin-Gen für Sichelzellenanämie, Zystische-Fibrose-Gen, Onkogene, cDNA und dergleichen.

[0092] Das Polynukleotid kann mittels Prozeduren, die in der Technik hinreichend bekannt sind, aus verschiedenen biologischen Materialien gewonnen werden. Wo es angebracht ist, kann das Polynukleotid gespalten werden, um ein Fragment zu erhalten, das eine Targetnukleotidsequenz enthält, z. B. durch Scheren oder durch eine Behandlung mit einer Restriktionsendonuklease oder anhand eines anderen Verfahrens einer stellspezifischen chemischen Spaltung.

[0093] Für die Zwecke der vorliegenden Erfindung ist das Polynukleotid, oder ein aus dem Polynukleotid gewonnenes gespaltenes Fragment, üblicherweise zumindest teilweise denaturiert oder einsträngig, oder es wird behandelt, damit es denaturiert oder einsträngig wird. Derartige Behandlungen sind in der Technik hinreichend bekannt und umfassen z. B. eine Wärme- oder Alkalibehandlung oder einen enzymatischen Aufschluss eines Stranges. Beispielsweise kann dsDNA über einen Zeitraum von etwa 1 bis 10 Minuten bei 90 bis 100°C erhitzt werden, um ein denaturiertes Material zu erzeugen.

[0094] Die Nukleinsäuren können mittels in-vitro-Replikation und/oder mittels Amplifikationsverfahren wie z. B. Polymerasekettenreaktion (PCR – polymerase chain reaction), der asymmetrischen PCR, der Ligasekettenreaktion (LCR – ligase chain reaction) und so weiter, erzeugt werden. Die Nukleinsäuren können entweder einsträngig oder doppelsträngig sein. Einsträngige Nukleinsäuren sind bevorzugt, da sie keine komplementären Stränge aufweisen, die während des Hybridisierungsschrittes des Verfahrens der Erfindung um die Oligonukleotid-Vorläufer konkurrieren.

[0095] Der Ausdruck „Targetnukleinsäuresequenz“ bezieht sich auf eine Sequenz von Nukleotiden, die zu identifizieren, zu erfassen oder auf andere Weise zu analysieren sind und die üblicherweise in einem Abschnitt eines Polynukleotids oder in einem ganzen Polynukleotid vorliegen. Bei der vorliegenden Erfindung kann die Identität der Targetnukleotidsequenz bekannt sein, muss aber nicht. Die Identität der Targetnukleotidsequenz kann bis zu einem Grad bekannt sein, der ausreichend ist, um eine Herstellung verschiedener Sequenzen, die mit der Targetnukleotidsequenz hybridisierbar sind, und von Oligonukleotiden, z. B. Sonden und Primern, und

anderen Molekülen zu ermöglichen, die zum Durchführen von Verfahren gemäß der vorliegenden Erfindung und so weiter notwendig sind.

[0096] Üblicherweise enthält die Targetsequenz von etwa 30 bis 5.000 oder mehr Nukleotide, vorzugsweise 50 bis 1.000 Nukleotide. Die Targetnukleotidsequenz ist allgemein ein Anteil eines größeren Moleküls, oder sie kann im Wesentlichen das ganze Molekül sein, z. B. ein Polynukleotid, wie oben beschrieben wurde. Die Mindestanzahl von Nukleotiden in der Targetnukleotidsequenz ist so gewählt, um zu gewährleisten, dass das Vorliegen eines Targetpolynukleotids in einer Probe ein spezifischer Indikator des Vorliegens von Polynukleotid in einer Probe ist. Die maximale Anzahl von Nukleotiden in der Targetnukleotidsequenz wird normalerweise von mehreren Faktoren geregelt: der Länge des Polynukleotids, von dem sie abgeleitet ist, der Tendenz eines derartigen Polynukleotids, während der Isolierung mittels Scheren oder anderer Prozesse gebrochen zu werden, der Effizienz jeglicher Prozeduren, die erforderlich sind, um die Probe für eine Analyse zu präparieren (z. B. Transkription einer DNA-Matrize zu RNA), und der Effizienz einer Identifikation, Erfassung, Amplifikation und/oder einer anderen Analyse der Targetnukleotidsequenz, wo dies angebracht ist.

[0097] Der Begriff „Oligonukleotid“ bezieht sich auf ein Polynukleotid, üblicherweise einsträngig, üblicherweise ein synthetisches Polynukleotid, das jedoch auch ein natürlich vorkommendes Polynukleotid sein kann. Die Länge eines Oligonukleotids wird allgemein durch die jeweilige Rolle desselben, z. B. Sonde, Primer, X-mer und dergleichen, geregelt. Zum Herstellen eines Oligonukleotids können verschiedene Techniken eingesetzt werden. Derartige Oligonukleotide können mittels biologischer Synthese oder mittels chemischer Synthese erhalten werden. Für kurze Oligonukleotide (bis zu etwa 100 Nukleotiden) ist die chemische Synthese im Vergleich zur biologischen Synthese häufig wirtschaftlicher. Zusätzlich zur Wirtschaftlichkeit liefert die chemische Synthese eine zweckmäßige Art und Weise, Verbindungen von niedriger relativer Molekülmasse und/oder modifizierte Basen während spezifischer Syntheseschritte zu integrieren. Ferner ist die chemische Analyse sehr flexibel bei der Wahl der Länge und der Region der Targetpolynukleotidbindungssequenz. Das Oligonukleotid kann mittels standardmäßiger Verfahren synthetisiert werden, z. B. mittels derjenigen, die bei kommerziellen automatisierten Nukleinsäuresynthetisierern verwendet werden. Die chemische Synthese einer DNA auf einem geeignet modifizierten Glas oder Harz kann zu einer kovalent an die Oberfläche gebundenen DNA führen. Dies kann Vorteile bezüglich des Waschens und des Handhabens der Probe bieten. Verfahren der Oligonukleotid-Synthese umfassen Phosphotriester- und Phosphodiester-Verfahren (Narang u. a. (1979), Meth. Enzymol 68: 90) und eine Synthese auf einem Träger (Beaucage u. a. (1981), Tetrahedron Letters 22: 1.859 – 1.862) sowie Phosphoramidit-Techniken (Caruthers, M. H., u. a., „Methods in Enzymology“, Bd. 154, S. 287 – 314 (1988)) und andere, die bei „Synthesis and Applications of DNA and RNA“, S.A. Narang, Herausgeber, Academic Press, New York, 1987, und den darin enthaltenen Bezugnahmen auf andere Dokumente beschrieben sind. Die chemische Synthese mittels eines photolithographischen Verfahrens von räumlich adressierbaren Arrays von Oligonukleotiden, die an Glasoberflächen gebunden sind, wird von A. C. Pease u. a., Proc. Nat. Acad. Sci. USA (1994) 91: 5.022 – 5.026, beschrieben.

[0098] Der Begriff „X-mer“ bezieht sich auf ein Oligonukleotid, das eine definierte Länge aufweist, die üblicherweise eine Sequenz einer Länge von zumindest 3 Nukleotiden, vorzugsweise 4 bis 14 Nukleotiden und üblicherweise 5 bis 7 Nukleotiden ist.

[0099] Der Ausdruck „X-mer-Vorläufer“, die manchmal als „Oligonukleotid-Vorläufer“ bezeichnet werden, bezieht sich auf eine Nukleinsäuresequenz, die zu einem Abschnitt der Targetnukleinsäuresequenz komplementär ist. Die Oligonukleotid-Vorläufer sind Sequenzen von Nukleosidmonomeren, die mittels Phosphorbindungen (z. B. Phosphodiester, Alkyl- und Arylphosphat, Phosphorthioat, Phosphotriester) oder Nicht-Phosphor-Bindungen (z. B. Peptid, Sulfamat und andere) verbunden werden. Sie können natürliche oder synthetische Moleküle einer einsträngigen DNA und einer einsträngigen RNA mit kreisförmigen, verzweigten oder linearen Formen sein und optional Bereiche bzw. Domänen umfassen, die in der Lage sind, stabile Sekundärstrukturen (z. B. Stamm-und-Schleife- und Schleife-Stamm-Schleife-Strukturen) zu bilden. Die Oligonukleotid-Vorläufer enthalten ein 3'- und ein 5'-Ende. Der Begriff wird durch ein ω bezeichnet.

[0100] Der Begriff „Gemisch“ bezieht sich auf ein physisches Gemisch von zwei oder mehr Substanzen. Der Ausdruck wird durch Ω bezeichnet.

[0101] Der Ausdruck „Oligonukleotidsonde“ bezieht sich auf ein Oligonukleotid, das verwendet wird, um sich an einen Abschnitt eines Polynukleotids wie z. B. ein weiteres Oligonukleotid oder eine Targetnukleotidsequenz zu binden. Der Entwurf und die Herstellung der Oligonukleotidsonden hängen allgemein von der Sequenz ab, an die sie sich binden.

[0102] Der Ausdruck „Oligonukleotid-Primer“ bezieht sich auf ein Oligonukleotid, das üblicherweise bei einer Kettenverlängerung auf einer Polynukleotidmatrix, beispielsweise bei einer Amplifikation einer Nukleinsäure, eingesetzt wird. Der Oligonukleotid-Primer ist üblicherweise ein synthetisches Nukleotid, das einsträngig ist, eine Sequenz an seinem 3'-Ende enthält, die in der Lage ist, mit einer definierten Sequenz des Targetpolynukleotids zu hybridisieren. Normalerweise weist ein Oligonukleotid-Primer eine Komplementarität von zumindest 80 %, vorzugsweise 90 %, stärker bevorzugt 95 %, am stärksten bevorzugt 100 %, zu einer definierten Sequenz oder einer Primer-Bindungsstelle auf. Die Anzahl von Nukleotiden in der hybridisierbaren Sequenz eines Oligonukleotid-Primers sollte derart sein, dass strenge Bedingungen, die zum Hybridisieren des Oligonukleotid-Primers verwendet werden, eine übermäßige willkürliche nicht-spezifische Hybridisierung verhindern. Üblicherweise ist die Anzahl von Nukleotiden in dem Oligonukleotid-Primer mindestens so groß wie die definierte Sequenz des Targetpolynukleotids, nämlich zumindest zehn Nukleotide, vorzugsweise zumindest 15 Nukleotide und allgemein von etwa 10 bis 200, vorzugsweise 20 bis 50, Nukleotide.

[0103] Der Ausdruck „Nukleosid-Triphosphate“ bezieht sich auf Nukleoside, die einen 5'-Triphosphat-Substituenten aufweisen. Die Nukleoside sind Pentosezuckerderivate von Stickstoffbasen entweder einer Purin- oder einer Pyrimidinableitung, kovalent an den 1'-Kohlenstoff des Pentosezuckers gebunden, der üblicherweise eine Desoxyribose oder eine Ribose ist. Die Purinbasen umfassen Adenin (A), Guanin (G), Inosin (I) und Derivate und Analoga derselben. Die Pyrimidinbasen umfassen Cytosin (C), Thymin (T), Uracil (U) und Derivate und Analoga derselben. Nukleosid-Triphosphate umfassen Desoxyribonukleosid-Triphosphate wie z. B. die vier üblichen Desoxyribonukleosid-Triphosphate dATP, dCTP, dGTP und dTTP und Ribonukleosid-Triphosphate wie z. B. die vier üblichen Triphosphate rATP, rCTP, rGTP und rUTP. Der Begriff „Nukleosid-Triphosphate“ umfasst ferner Derivate und Analoga derselben, die beispielhaft durch diejenigen Derivate veranschaulicht werden, die auf ähnliche Weise wie die nicht-abgeleiteten Nukleosid-Triphosphate erkannt und polymerisiert werden.

[0104] Der Begriff „Nukleotid“ oder „Nukleotidbase“ oder „Base“ bezieht sich auf eine Base-Zucker-Phosphat-Kombination, die die monomere Einheit von Nukleinsäurepolymeren, d. h. DNA und RNA, ist. Gemäß seiner Verwendung in dem vorliegenden Dokument umfasst der Begriff modifizierte Nukleotide, wie sie nachfolgend definiert werden. Allgemein bezieht sich der Begriff auf jegliche Verbindung, die einen zyklischen Zucker vom Furanosidtyp (β -D-Ribose in RNA und β -D-2'-Desoxyribose in DNA) enthält, der an der 5'-Position phosphoryliert ist und entweder eine Base vom Purintyp oder vom Pyrimidintyp aufweist, die über eine β -Glycosid-C1'-N-Bindung an die C-1'-Zucker-Position angelagert ist. Diese Begriffe sind austauschbar und werden durch ein b bezeichnet. Das Nukleotid kann natürlich oder synthetisch sein, einschließlich eines Nukleotids, das massenmodifiziert wurde, einschließlich, unter anderem, Nukleotiden, die modifizierte Nukleoside mit modifizierten Basen (z. B. 5-Methylcytosin) und modifizierten Zuckergruppen (z. B. 2'-O-Methylribosyl, 2'-O-Methoxyethylribosyl, 2'-Fluorribosyl, 2'-Aminoribosyl und dergleichen) aufweisen.

[0105] Der Begriff „DNA“ bezieht sich auf Desoxyribonukleinsäure.

[0106] Der Begriff „RNA“ bezieht sich auf Ribonukleinsäure.

[0107] Der Begriff „natürliches Nukleotid“ bezieht sich auf diejenigen Nukleotide, die die fundamentalen Baublöcke einer zellulären DNA bilden, die definitionsgemäß Desoxycytidylsäure (pdC), Desoxyadenylsäure (pdA), Desoxyguanylsäure (pdG) und Desoxythymidylsäure (pdT) umfassen, und die die fundamentalen Baublöcke einer zellulären RNA bilden, die definitionsgemäß Desoxycytidylsäure (pdC), Desoxyadenylsäure (pdA), Desoxyguanylsäure (pdG) und Desoxyuridylsäure (pdU) umfassen. pdU wird als natürliches Äquivalent von pdT betrachtet.

[0108] Der Begriff „natürliche Nukleotidbase“ bezieht sich auf Basen vom Purin- und Pyrimidin-Typ, die in zellulärer DNA zu finden sind und Cytosin (C), Adenin (A), Guanin (G) und Thymin (T) umfassen, und die in zellulärer RNA zu finden sind und Cytosin (C), Adenin (A), Guanin (G) und Uracil (U) umfassen. U wird als natürliches Äquivalent von T betrachtet.

[0109] Der Ausdruck „modifiziertes Nukleotid“ bezieht sich auf eine Einheit in einem Nukleinsäurepolymer, die eine modifizierte Base-, Zucker- oder Phosphatgruppe enthält. Das modifizierte Nukleotid kann mittels einer chemischen Modifikation des Nukleotids entweder als Teil des Nukleinsäurepolymers oder vor der Integration des modifizierten Nukleotids in das Nukleinsäurepolymer hergestellt werden. Beispielsweise können die oben erwähnten Verfahren zur Synthese eines Oligonukleotids eingesetzt werden. Bei einem anderen Lösungsansatz kann ein modifiziertes Nukleotid hergestellt werden, indem ein modifiziertes Nukleosid-Triphosphat während einer Amplifikationsreaktion in die Polymerkette integriert wird. Beispiele modifizierter Nukleotide umfas-

sen, im Sinne der Veranschaulichung und nicht der Einschränkung, Dideoxynukleotide, Derivate oder Analoga, die biotinyliert, aminmodifiziert, alkyliert, mit Fluorophor markiert und dergleichen werden und ferner Phosphorthioat, Phosphit, ringatommodifizierte Derivate usw. umfassen.

[0110] Der Ausdruck „Basenpaarung gemäß Watson-Crick“ bezieht sich auf die Wasserstoffbrückenbindung zwischen zwei Basen, wobei spezifische Muster von Wasserstoffbrückenbindungsdonatoren und -akzeptoren die standardmäßigen Geometrien aufweisen, die bei „Principles of Nucleic Acid Structure“, Wolfram Saenger, Springer-Verlag, Berlin (1984), definiert sind.

[0111] Der Ausdruck „Basenpaarungsspezifität“ einer Nukleotidbase b bezieht sich auf die Anzahl von natürlichen Nukleotidbasen, mit denen die Base Watson-Crick-Basenpaare bildet. Der Begriff wird mit $S_{bp}(b)$ angegeben. Beispielsweise lautet das $S_{bp}(b)$ für die vier natürlichen Nukleotide wie folgt: $S_{bp}(A) = 1$, $S_{bp}(G) = 1$, $S_{bp}(C) = 1$ und $S_{bp}(T) = 1$.

[0112] Der Ausdruck „natürliches Komplement eines Nukleotids“ bezieht sich auf das natürliche Nukleotid, mit dem ein Nukleotid am liebsten ein Basenpaar gemäß den Watson-Crick-Basenpaarungsregeln bildet. Wenn das Nukleotid mit gleicher Affinität mit mehr als einem natürlichen Nukleotid Basenpaare bilden kann oder sich in verschiedenen Umgebungen am liebsten mit verschiedenen natürlichen Nukleotiden paart, so wird das Nukleotid als mehrere natürliche Nukleotidkomplemente aufweisend betrachtet.

[0113] Der Ausdruck „natürliches Äquivalent eines Nukleotids“ bezieht sich auf das natürliche Komplement des natürlichen Komplements des Nukleotids. In Fällen, in denen ein Nukleotid mehrere natürliche Komplemente aufweist, wird es als mehrere natürliche Äquivalente aufweisend angesehen.

[0114] Der Ausdruck „natürliches Äquivalent eines Oligonukleotid-Vorläufers“ bezieht sich auf einen Oligonukleotid-Vorläufer, bei dem jedes Nukleotid durch sein natürliches Nukleotid-Äquivalent ersetzt wurde. In Fällen, in denen ein oder mehr der ursprünglichen Nukleotide mehrere natürliche Äquivalente aufweist bzw. aufweisen, werden die Oligonukleotid-Vorläufer als mehrere natürliche Äquivalente aufweisend angesehen, wobei Äquivalente aus allen möglichen Ersatzkombinationen ausgewählt werden. Der Ausdruck wird mit $NE(\omega)$ angegeben.

[0115] Der Begriff „Nukleosid“ bezieht sich auf eine Base-Zucker-Kombination oder ein Nukleotid, dem ein Phosphatanteil fehlt.

[0116] Ein „Kettenabbruch-Nukleosid-Triphosphat“ ist ein Nukleosid-Triphosphat, das in der Lage ist, in einer Kettenverlängerungsreaktion zu einem Oligonukleotid-Primer hinzugefügt zu werden, jedoch nicht in der Lage ist, einer Kettenverlängerung unterzogen zu werden. Beispiele, die lediglich veranschaulichend und nicht einschränkend sind, umfassen die vier standardmäßigen Dideoxynukleotid-Triphosphate, massenmodifizierte Dideoxynukleotid-Triphosphat-Analoga, Thio-Analoga von natürlichen und massenmodifizierten Dideoxynukleotid-Triphosphaten, Arabanose, 3'-Amino-, 3'-Azido-, 3'-Fluor-Derivate und dergleichen.

[0117] Der Ausdruck „Dideoxynukleosid-Triphosphat“ bezieht sich auf und umfasst die vier natürlichen Dideoxynukleosid-Triphosphate (ddATP, ddGTP, ddCTP und ddTTP für DNA und ddATP, ddGTP, ddCTP und ddUTP für RNA) und massenmodifizierten Dideoxynukleosid-Triphosphate. Der Begriff kann durch ddNTP angegeben werden.

[0118] Der Ausdruck „Verlängerungs-Nukleosid-Triphosphate“ bezieht sich auf und umfasst natürliche Desoxynukleosid-Triphosphate, modifizierte Desoxynukleotid-Triphosphate, massenmodifizierte Desoxynukleosid-Triphosphate, 5'-(α)-Phosphothioat und 5'-N-(α -Phosphoramidat)-Analoga von natürlichen und massenmodifizierten Desoxy- und Ribonukleosid-Triphosphaten und dergleichen, z. B. diejenigen, die in der U.S.-Patentschrift Nr. 5,171,534 und in der U.S.-Patentschrift Nr. 5,547,835, wobei die relevanten Abschnitte derselben durch Bezugnahme in das vorliegende Dokument aufgenommen sind, offenbart sind.

[0119] Der Ausdruck „Nukleotidpolymerase“ bezieht sich auf einen Katalysator, üblicherweise ein Enzym, zum Bilden einer Verlängerung eines Polynukleotids entlang einer DNA- oder RNA-Matrize, wobei die Verlängerung dazu komplementär ist. Die Nukleotidpolymerase ist eine matrizenabhängige Polynukleotidpolymerase und verwendet Nukleosid-Triphosphate als Baublocke zum Verlängern des 3'-Endes eines Polynukleotids, um eine zu der Polynukleotidmatrize komplementäre Sequenz zu liefern. Üblicherweise sind die Katalysatoren Enzyme, z. B. DNA-Polymerasen, beispielsweise prokaryotische DNA-Polymerase (I, II oder III), T4-DNA-Polymerase, T7-DNA-Polymerase, E.coli-DNA-Polymerase (Klenow-Fragment, 3'-5'-Exo-), Revertase,

Vent-DNA-Polymerase, Pfu-DNR-Polymerase, Taq-DNA-Polymerase, Bst-DNA-Polymerase und dergleichen, oder RNA-Polymerasen, z. B. T3- und T7-RNA-Polymerasen. Polymeraseenzyme können von jeglicher Quelle wie z. B. Zellen, Bakterien wie z. B. *E. coli*, Pflanzen, Tieren, Viren, thermophilen Bakterien usw., gewonnen werden.

[0120] Eine „Amplifikation“ von Nukleinsäuren oder Polynukleotiden ist jegliches Verfahren, das zur Bildung von einer oder mehr Kopien eines Nukleinsäure- oder Polynukleotidmoleküls (exponentielle Amplifikation) oder zur Bildung von einer oder mehr Kopien lediglich des Komplements eines Nukleinsäure- oder Polynukleotidmoleküls (lineare Amplifikation) führt. Amplifikationsverfahren umfassen die Polymerasekettenreaktion (PCR), die auf wiederholten Denaturierungszyklen, Oligonukleotid-Primer-Tempern und Primer-Verlängerung anhand einer thermophilen matrizenabhängigen Polynukleotidpolymerase beruht, was zur exponentiellen Zunahme von Kopien der gewünschten Sequenz des durch die Primer flankierten Polynukleotidanalyten führt. Die zwei verschiedenen PCR-Primer, die zu gegenüberliegenden Strängen der DNA temvern, sind so positioniert, dass das Polymerasekatalysierte Verlängerungsprodukt eines Primers als Matrizenstrang für den anderen dienen kann, was zur Akkumulierung eines charakteristischen doppelsträngigen Fragments führt, dessen Länge durch den Abstand zwischen den 5'-Enden der Oligonukleotid-Primer definiert wird. Die Reagenzien zum Durchführen einer derartigen Amplifikation umfassen Oligonukleotid-Primer, eine Nukleotidpolymerase und Nukleosid-Triphosphat wie z. B. Desoxyadenosin-Triphosphat (dATP), Desoxyguanosin-Triphosphat (dGTP), Desoxycytidin-Triphosphat (dCTP) und Desoxythymidin-Triphosphat (dTTP). Andere Verfahren zur Amplifikation umfassen eine Amplifikation eines einsträngigen Polynukleotids unter Verwendung eines einzigen Oligonukleotid-Primers, die Ligasekettenreaktion (LCR), die auf einer Nukleinsäuresequenz basierte Amplifikation (NAS-BA), das Q-Beta-Replikase-Verfahren und 3SR.

[0121] Die Begriffe „Hybridisierung (Hybridisieren)“ und „Binden“ im Kontext von Nukleotidsequenzen werden in dem vorliegenden Dokument austauschbar verwendet. Die Fähigkeit zweier Nukleotidsequenzen, miteinander zu hybridisieren, beruht auf dem Grad der Komplementarität der beiden Nukleotidsequenzen, der wiederum auf dem Anteil übereinstimmender komplementärer Nukleotidpaaren beruht. Je mehr Nukleotide in einer gegebenen Sequenz zu einer anderen Sequenz komplementär sind, desto strenger können die Bedingungen für die Hybridisierung sein, und desto spezifischer ist die Bindung der zwei Sequenzen. Eine erhöhte Strenge wird erzielt, indem die Temperatur erhöht, das Verhältnis von Hilfslösungsmitteln erhöht, die Salzkonzentration verringert wird und dergleichen.

[0122] Der Begriff „komplementär“, „Komplement“ oder „komplementäre Nukleinsäuresequenz“ bezieht sich auf den Nukleinsäurestrang, der anhand der Watson-Crick-Basenpaarungsregeln auf die Basensequenz in einem anderen Nukleinsäurestrang bezogen ist. Allgemein sind zwei Sequenzen komplementär, wenn sich die Sequenz des Einen in einer antiparallelen Richtung an die Sequenz des Anderen binden kann, wobei sich das 3'-Ende jeder Sequenz an das 5'-Ende der anderen Sequenz bindet, und jedes A, T(U), G und C einer Sequenz anschließend mit T(U), A, C bzw. G der anderen Sequenz ausgerichtet wird. RNA-Sequenzen können auch komplementäre G/U- oder U/G-Basenpaare umfassen.

[0123] Der Begriff „Hybrid“ bezieht sich auf ein doppelsträngiges Nukleinsäuremolekül, das durch eine Wasserstoffbrückenbindung zwischen komplementären Nukleotiden gebildet wird. Der Begriff „Hybridisieren“ bezieht sich auf den Prozess, anhand dessen einzelne Stränge von Nukleinsäuresequenzen durch eine Wasserstoffbrückenbindung zwischen komplementären Nukleotiden doppelhelixförmige Segmente bilden.

[0124] Der Begriff „massenmodifiziert“ bezieht sich auf eine Nukleinsäuresequenz, deren Masse entweder durch eine interne Veränderung – d. h. durch eine Addition, eine Deletion oder eine Substitution eines chemischen Anteils – ihrer chemischen Struktur oder durch eine externe Veränderung – d. h. durch die Addition eines chemischen Anteils (Atom oder Molekül), der kovalent angelagert wurde – ihrer chemischen Struktur verändert wurde. Der chemische Anteil wird somit als Massenmodifizierungsanteil bezeichnet.

[0125] Der Ausdruck „Massenzahl eines Atoms“ bezieht sich auf die Nukleonenzahl des häufigsten Isotops des interessierenden Elements.

[0126] Die berichtete Masse für alle Nukleinsäuren (d. h. Nukleotide, Nukleotidvorläufer, Oligonukleotide, X-mer und X-mer-Produkte) wird unter Verwendung der Massenzahlen für die am häufigsten auftretenden Isotope der konstituierenden Atome (d. h. C12, N14, H1, O16, P31, I127) und eines Protonierungszustands, der in einer wässrigen Lösung bei einem pH-Wert von 7 stabil ist, berechnet.

[0127] Der Ausdruck „Massenzahl eines Oligonukleotid-Vorläufers“ bezieht sich auf die Summe der Massen-

zahlen der konstituierenden Atome der Oligonukleotid-Vorläufer. Der Ausdruck wird mit $z(\omega)$ bezeichnet.

[0128] Der Ausdruck „Massenzahlhistogramm eines Gemischs von Oligonukleotid-Vorläufern“ Ω bezieht sich auf eine Funktion h von den natürlichen Zahlen bis zu den natürlichen Zahlen, die durch $h(z)$ definiert sind, wobei $h(z)$ die Anzahl von Oligonukleotid-Vorläufern in dem Gemisch Ω ist, wofür gilt: $z(\omega) = z$.

[0129] Der Ausdruck „durchschnittliche Mehrdeutigkeit eines Gemischs von Oligonukleotid-Vorläufern“ $(A(\Omega))$ bezieht sich auf die Summe der Quadrate der Werte des Massenzahlhistogramms des Gemischs von Oligonukleotid-Vorläufern geteilt durch die Anzahl von Oligonukleotid-Vorläufern in dem Gemisch, und kann mathematisch wie folgt ausgedrückt werden:

$$A(\Omega) = 1 / N \sum_z h(z)^2$$

[0130] Der Ausdruck „Massenzahlkomplexität“ (MNC – mass number complexity) bezieht sich auf die Anzahl von Oligonukleotid-Vorläufern in dem Gemisch geteilt durch die durchschnittliche Mehrdeutigkeit des Gemischs von Oligonukleotid-Vorläufern, und kann mathematisch wie folgt definiert werden:

$$MNC(\Omega) = N/A(\Omega)$$

[0131] Der Ausdruck „Oligonukleotid-Bedeckungskomplexität“ $CC_o(\omega)$ kann mathematisch wie folgt ausgedrückt werden:

$$CC_o(\omega) = 1 / 4^L \prod_{i=1}^L S_{bp}(b_i)$$

wobei L die Anzahl von Nukleotidbasen in dem Oligonukleotid-Vorläufer ist und b_i die i -te Einheit des Oligonukleotid-Vorläufers darstellt.

[0132] Der Ausdruck „Gemischbedeckungskomplexität“ ($CC_M(\Omega)$) bezieht sich auf die Summe der Bedeckungskomplexitäten jedes der Oligonukleotid-Vorläufer in dem Gemisch und kann mathematisch wie folgt ausgedrückt werden:

$$CC_M(\Omega) = \sum_{\omega \in \Omega} CC_o(\omega)$$

[0133] Der Begriff „Kategorisierung“ bezieht sich auf die Einteilung eines Gemischs in definierte Teilsatzgemische, wobei jedes einzelne Oligonukleotid des Gemischs in zumindest einem Teilsatzgemisch erscheint.

[0134] Der Ausdruck „zusammengesetzte Gemischbedeckungskomplexität“ bezieht sich auf die Bedeckungskomplexität eines Satzes von Gemischen, die durch eine Kategorisierung erzeugt wird und gleich der Gemischbedeckungskomplexität des ursprünglichen, nicht-kategorisierten Gemisches ist.

[0135] Der Ausdruck „zusammengesetzte Massenzahlkomplexität“ bezieht sich auf die Massenzahlkomplexität eines Satzes von Gemischen, die anhand einer Kategorisierung erzeugt wird und die gleich der Summe der Massenzahlkomplexitäten der Teilsatzgemische ist.

[0136] Der Ausdruck „direkte Massenspektralanalyse“ bezieht sich auf ein Verfahren einer Massenspektralanalyse, das entweder die Targetnukleinsäuresequenz selbst oder das Komplement der Targetnukleinsäuresequenz analysiert. Die Targetnukleinsäuresequenz selbst oder ihr Komplement kann massenmodifiziert werden, zusätzliche Nukleotidbasen modifiziert werden, zusätzliche Nukleotidbasen enthalten oder auf andere Weise modifiziert sein, vorausgesetzt, dass die Targetnukleinsäuresequenz oder ihr Komplement tatsächlich einer Massenanalyse unterzogen wird. Jedoch umfasst der Ausdruck keine Massenspektralanalyse, bei der ein Massenmarkierungsanteil, der das Vorhandensein einer Targetnukleinsäuresequenz angibt, analysiert wird, z. B. diejenigen indirekten Verfahren, die in der PCT-Anmeldung WO 95/04160 beschrieben sind.

[0137] Der Begriff „Generizität“ oder „generisch“ bezieht sich, wenn er auf ein Verfahren angewendet wird, auf ein Verfahren einer Massenspektralanalyse, das ohne Bezugnahme auf bestimmte Informationen angewendet werden kann. Der Ausdruck „positionsbezogene Generizität“ bezieht sich auf Massenspektralanalyse-Verfahren, die keine Vorab-Informationen über das Vorliegen, die Position oder die Identität von Mutationen in der Targetnukleinsäuresequenz erfordern. Der Ausdruck „Targetgenerizität“ bezieht sich auf Massenspektralanalyse-Verfahren, die keine Vorab-Informationen über die Targetnukleinsäure erfordern.

[0138] Der Begriff „Träger“ oder „Oberfläche“ bezieht sich auf ein poröses oder nicht-poröses wasserunlösliches Material. Die Oberfläche kann eine beliebige einer Anzahl von Formen aufweisen, z. B. Streifen, Auflage, Platte, Stab, Partikel, einschließlich Kugelchen, und dergleichen. Der Träger kann hydrophil sein oder er kann in der Lage sein, hydrophil gemacht zu werden, und umfasst anorganische Pulver wie z. B. Siliziumdioxid, Magnesiumsulfat und Aluminiumoxid; natürliche polymere Materialien, insbesondere zellulose Materialien und aus Zellulose gewonnene Materialien, z. B. faserhaltige Papiere, z. B. Filterpapier, chromatographisches Papier usw.; synthetische oder modifizierte natürlich vorkommende Polymere, z. B. Nitrozellulose, Zelluloseacetat, Poly(vinylchlorid), Polyacrylamid, vernetztes Dextran, Agarose, Polyacrylat, Polyethylen, Polypropylen, Poly(4-methylbuten), Polystyren, Polymethacrylat, Poly(ethylenterephthalat), Nylon, Poly(vinylbutyrat) usw.; die entweder alleine oder in Verbindung mit anderen Materialien verwendet werden; Glas, das als Bioglass erhältlich ist, Keramik, Metalle und dergleichen. Natürliche oder synthetische Vereinigungen wie z. B. Liposome, Phospholipidbläschen und Zellen können ebenfalls verwendet werden. Ein Binden von Oligonukleotiden an einen Träger oder eine Oberfläche kann anhand von hinreichend bekannten Techniken, die gemeinhin in der Literatur zur Verfügung stehen, bewerkstelligt werden. Siehe z. B. A. C. Pease u. a., Proc. Nat. Acad. Sci. USA, 91: 5.022 – 5.026 (1994).

[0139] Der Begriff „Mutation“ bezieht sich auf eine auf Nukleotide bezogene Variation zwischen zwei Polynukleotiden, z. B. bei Einzelnukleotid-Polymorphismen. Allgemein treten die Variationen von Individuum zu Individuum auf. Die Mutation kann eine Veränderung der Sequenz von Nukleotiden einer auf normale Weise konservierten Nukleinsäuresequenz sein, die zur Bildung einer Mutanten führt, wie sie von der normalen (unveränderten) oder Wildtyp-Sequenz unterscheiden wird. Mutationen können allgemein in zwei allgemeine Klassen unterteilt werden, nämlich Basenpaar-Substitutionen und Frameshift-Mutationen. Die Letztgenannten beinhalten die Einfügung oder Deletion von einem bis mehreren Nukleotidpaaren. Eine Differenz eines einzigen Nukleotids kann bedeutend sein, um den Phänotyp von der Normalität zur Abnormalität zu verändern, z. B. im Fall der Sichelzellenanämie.

Allgemeine Anmerkungen

[0140] Die vorliegende Erfindung liefert Verfahren und Reagenzien, um den Bedarf an empfindlicheren, genaueren und einen höheren Durchsatz ermöglichen Analysen von Targetnukleinsäuresequenzen zu decken. Die Verfahren und Reagenzien können generisch auf allgemein jegliche Targetnukleinsäuresequenz angewendet werden und erfordern keine Vorab-Informationen über das Vorliegen, die Position oder Identität von Mutationen in der Targetnukleinsäuresequenz.

[0141] Die Reagenzien der Erfindung, die für eine direkte Massenspektralanalyse von Nukleinsäuren nützlich sind, sind Gemische, die natürliche und massenmodifizierte X-mer-Vorläufer umfassen, die eine minimale Länge von drei Nukleotiden aufweisen. Die minimale Gemischbedeckungskomplexität (CC_M) der Gemische ist 56 geteilt durch die Anzahl von gesonderten X-meren in dem Gemisch. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Jeder der X-mer-Vorläufer in einem Gemisch wird durch eine einzelne chemische Spezies dargestellt.

[0142] Die Verfahren und Reagenzien der vorliegenden Erfindung verringern die bei der Massenspektralanalyse einer Targetnukleinsäuresequenz vorliegenden Mehrdeutigkeiten und erhöhen somit bei allen Anwendungen, die die Massenspektrometrie verwenden, die Leistungsfähigkeit, die Sequenz der Targetnukleinsäure zu analysieren. Diese Verringerung wird bewerkstelligt, indem ein Gemisch von natürlichen und massenmodifizierten Oligonukleotid-Vorläufern eingesetzt wird, die ein hohes Maß an Massen- und Bedeckungskomplexität aufweisen. Diese Verringerung kann durch eine „Kategorisierung“, d. h. durch Verwenden von Teilsätzen der Gemische in zumindest zwei Reaktionsgemischen, weiter verbessert werden. Die Ergebnisse der separaten Abfragung bei den Teilsatzgemischen könnten anschließend kombiniert werden. Auf diese Weise wird das Ausmaß einer Massenüberlappung zwischen X-mer-Produkten bei einer gegebenen Massenanalyse verringert, während gleichzeitig ein hoher Grad an Gesamtbedeckungskomplexität des Targets aufrechterhalten wird.

[0143] Die Gemische der Erfindung sind insofern generisch oder universell, als sie bei jeglicher Anwendung verwendet werden können, deren Ziel darin besteht, Sequenzinformationen einer Targetnukleinsäure zu bestimmen. Ferner können die Gemische ohne Bezugnahme auf jegliche Vorab-Informationen über die Targetnukleinsäuresequenz, einschließlich z. B. des Vorliegens, der Position oder der Identität einer Mutation, entworfen werden. Dies soll jedoch nicht implizieren, dass die Gemische beim Analysieren von Targetnukleinsäuresequenzen nicht nützlich wären, bei denen manche Informationen über die Sequenz vorab bekannt sind. Noch impliziert dies, dass Vorab-Informationen über das Target bei einer Analyse der sich ergebenden Mas-

senspektra nicht sinnvoll eingesetzt werden können.

Reagenzien der Erfindung

Oligonukleotid-Vorläufer (X-mer-Vorläufer)

[0144] Die Oligonukleotid-Vorläufer-Reagenzien (X-mer-Vorläufer-Reagenzien) der Erfindung sind Gemische von natürlichen und massenmodifizierten X-mer-Vorläufern, die eine Mindestlänge von drei Nukleotiden und eine Gemischbedeckungskomplexität von etwa 7/8 aufweisen, wenn das Gemisch zumindest 56 gesonderte X-mer-Vorläufer enthält. Wenn die durchschnittliche Länge des X-mer-Vorläufers zunimmt, nimmt auch die Anzahl gesonderter X-mere in den Gemischen der vorliegenden Erfindung zu, und die Gemischbedeckungskomplexität mag abnehmen. Die Untergrenze der Gemischbedeckungskomplexität ist gleich 56 geteilt durch die Anzahl von X-meren in dem Gemisch. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden.

[0145] Die jeweilige Zusammensetzung des Gemischs wird von Fall zu Fall bestimmt und hängt von den Anforderungen der gegebenen Anwendung ab. Die Zusammensetzung eines Gemischs wird durch die hierin dargelegten Gleichungen definiert. Die Gemischbedeckungskomplexität ist wie folgt definiert:

$$CC_M(\Omega) = \sum_{\omega \in \Omega} CC_o(\omega)$$

wobei CC_o die Oligonukleotid-Bedeckungskomplexität jedes der Oligonukleotid-Vorläufer in dem Gemisch ist und wie folgt definiert ist:

$$CC_o(\omega) = 1 / 4^L \prod_{i=1}^L S_{bp}(b_i)$$

wobei L die Anzahl von Nukleotidbasen in dem Oligonukleotid-Vorläufer ist, S_{bp} die Basenpaarungsspezifität ist und b_i die i-te Einheit des Oligonukleotid-Vorläufers darstellt.

[0146] Beispiele von Gemischen, die die oben beschriebenen Spezifikationen aufweisen, umfassen veranschaulichungs- und nicht einschränkungshalber: (1) ein Gemisch Ω_1 , das aus 60 der möglichen 64 3-mere besteht ($CC_M(\Omega_1) = 7/8$); (2) ein Gemisch Ω_2 , das aus 128 der möglichen 256 4-mere besteht ($CC_M(\Omega_2) = 1/2$); (3) ein Gemisch Ω_3 , das aus 256 der möglichen 1.024 5-mere besteht ($CC_M(\Omega_3) = 1/4$); (4) ein Gemisch Ω_4 , das aus 512 der möglichen 4.096 6-mere besteht ($CC_M(\Omega_4) = 1/8$); (5) ein Gemisch Ω_5 , das aus 1.024 der möglichen 4.096 6-mere besteht ($CC_M(\Omega_5) = 1/4$); (6) ein Gemisch Ω_6 , das aus 48 5-meren und 512 6-meren besteht ($CC_M(\Omega_6) = 11/64$); (7) ein Gemisch Ω_7 , das aus 128 5-meren, 512 6-meren und 128 7-meren besteht ($CC_M(\Omega_7) = 33/128$); (8) ein Gemisch Ω_8 , das aus 256 5-meren, 1.000 6-meren und 96 7-meren besteht ($CC_M(\Omega_8) = 1/2$).

[0147] Beispiele von Gemischen, die den obigen Spezifikationen nicht entsprechen, umfassen veranschaulichungs- und nicht einschränkungshalber: (1) ein Gemisch Ω_9 , das aus 64 der möglichen 256 4-mere besteht ($CC_M(\Omega_9) = 1/4 < 56/64$), (2) ein Gemisch Ω_{10} , das aus 128 der möglichen 1.024 5-mere besteht ($CC_M(\Omega_{10}) = 1/8 < 56/128$), (3) ein Gemisch Ω_{11} , das aus 384 6-meren und 128 7-meren besteht ($CC_M(\Omega_{11}) = 13/128 < 56/512$), (4) ein Gemisch Ω_{12} , das aus 64 5-meren, 256 6-meren und 64 7-meren besteht ($CC_M(\Omega_{12}) = 33/256 < 56/384$).

[0148] Außerdem sind die Reagenzien der Erfindung Gemische von natürlichen und massenmodifizierten X-mer-Vorläufern, wobei die Massenzahlkomplexität (MNC) des Gemischs größer ist als die Massenzahlkomplexität jedes natürlichen Äquivalents des Gemisches. Die Massenzahlkomplexität bezieht sich auf die Anzahl von X-mer-Vorläufern in dem Gemisch geteilt durch die durchschnittliche Mehrdeutigkeit des Gemischs von X-mer-Vorläufern, und kann mathematisch wie folgt definiert werden:

$$MNC(\Omega) = N/A(\Omega)$$

[0149] Die durchschnittliche Mehrdeutigkeit des Gemisches von X-mer-Vorläufern ($A(\Omega)$) bezieht sich auf die Summe der Quadrate der Werte des Massenzahlhistogramms des Gemischs von X-mer-Vorläufern geteilt durch die Anzahl von X-mer-Vorläufern in dem Gemisch, und kann mathematisch wie folgt ausgedrückt werden:

$$A(\Omega) = 1 / N \sum_z h(z)^2$$

[0150] Das Massenzahlhistogramm eines Gemisches von X-mer-Vorläufern ($h(z)$) bezieht sich auf eine Funktion h von den natürlichen Zahlen zu den durch $h(z)$ definierten natürlichen Zahlen, wobei $h(z)$ die Anzahl von X-mer-Vorläufern in dem Gemisch Ω ist, wofür gilt: $z(\omega) = z$.

[0151] Üblicherweise beträgt die MNC des Gemisches zumindest etwa das Doppelte, noch üblicher zumindest etwa das Zehnfache und am stärksten bevorzugt zumindest etwa das Fünfzigfache der Massenzahlkomplexität jeglichen natürlichen Äquivalents des Gemisches. Beispielsweise weist das Gemisch aller natürlichen 4.096 6-mere eine MNC von 53 auf (siehe nachfolgende Erörterung und Tabelle 1). Ein Gemisch, das alle 4.096 6-mere enthält, die auf kombinatorische Weise synthetisiert werden, kann eine MNC von 348 aufweisen, was etwa das 6,5-fache des natürlichen Äquivalents ist. Ein anderes Gemisch, bei dem jedes X-mer individuell synthetisiert wird, kann eine MNC von 559 aufweisen, was etwa das Zehnfache des natürlichen Äquivalents ist. Ein Gemisch, bei dem jedes der 4.096 6-mere eine eindeutige Masse besitzt, würde eine MNC von 4.096 aufweisen, was etwa das 77-fache des natürlichen Äquivalents ist.

[0152] Die bei dem Verfahren der Erfindung nützlichen X-mer-Vorläufer weisen eine Länge von zumindest drei Nukleotideinheiten auf. Vorzugsweise weisen die X-mer-Vorläufer eine Länge von zumindest vier Nukleotideinheiten, stärker bevorzugt zumindest fünf Nukleotideinheiten und am stärksten bevorzugt zumindest sechs Nukleotideinheiten auf. Die Länge des X-mer-Vorläufers kann für jeden X-mer-Vorläufer in dem Gemisch unabhängig ausgewählt werden. Somit ist es möglich, ein einziges Gemisch von X-mer-Vorläufern zu haben, die Längen von fünf, sechs und sieben Nukleotiden aufweisen. Wie aus der obigen Erörterung hervorgeht, nehmen der Wert der Gemischbedeckungskomplexität und somit die Anforderungen an dieselbe mit zunehmender Länge des X-mer-Vorläufers ab. In Fällen, in denen ein einziges Gemisch mehr als eine Länge besitzt, wird die Bedeckungskomplexität des Gemisches erhalten, indem die Bedeckungssindizes der einzelnen Oligonukleotide summiert werden. Somit würde in diesem Fall der Beitrag jedes Oligonukleotids zu der Bedeckungskomplexität des Gemisches von seiner Länge abhängen: kürzere Oligonukleotide tragen mehr bei. Man sollte beachten, dass eine Verwendung von langen Oligonukleotiden zu einem Verlust an Generizität führen kann. Niedrigere Gemischbedeckungskomplexitätswerte mögen nur dann verwendet werden, wenn ein Verlust an Generizität toleriert werden kann. Ferner können die Reagenzien einen Satz von Gemischen von Oligonukleotid-Vorläufern umfassen. In diesem Fall kann die Gemischbedeckungskomplexität eines beliebigen Angehörigen des Satzes niedriger als die oben beschriebene sein, solange die Gesamtkomplexität des Gemisches der obigen Beschreibung entspricht.

[0153] Die bei dem Verfahren der Erfindung nützlichen X-mer-Vorläufer können jeweils durch eine einzige chemische Spezies dargestellt werden, statt durch eine Anzahl von Varianten von ähnlichen chemischen Spezies dargestellt zu werden, z. B. die Leiter von Reporterprodukten, die verwendet werden, um die Nukleotidsequenz in dem Oligonukleotid darzustellen, das in der PCT-Anmeldung WO 95/04160 (Southern) beschrieben ist. Somit besitzt jeder X-mer-Vorläufer in dem Gemisch der Erfindung eine einzige Masse, wohingegen jedem Oligonukleotid in dem Gemisch der WO 95/04160 ein Spektrum von Massen zugeordnet ist, die die interessierende Nukleotidsequenz darstellen, wie oben erörtert wurde. Es ist wichtig zu erkennen, dass der von Southern offenbarte Massenmarkierung-Lösungsansatz spaltbare Massenmarken bzw. Massenmarkierungen verwendet, bei denen lediglich der markierte Abschnitt der markierten Oligonukleotide in dem Massenspektrometer analysiert wird. Wie aus der Offenbarung des vorliegenden Dokuments hervorgeht, steht dies im Gegensatz zur vorliegenden Erfindung, die sich auf ein Erzeugen eines Massenspektrums der Oligonukleotidprodukte selbst stützt, die aus einem mittels eines Targets bewirkten enzymatischen Prozess resultieren. Überdies ist das Gemisch von massenmodifizierten X-meren bei der vorliegenden Erfindung so entworfen, dass jegliche gegebene Oligonukleotidsequenz lediglich eine einzige Masse besitzt, d. h. eine einzige chemische Spezies darstellt. Dies ist bei dem von Southern offenbarten Massenmarkierung-Lösungsansatz nicht der Fall. Aufgrund des „Leitermarke“-Entwurfs des Lösungsansatzes von Southern ist jeder diskreten Oligonukleotidsequenz in dem Gemisch ein „Spektrum“ von Massenentitäten zugeordnet.

[0154] Um bei den Verfahren der vorliegenden Erfindung nützlich zu sein, ist es wünschenswert und oft notwendig, zu wissen, welche X-mer-Vorläufer in dem Gemisch vorliegen. Jedoch ist es nicht absolut notwendig, den Spiegel jedes X-mer-Vorläufers zu kennen. Trotzdem ist es vorteilhaft, in der Lage zu sein, die Konzentration jedes X-mers in dem Gemisch zu steuern, um Unterschiede bei Doppel-Thermostabilitäten zu kompensieren (siehe nachfolgende Erörterung).

[0155] Bei einem bevorzugten Ausführungsbeispiel bestehen die Vorläufer-X-mer-Gemische sowohl aus natürlichen als auch aus massenmodifizierten Nukleotiden. Die Identität und die Position von massenmodifizier-

ten Nukleotiden in den X-mer-Vorläufern hängen von einer Anzahl von Faktoren ab. Diese umfassen: die gewünschte Gesamtmassekomplexität des X-mer-Vorläufer-Gemischs, die gewünschten thermodynamischen Eigenschaften des X-mer-Vorläufers, die Fähigkeit eines Enzyms oder eines Satzes von Enzymen (d. h. Polymerasen und Ligasen), massenmodifizierte Nukleotide in dem X-mer-Vorläufer zu beherbergen, und die Einschränkungen, die das jeweilige Syntheseverfahren des X-mer-Vorläufer-Gemischs auferlegt.

[0156] Die X-mer-Vorläufer können entweder durch eine interne Veränderung – d. h. durch eine Addition, Deletion oder Substitution eines chemischen Anteils – ihrer chemischen Struktur oder durch eine externe Veränderung – d. h. durch die Addition eines chemischen Anteils (Atom oder Molekül), der kovalent angelagert wird – ihrer chemischen Struktur massenmodifiziert werden. Ein X-mer-Vorläufer kann sowohl eine interne Veränderung als auch eine externe Veränderung, mehr als eine interne Veränderung, mehr als eine externe Veränderung oder eine bestimmte Kombination derselben aufweisen.

[0157] Geeignete interne Massenmodifizierungen umfassen zumindest eine chemische Modifizierung der Internukleosidbindung, der Zuckerhauptkette oder der Nukleosidbase des X-mer-Vorläufers. Beispiele von geeigneten, intern massenmodifizierten X-mer-Vorläufern umfassen, als Veranschaulichung und nicht als Einschränkung, diejenigen, die 2'-Desoxy-5-methylcytidin, 2'-Desoxy-5-fluorcytidin, 2'-Desoxy-5-iodocytidin, 2'-Desoxy-5-fluoruridin, 2'-Desoxy-5-iodouridin, 2'-O-Methyl-5-fluoruridin, 2'-Desoxy-5-ioduridin, 2'-Desoxy-5(1-propynyl)uridin, 2'-O-Methyl-5(1-propynyl)uridin, 2-Thiothymidin, 4-Thiothymidin, 2'-Desoxy-5(1-propynyl)cytidin, 2'-O-Methyl-5(1-propynyl)cytidin, 2'-O-Methyladenosin, 2'-Desoxy-2,6-diaminopurin, 2'-O-Methyl-2,6-diaminopurin, 2'-Desoxy-7-deazadenosin, 2'-Desoxy-6methyladenosin, 2'-Desoxy-8-oxoadenosin, 2'-O-Methylguanosin, 2'-Desoxy-7-deazaguanosin, 2'-Desoxy-8-oxoguanosin, 2'-Desoxyinosin oder dergleichen umfassen.

[0158] Geeignete externe Massenmodifizierungen umfassen ein Massenmarkieren der X-mer-Vorläufer oder Dideoxy-Abbrecher. Externe Massenmodifizierungsanteile können an das 5'-Ende des X-mers, an die Nukleotidbase (oder -basen), an die Phosphathauptkette, an die 2'-Position des Nukleosids (der Nukleoside), an die endständige 3'-Position und dergleichen angelagert werden. Geeignete externe Massenmodifizierungsanteile umfassen beispielsweise ein Halogen, ein Azido, Nitro, Alkyl, Aryl, Schwefel, Silber, Gold, Platin, Quecksilber, Massenanteile des Typs W-R, wobei W eine Bindungsgruppe ist und R ein Massenmodifizierungsanteil ist, und dergleichen.

[0159] Die Bindungsgruppe W ist an der kovalenten Bindung zwischen dem Nukleotid oder Nukleosid und R beteiligt und variiert je nach der Beschaffenheit der Moleküle. Funktionelle Gruppen, die normalerweise an den Molekülen vorliegen oder eingeführt werden, werden zum Binden bzw. Verknüpfen verwendet. Die Bindungsgruppen können von einer Bindung an eine Kette von 1 bis 100 Atomen, üblicherweise von etwa 1 bis 60 Atomen, vorzugsweise 1 bis 40 Atomen, stärker bevorzugt 1 bis 20 Atomen, variieren, wobei Jede unabhängig aus der Gruppe ausgewählt ist, die normalerweise aus Kohlenstoff, Wasserstoff, Sauerstoff, Schwefel, Stickstoff, Halogen und Phosphor besteht. Die Atome in der Kette können mit anderen Atomen als Wasserstoff substituiert werden. Als allgemeine Regel kann die Länge einer bestimmten Bindungsgruppe willkürlich ausgewählt werden, um eine Zweckmäßigkeit bezüglich der Synthese und der Integration jeglicher gewünschten Gruppe zu liefern. Die Bindungsgruppen können aliphatisch oder aromatisch sein, obwohl bei Diazo-Gruppen üblicherweise aromatische Gruppen beteiligt sind. Die massenmodifizierende R-Gruppe kann aus der Bindungsgruppe selbst bestehen, oder ein separater Massenmodifizierungsanteil kann anhand von Verfahren, die Fachleuten bekannt sind, angelagert sein. Übliche Funktionalitäten, die in den Bindungsgruppen vorliegen, die eine kovalente Bindung zwischen dem Massenmodifizierungsanteil R und dem zu konjugierenden Molekül bilden, umfassen Alkylamin, Amidin, Thioamid, Ether, Carbamat, Harnstoff, Thioharnstoff, Guanidin, Azo, Thioether und Carboxylat, Sulfonat und Phosphatester, Amide und Thioether. Wenn beispielsweise ein Amin und eine Carbonsäure oder ihr Stickstoffderivat oder ihre Phosphorsäure verbunden werden, entstehen Amide, Amidine und Phosphoramidate. Wenn Mercaptan und aktiviertes Olefin verbunden werden, entstehen Thioether. Wenn ein Mercaptan und ein Alkylierungsmittel verbunden werden, entstehen Thioether. Wenn Aldehyd und ein Amin unter Reduktionsbedingungen verbunden werden, entsteht ein Alkylamin. Wenn eine Carbonsäure oder eine Phosphatsäure und ein Alkohol verbunden werden, entstehen Ester.

[0160] Fachleuten leuchten andere geeignete Massenmodifizierungen ein, einschließlich derjenigen, die bei Oligonucleotides and Analogues, A Practical Approach, F. Eckstein (Herausgeber), IRL Press, Oxford (1991); in der U.S.-Patentschrift Nr. 5,605,798; und in der japanischen Patentschrift Nr. 59-131909, die durch Bezugnahme in das vorliegende Dokument aufgenommen sind, offenbart sind.

[0161] Ein vorrangiges Ziel der Erfindung besteht darin, entweder vollständige Gemische oder Sätze von Ge-

mischen zu erzeugen, die die verfügbare Massenbandbreite des Massenspektrometers verwenden, mit dem gleichzeitigen Ziel, eine Massenüberlappung zwischen Oligonukleotiden, die unterschiedliche Basenpaarungsmuster (Sequenzen) aufweisen, zu verringern. Wie aus der Erörterung des vorliegenden Dokuments hervorgehen sollte, geben die Menge und die Art der Informationen, die bei einer gegebenen Analyse gesucht werden, die Art des benötigten X-mer-Gemisches an.

[0162] Als Nächstes werden drei Verfahren zum Synthetisieren von X-mer-Vorläufer-Gemischen beschrieben. Dies erfolgt der Veranschaulichung, nicht der Einschränkung halber. Jedes der hierin beschriebenen Verfahren weist je nach dem Grad der synthetischen Steuerung der einzelnen Oligonukleotide, die erforderlich ist, bestimmte Vorteile auf. Alle drei Verfahren verwenden standardmäßige Phosphoramiditchemien oder enzymatische Reaktionen, die in der Technik bekannt sind. Es wird betrachtet, dass unterschiedliche Arten von massenmodifizierten Nukleotid-Vorläufer-Gemischen für definierte Arten von Anwendungen synthetisiert werden können. Beispielsweise kann ein definiertes Gemisch, das problemlos und kostengünstig herzustellen ist, für Untersuchungen vom Typ eines hohen Durchsatzes und einer niedrigen Auflösung verwendet werden. Komplexe Gemische, die kostspieliger herzustellen sind, können für Anwendungen vom Typ einer höheren Auflösung reserviert werden.

[0163] Die X-mer-Vorläufer können mittels herkömmlicher Techniken synthetisiert werden, einschließlich Verfahren, die eine Phosphoramidit-Chemie einsetzen, einschließlich sowohl 5'-bis-3'- als auch 3'-bis-5'-Syntheserouten. Beispielsweise sind zum Synthetisieren aller 6-mer 4.096 separate Synthesen notwendig. Eine Verwendung eines automatisierten Roboter-Arbeitsplatzes erleichtert diesen Prozess. Dieses Verfahren ermöglicht eine vollständige synthetische Steuerung jedes einzelnen X-mer-Vorläufers bezüglich der Zusammensetzung und Länge. Dies ist notwendig, um X-mer-Gemische zu erzeugen, die die größtmögliche MNC aufweisen. Eine individuelle Synthese ermöglicht ferner eine QC-Analyse jedes X-mers, was die Endprodukt herstellung unterstützt. Einzelne Proben jedes X-mers aufzuweisen ermöglicht ferner, dass definierte Teilsatzgemische erzeugt werden, um die Zusammensetzungsauflösung zu erhöhen. Ferner ermöglicht es, dass jedes X-mer in einer festgelegten Konzentration in dem Gemisch vorliegt. Dies kann potentiell hilfreich dabei sein, unterschiedliche Wärmestabilitäten, die für jedes X-mer/Target-Duplex erwartet werden, zu kompensieren.

[0164] Unter Verwendung einer standardmäßigen Feststoffträger-Phosphoramidit-Chemie und einer definierten Serie von 25 %igen Gemischen jedes Phosphoramidits vom Typ A, C, G und T können die X-mer-Vorläufer parallel oder in einer Einzelsynthese hergestellt werden. Beispielsweise kann eine Synthese schrittweise durchgeführt werden, wobei von einem 25 %igen Gemisch aus jeweils einem 3'-CPG-gebundenen 5'DMT-geschützten A-, G-, C- und T-Nukleosid ausgegangen wird. Für die Synthese eines Gemisches aus allen 4.096 6-meren werden zur Verwendung bei jeder der fünf Kondensationsreaktionen fünf Flaschen hergestellt, die ein 25 %iges Gemisch jeweils eines Phosphoramidits vom Typ A, G, C bzw. T enthalten. Beispielsweise enthält die Flasche für den ersten Kondensationsschritt ein 25 %iges Moläquivalent der Phosphoramidite, die 2'-O-Methyl-2,6-diaminopurin, 2'-O-Methylguanosin, 2'-Desoxy-5-iodocytidin und Thymidin entsprechen. Die Flasche für die zweite Kondensationsreaktion enthält ein 25 %iges Moläquivalent der Phosphoramidite, die 2'-Desoxyadenosin, 2'-Desoxy-7-deazaguanosin, 2'-O-Methyl-5(1-propynyl)cytidin und 2'-Desoxy-5-fluoruridin entsprechen. Ähnliche 25 %ige Gemische anderer Arten von modifizierten A-, G-, C- und T-Phosphoramiditen werden für die drei verbleibenden Kondensationsschritte erzeugt.

[0165] Obwohl dies von einem Standpunkt der Synthese aus betrachtet ein relativ einfacher Lösungsansatz ist, erlegt er dem resultierenden abschließenden X-mer-Gemisch gewisse wechselseitige Massenabhängigkeiten auf. Bei dem obigen Beispiel weisen alle 1.024 6-mer-Sequenzen, die ein A-Typ-Nukleotid in der zweiten Position erfordern, in dieser Position ein 2'-O-Methyl-2,6-diaminopurin auf. Alle Sequenzen, die in der dritten Position ein G erfordern, weisen ein 7-Deazaguanosin auf, und so weiter. Obwohl dieses Syntheseschema die Mehrdeutigkeit relativ zur Verwendung von natürlichen Basen insgesamt verringert, schränken die sich ergebenden positionsbezogenen wechselseitigen Abhängigkeiten in dem Gemisch die Auflösungsleistung für die abschließende Untersuchung ein. Dies geht aus den berechneten MNC-Werten für PEA, die in der Tabelle I angegeben sind, hervor. Die MNC für das kombinatorische 6-mer-Gemisch ist beträchtlich größer als die für das natürliche 6-mer-Gemisch. Sie ist jedoch etwa um das Zweifache niedriger als die, die für den einzeln optimierten Satz erhalten wird, der definitionsgemäß einzeln synthetisiert werden muss. Es ist jedoch wichtig zu beachten, dass mehrere kombinatorische Synthesen durchgeführt werden können, so dass bei jeder getrennten Synthese definierte Teilsätze der X-mer erzeugt werden. Die Produkte aus den getrennten Synthesen werden anschließend zusammengemischt, um ein vollständiges Gemisch zu ergeben. Obwohl in einem gegebenen Synthesegemisch immer noch eine definierte positionsbezogene wechselseitige Abhängigkeit vorliegt, kann das gesamte kombinierte Gemisch eine größere Massenkomplexität aufweisen als jedes einzelne oben beschriebene kombinatorische Gemisch.

[0166] Bei einem weiteren Lösungsansatz können die massenmodifizierten Oligonukleotide einzeln synthetisiert werden, wie in dem ersten Verfahren beschrieben ist, worauf eine chemische Modifikation ihrer 5'-Endungen mit einer gewissen Art von Massenmarkenanteil folgt. Es ist lediglich eine geringe Anzahl von diskreten Massenmarken notwendig, um die Massen eines sich ergebenden natürlichen Oligonukleotidgemischs über die gesamte verwendbare Massenspektrometermassenbandbreite zu verteilen. Dieses Verfahren ähnelt dem in der U.S.-Patentschrift Nr. 5,605,798, deren relevante Offenbarung durch Bezugnahme in das vorliegende Dokument aufgenommen ist, Beschriebenen. Es ist zu beachten, dass die zuvor erwähnte Patentschrift zwar eine ähnliche Synthese beschreibt, jedoch nicht die Verwendung von massenmarkierten Oligonukleotiden für die in der vorliegenden Erfindung beschriebene Art von Massensignaturanalyse beschreibt oder nahe legt.

Auswirkungen von X-mer-Modifikationen auf massenspektroskopische, thermodynamische und enzymatische Eigenschaften

[0167] Die Zusammensetzung der X-mer-Vorläufer beeinflusst direkt die Gesamtspezifität und Empfindlichkeit der Untersuchung. Die Kontrolle sowohl über deren Entwurf als auch über deren Synthesemodus zu haben, ermöglicht außerdem die Einbindung von Modifikationen, die bei deren Verwendung bei den Verfahren der Erfindung helfen. Beispielsweise kann die Internukleosidbindung an der Phosphodiesterhauptkette der X-mer-Vorläufer modifiziert werden. Bei einem Ausführungsbeispiel ist bevorzugt, dass eine derartige chemische Modifikation die Phosphodiesterbindung widerstandsfähig gegenüber einem Nukleaseaufschluss macht. Geeignete Modifikationen umfassen ein Integrieren von nicht Brücken bildenden Thiophosphathauptketten, 5'-N-Phosphoamiditinternukleotidbindungen und dergleichen.

[0168] Die Massenmodifikation kann die thermodynamische Stabilität der Hybride, die zwischen dem X-mer-Vorläufer und dem Targetnukleinsäuresequenzanalyten gebildet werden, erhöhen, um die thermodynamische Stabilität der Hybride in dem Gemisch zu normieren. Beispielsweise bildet 2,6-Diaminopurin stabilere Basenpaare mit Thymidin, als Adenosin dies tut. Zusätzlich erhöht ein Integrieren von 2-Fluorthymidin die Stabilität von A-T-Basenpaaren, wohingegen ein Integrieren von 5-Brom- und 5-Methylcytidin die Stabilität von G-C-Basenpaaren erhöht.

[0169] Die Massenmodifikation kann die thermodynamische Stabilität der Hybride, die zwischen dem X-mer-Vorläufer und dem Targetnukleinsäuresequenzanalyten gebildet werden, verringern, um die thermodynamische Stabilität der Hybride in dem Gemisch zu normieren. A-T-Basenpaare können durch ein Integrieren von 2'-Aminonukleosiden destabilisiert werden. Statt Guanosin kann auch Inosin verwendet werden, um G-C-Basenpaare zu destabilisieren. Es hat sich gezeigt, dass die Integration von N-4-Ethyl-2'-desoxycytidin die Stabilität von G-C-Basenpaaren verringert. Ein Integrieren des Letztgenannten kann die Stabilität einer beliebigen gegebenen Duplexsequenz in einem Umfang normieren, bei dem deren Stabilität von dem A-T- und dem G-C-Gehalt unabhängig gemacht wird (Nguyen u. a., Nucleic Acids Res. 25, 3.095 (1997)).

[0170] Modifikationen, die die Fragmentierung des Oligonukleotids aufgrund der Ionisierungsprozesse bei der Massenspektrometrie verringern, können ebenfalls eingeführt werden. Beispielsweise ist ein Lösungsansatz eine 7-Deaza-Modifikation von Purinen, um die N-glycosidische Bindung zu stabilisieren und somit die Fragmentierung von Oligonukleotiden während des Ionisierungsprozesses zu verringern (siehe z. B. Schneider und Chait, Nucleic Acids Res. v23, 1.570 (1995)). Eine Modifikation der 2'-Position des Ribosering mit einer Elektronen ziehenden Gruppe wie z. B. Hydroxyl oder Fluor kann verwendet werden, um die Fragmentierung zu verringern, indem die N-glycosidische Bindung stabilisiert wird (siehe z. B. Tang u. a., J Am Soc Mass Spectrom, 8, 218 – 224, 1997).

Massenmarkierte Kettenabbruch-Nukleotide

[0171] Die Verwendung von Kettenabbruch-Nukleosid-Triphosphaten wie z. B. Dideoxynukleosid-Triphosphaten bei der vorliegenden Erfindung für das PEA-Verfahren unterscheidet sich grundlegend von der in der Technik bekannten. Das vorliegende PEA-Verfahren verwendet Kettenabbruch-Nukleotide als Mittel zum „Bewerten“ von Hybridisierungseignissen zwischen der Targetnukleinsäure und einer Vielzahl von massenmodifizierten X-meren, indem es die Masse der resultierenden Verlängerungs- bzw. Erweiterungs- bzw. Extensionsprodukte aus der Massenbandbreite der massenmodifizierten X-mer-Vorläufer heraus verschiebt. Diese spezifische Funktion gibt vor, dass die absolute Masse der Kettenabbruch-Nukleotide größer sein soll als die Massenbandbreite, die durch den leichtesten und den schwersten massenmodifizierten X-mer-Vorläufer in dem Gemisch definiert wird. Beispielsweise liegt die Massenbandbreite für ein X-mer-Vorläufer-Gemisch, das aus allen 6-X-meren zusammengesetzt ist, die aus den vier natürlichen Desoxynukleotiden hergestellt sind, zwischen 1.667 atomaren Masseneinheiten (amu) für (C₆) und 1.907 amu für (G₆). Dies ergibt eine Massen-

bandbreitendifferenz von 240 amu. Die Massen der einzelnen natürlichen Dideoxynukleotide (die Monophosphatform minus der Masse eines Wassermoleküls) betragen 296, 312, 272 bzw. 287 amu für pddA, pddG, pddC bzw. pddT. Da also die absolute Masse jedes Dideoxynukleotids größer ist als die Massenbandbreite für das natürliche 6-mer-Gemisch, sind sie ausreichend zum Partitionieren der Massen der X-mer-Vorläufer und der X+1-mer-Verlängerungsprodukte. Wenn jedoch die Massenbandbreite der X-mer-Vorläufer erhöht wird, beispielsweise durch die Einführung von Massenmodifikationen oder durch Verwendung von X-meren gemischter Längen, dann ist es wünschenswert, das Kettenabbruch-Nukleotid mit einer Massenmarke zu versehen, so dass die Massen aller Verlängerungsprodukte größer sind als die aller X-mer-Vorläufer.

[0172] Bei einem Ausführungsbeispiel der vorliegenden Erfindung können die massenmarkierten Dideoxynukleosid-Triphosphate auch eine zusätzliche chemische Komponente besitzen, die die Ionisierungseffizienz des gewünschten verlängerten X-mers relativ zu den nicht-verlängerten X-mer-Vorläufern oder jeglichen anderen unerwünschten Komponenten, die in dem Probengemisch vorliegen, erhöht. Üblicherweise wird die Ionisierungseffizienz zumindest um den Faktor 2 erhöht, noch üblicher um den Faktor 4 und vorzugsweise um den Faktor 10. Dort, wo also beispielsweise 6-mere als X-mer-Vorläufer verwendet werden, unterstützt die obige zusätzliche Komponente das Ermöglichen der Analyse der 6-mer-Vorläufer und der 7-mer-Verlängerungsprodukte. Exemplarisch für derartige zusätzliche chemische Komponenten sind primäre Amine, die als Protonierungsstellen fungieren und somit einzelne positive Ionenspezies für eine MALDI-Analyse unterstützen können (Tang u. a., 1997, *supra*). Es ist auch möglich, quaternäre Amine zu integrieren, die eine feststehende positive Ladung aufweisen. Diese Klasse von chemischen Gruppen kann unter Verwendung einer NHS-Esterchemie, die der bei Gut u. a. in der WO 96/27681 offenbarten ähnelt, in nicht-spaltbare Massenmarken integriert werden. Kurz gesagt wird der Succinimidester einer mit quaternärem Ammonium geladenen Spezies, z. B. Trimethylammoniumhexyryl-N-hydroxysuccinimidylester, mit einem Nukleosiddervat, das eine primäre aliphatische Aminogruppe aufweist, zur Reaktion gebracht. Ein geeignetes Nukleosid ist beispielsweise ein Terminator wie z. B. die 3'-Aminoderivate der 2'-Desoxynukleoside. Andere geeignete Nukleoside wären die 5-[3-Amino-1-propynyl]-pyrimidin- und 7-Deaza-[3-amino-1-propynyl]-purin-Derivate, die ähnlich denjenigen sind, die verwendet werden, um die fluoreszierend markierten ddNTPs zu erzeugen, die von Prober u. a. (Science, 238, 336 (1987)) beschrieben sind.

Verfahren der Erfindung

Erzeugen von Kurzwortinhalt-Darstellungen von Targetnukleinsäuren

[0173] Die Erfindung ist auf Verfahren und Reagenzien zum Rekapitulieren einer Targetnukleinsäure in Form eines Satzes von Oligonukleotiden (X-meren), die zu der Targetsequenz komplementär sind, und zum Analysieren des Satzes mittels Massenspektroskopie gerichtet. Der Satz von Oligonukleotiden stellt den „Kurzwort“-Inhalt des Targets dar, der definierte Sequenzinformationen über das Target liefert. Der Satz von Oligonukleotiden, die ein Target darstellen, kann einer von drei allgemeinen Typen sein (**Fig. 1**). Der verschachtelte Satz von überlappenden X-meren (**Fig. 1a**) ist dadurch gekennzeichnet, dass er eine umfassende Überlappung zwischen den X-meren in dem Satz aufweist. Der verschachtelte Satz von halb-überlappenden X-meren (**Fig. 1b**) weist eine geringere Überlappung unter den X-meren auf, wohingegen der nicht-überlappende Satz von X-meren (**Fig. 1c**) keine Überlappung aufweist. Bei allen drei Arten von Sätzen muss die X-mer-Länge in einem gegebenen Satz nicht konstant sein. Allgemein weisen die X-mere in dem verschachtelten Satz von überlappenden X-meren eine Länge von etwa 3 bis etwa 18, üblicherweise etwa 5 bis etwa 14, Nukleotiden auf. Für diesen Satz umfasst die Überlappung entlang der gesamten Länge der Targetnukleinsäuresequenz alle Nukleotide bis auf eines. Allgemein weisen die X-mere in dem verschachtelten Satz von halb-überlappenden X-meren eine Länge von etwa 3 bis etwa 18, üblicherweise etwa 5 bis etwa 14, Nukleotiden auf. Allgemein weisen die X-mere in dem verschachtelten Satz von nicht-überlappenden X-meren eine Länge von etwa 3 bis etwa 18, üblicherweise etwa 4 bis etwa 14, Nukleotiden auf. Bei allen drei Lösungsansätzen erfassen die X-mere die gesamte Länge der Targetnukleotidsequenz. Die tatsächliche Anzahl von erzeugten X-meren wird allgemein durch die Länge der Targetnukleotidsequenz und das gewünschte Ergebnis bestimmt. Die Anzahl von X-meren sollte ausreichend sein, um die Ziele der definierten Anwendung zu erzielen. Wenn beispielsweise das Ziel darin besteht, eine Mutationserfassung durchzuführen, dann werden ausreichend viele X-mere benötigt, um das X-mer oder den Satz von X-meren, die die Mutation umfassen, zu unterscheiden.

Allgemeine Beschreibung der Verfahren

[0174] Ein Aspekt der vorliegenden Erfindung ist ein Verfahren zum Analysieren einer Targetnukleinsäuresequenz. Ein Gemisch von X-mer-Vorläufern wird zu den Targetnukleinsäuresequenzen hybridisiert. Das Gemisch umfasst natürliche und massenmodifizierte X-mer-Vorläufer, die eine minimale Länge von drei Nukleo-

tiden aufweisen. Das Gemisch weist eine Gemischbedeckungskomplexität von etwa 7/8 auf, wenn das Gemisch zumindest 56 gesonderte X-mer-Vorläufer enthält. Mit zunehmender durchschnittlicher Länge des X-mer-Vorläufers nimmt auch die Anzahl von gesonderten X-meren in den Gemischen dieser Erfindung zu, und die Gemischbedeckungskomplexität mag abnehmen. Die Untergrenze der Gemischbedeckungskomplexität ist gleich 56 geteilt durch die Anzahl von X-meren in dem Gemisch. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Die Massenzahlkomplexität (MNC) des Gemisches kann größer sein als die jedes natürlichen Äquivalentgemisches. Die Hybride werden verarbeitet, um die Masse der X-mer-Vorläufer-Abschnitte der Hybride in einer mittels einer Targetsequenz durchgeföhrten Reaktion zu verändern, und das Produkt wird anhand der Massenspektrometrie analysiert. Die ersten zwei Schritte des Verfahrens können in Lösung oder mit oberflächengebundenen Nukleinsäuren wie z. B. in einem Array durchgeföhr werden. Lösungsbasierte Systeme können bevorzugt sein, da sie durch standardmäßige Lösungsmassewirkungs- und -diffusionsprozesse geregelt sind.

Herstellung des X-mer-Gemisches:

[0175] Der erste Schritt des Verfahrens der Erfindung ist das Herstellen eines Gemisches aus natürlichen und massenmodifizierten X-mer-Vorläufern, die eine entsprechende Massenzahl und Bedeckungskomplexität für die gegebene Anwendung aufweisen. Das X-mer-Vorläufer-Gemisch kann auch die hierin beschriebenen Attribute bezüglich Ionisierung und thermodynamischer Eigenschaften besitzen. Der Entwurf und die Herstellung des X-mer-Vorläufer-Gemischs kann wie hierin beschrieben durchgeföhr werden.

Verarbeitungsschritt

[0176] Der zweite Schritt des Verfahrens der Erfindung besteht darin, die Hybride zu verarbeiten, um die Masse der X-mer-Vorläufer-Abschnitte der Hybride gemäß der Beschreibung hierin zu verändern. Diese Veränderung kann entweder mittels einer enzymatischen oder einer chemischen Reaktion bewerkstelligt werden. Geeignete enzymatische Techniken umfassen eine Polymeraseverlängerungsuntersuchung, eine Ligaseuntersuchung und dergleichen. Geeignete chemische Techniken umfassen eine Kondensation aktiver X-mer-Vorläufer unter Verwendung von Carbodiimiden und Cyanbromidderivaten und dergleichen. Die folgende Erörterung ist eine kurze Beschreibung mancher der verschiedenen Prozesse; eine ausführlichere Erläuterung wird nachstehend dargelegt.

Polymeraseverlängerungsuntersuchung

[0177] Für die Polymeraseverlängerungsuntersuchung (PEA) werden die hybridisierten X-mer-Vorläufer verlängert, indem unter Verwendung einer Nukleotidpolymerase ein einzelnes Nukleotid an dem 3'-Ende der hybridisierten X-mer-Vorläufer polymerisiert wird (siehe [Fig. 2](#)). Für die Polymeraseverlängerungs- und -spaltungsuntersuchung (PEACA), eine Form der PEA, werden die hybridisierten X-mer-Vorläufer zuerst verlängert, indem ein oder mehr Nukleotide an dem 3'-Ende der hybridisierten X-mer-Vorläufer unter Verwendung einer Nukleotidpolymerase polymerisiert werden und durch Aufschließen des Produktes mit einer 5'- bis 3'-Exonuklease, – Endonuklease oder einem chemischen Reagensmittel, je nach den Besonderheiten des durchgeföhrten Verfahrens, erneut verkürzt werden (siehe [Fig. 3](#) und [Fig. 4](#)).

Ligaseuntersuchung

[0178] Für die X-mer-Ligationsuntersuchung (XLA) werden benachbarte hybridisierte X-mer-Vorläufer vor einer Analyse unter Verwendung einer Ligase miteinander ligiert (siehe [Fig. 5](#)). Es ist bevorzugt, dass die X-mer-Vorläufer eine ausreichende Länge aufweisen, um als gute Substrate zur Ligation durch die Ligase dienen, jedoch nicht zu lang sind, um als Schablonen für eine Ligation von komplementären X-mer-Vorläufern in dem Reaktionsgemisch zu dienen. Man sollte beachten, dass, obwohl es zwar vorzuziehen ist, dass alle benachbarten hybridisierten X-mer-Vorläufer ligiert sind, dies kein Erfordernis ist.

[0179] Die Ligationsuntersuchung kann mit oberflächengebundenen Arrays durchgeföhr werden (siehe [Fig. 6](#)). Die Arrays weisen eine Oberfläche und eine Vielzahl von an dieselbe angelagerten Oligonukleotidsonden auf. Die Sonden enthalten:

- (a) einen an die Oberfläche angelagerten spaltbaren Linker; und
- (b) eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist.

[0180] Das Verfahren umfasst die folgenden Schritte:

- (1) Hybridisieren der Targetnukleinsäuresequenz zu den Sonden;
- (2) Hinzufügen des Gemisches aus X-mer-Vorläufern zu der Targetnukleinsäuresequenz;
- (3) Ligieren der hybridisierten X-mer-Vorläufer, die zu dem endständigen 5'-Phosphat benachbart sind, mit der oberflächengebundenen Sonde, um einen Komplex aus hybridisiertem Vorläufer/Sonde zu bilden, wobei die Targetnukleinsäuresequenz daran angelagert ist;
- (4) Spalten des Komplexes an dem spaltbaren Linker; und
- (5) Analysieren des Komplexes in jedem Sondenmerkmal oder Satz von Merkmalen mittels Massenspektrometrie.

Ausführliche Beschreibung der Verfahren

[0181] Die folgende Beschreibung ist auf drei allgemeine Verfahren zum Erzeugen von Oligonukleotidsätzen gerichtet, die den Kurzwortinhalt des Targets darstellen. Jedes Verfahren kann je nach den verwendeten Reagenzien einen oder mehr Typen von Oligonukleotidsätzen erzeugen. Diese Beschreibung erfolgt mittels Veranschaulichung und stellt keine Einschränkung dar. Wie oben erwähnt wurde, wird das erste Verfahren als „Polymeraseverlängerungsuntersuchung“ (PEA) bezeichnet, das zweite wird als Polymeraseverlängerungs- und -spaltungsuntersuchung (PEACA) bezeichnet, und das dritte wird als „X-mer-Ligationsuntersuchung“ (XLA) bezeichnet.

[0182] Alle Verfahren beruhen auf Oligonukleotidgemischen (X-mer-Gemischen), die aus natürlichen und/oder massenmodifizierten Nukleotiden zusammengesetzt sind. Man sollte verstehen, dass die unterschiedlichen Sätze von Gemischen entworfen sein können, um die unterschiedlichen Typen von Sätzen zu erzeugen und somit verschiedene Mengen an Targetsequenzinformationen zu liefern. Durch eine Analyse der Massenspitzen, die in den bei den obigen Verfahren erzeugten Massenspektren vorliegen, und mittels einer Korrelation dieser Spitzen mit Informationen über die X-mer-Vorläufer in dem Gemisch, die für jedes Massenspektrum verantwortlich sind, und möglicherweise mit Vorab-Informationen über die Targetsequenz werden die aus dem Target gewünschten Informationen ermittelt.

[0183] Die PEA ist ein generisches Verfahren zum Erzeugen von verschachtelten Sätzen von überlappenden und halbüberlappenden X-meren. Es gibt drei grundlegende Schritte bei diesem Verfahren ([Fig. 2](#)). Bei Schritt 1 darf ein Gemisch (oder ein Satz von Gemischen) von X-meren, die entweder alle möglichen X-mer-Sequenzen oder Teilsätze derselben darstellen, an zufälligen Positionen entlang der Targetnukleinsäuresequenz gemäß Watson-Crick-Basenpaarungsregeln hybridisieren. Bei Schritt 2 werden die hybridisierten X-mera unter Verwendung einer Nukleotidpolymerase wie z. B. einer DNA- oder RNA-abhängigen DNA-Polymerase und eines Gemischs aus einem oder mehr Kettenabbruch-Nukleosid-Triphosphaten wie z. B. Dideoxynukleotid-Triphosphaten (MT-ddNTPs) um ein einziges Nukleotid verlängert. Bei Schritt 3 werden die resultierenden verlängerten X-mera, d. h. X+1-mer-Verlängerungsprodukte, mittels Massenspektroskopie analysiert.

[0184] Das Ausmaß der Überlappung zwischen den X+1-mer-Produkten hängt von der Sequenzvollständigkeit des abfragenden X-mer-Gemisches ab. Wenn beispielsweise alle 4.096 6-mera und alle vier ddNTPs in dem abfragenden Gemisch vorliegen, so ist die maximale Überlappung zwischen den resultierenden 7-mer-Produkten möglich. Das Bereitstellen eines Teilsatzes der 4.096 möglichen 6-mera und/oder eines Teilsatzes der vier ddNTPs führt zu einer geringeren Überlappung zwischen den 7-mer-Produkten und potentiellen Zwischenräumen in der Sequenzbedeckung.

[0185] Es ist auch wichtig, dass die Kettenabbruch-Nukleotide eine ausreichende Masse aufweisen, um das X-mer-Vorläufer-Gemisch und X+1-mer-Verlängerungsprodukte auf effektive Weise zu partitionieren. Es sollte beachtet werden, dass, obwohl es vorzuziehen ist, dass alle der hybridisierten X-mer-Vorläufer verlängert sind, dies kein Erfordernis darstellt. Bei der vorliegenden Erfindung ist die Bestimmung um so genauer, je größer die Anzahl von verlängerten hybridisierten X-mer-Vorläufern ist.

[0186] Die Kombination von Reagenzien ist Bedingungen unterworfen, unter denen die X-mera zu der Targetnukleinsäure hybridisieren und in der Gegenwart eines Kettenabbruch-Nukleosid-Triphosphats, das zu einem Nukleotid des Targets, das neben dem hybridisierten X-mer liegt, komplementär ist, um ein Nukleotid verlängert werden. Allgemein wird ein wässriges Medium verwendet. Es können auch andere polare Hilfslösungsmittel verwendet werden, üblicherweise mit Sauerstoff angereicherte organische Lösungsmittel von 1 – 6, üblicherweise von 1 – 4, Kohlenstoffatomen, einschließlich Alkoholen, Ethern und dergleichen. Üblicherweise liegen diese Hilfslösungsmittel, falls sie verwendet werden, zu weniger als etwa 70 Gewichtsprozent, noch üblicher zu weniger als etwa 30 Gewichtsprozent, vor.

[0187] Der pH-Wert für das Medium liegt üblicherweise im Bereich von etwa 4,5 bis 9,5, noch üblicher im Bereich von etwa 5,5 bis 8,5, und vorzugsweise im Bereich von etwa 6 bis 8. Es können verschiedene Puffer verwendet werden, um den gewünschten pH-Wert zu erreichen und um den pH-Wert während der Bestimmung aufrechtzuerhalten. Veranschaulichende Puffer umfassen Borat, Phosphat, Carbonat, Tris, Barbital und der gleichen. Der jeweilige verwendete Puffer ist bei dieser Erfindung nicht kritisch, bei einzelnen Verfahren kann jedoch ein Puffer gegenüber einem anderen bevorzugt sein.

[0188] Die Reaktion wird ausreichend lange durchgeführt, um die verlängerten X+1-mere zu erzeugen, die ein Kettenabbruch-Nukleosid-Triphosphat enthalten. Allgemein beträgt die Zeitdauer zum Durchführen des gesamten Verfahrens zwischen etwa 10 und 200 Minuten. Üblicherweise ist es wünschenswert, den Zeitraum zu minimieren.

[0189] Die Konzentration der Nukleotidpolymerase wird üblicherweise empirisch bestimmt. Vorzugsweise wird eine Konzentration verwendet, die ausreichend ist, um die meisten, wenn nicht alle, der Vorläufer-X-mere, die spezifisch zu der Targetnukleinsäure hybridisieren, zu verlängern (siehe unten). Die primären Begrenzungsfaktoren sind allgemein die Reaktionszeit und die Kosten des Reagens.

[0190] Die Anzahl der Targetnukleinsäuremoleküle kann lediglich 10^6 in einer Probe betragen, kann jedoch allgemein zwischen etwa 10^6 und 10^{13} , noch üblicher zwischen etwa 10^8 und 10^{12} Molekülen in einer Probe, vorzugsweise zumindest 10^{-13} M in der Probe, variieren und kann 10^{-13} bis 10^{-6} M, noch üblicher 10^{-11} bis 10^{-7} M, betragen. Allgemein sind die Reagenzien für die Reaktion in Mengen vorgesehen, um eine Verlängerung der hybridisierten X-mere zu erzielen. Die Anzahl jedes X-mer-Vorläufer-Moleküls beträgt allgemein 10^{10} und üblicherweise etwa 10^{10} bis etwa 10^{13} , vorzugsweise etwa 10^{11} bis etwa 10^{12} für eine Probengröße, die etwa 10 Mikroliter beträgt. Die Konzentration jedes X-mer-Vorläufers kann gemäß seiner Thermostabilität angepasst werden, wie oben erörtert wurde. Das absolute Verhältnis von Target zu X-mer-Vorläufer soll empirisch bestimmt werden. Die Konzentration der Kettenabbruch-Nukleosid-Triphosphate in dem Medium kann je nach der Affinität der Nukleosid-Triphosphate für die Polymerase variieren. Vorzugsweise liegen diese Reagenzien in einer überschüssigen Menge vor. Die Nukleosid-Triphosphate liegen üblicherweise zu etwa 10^{-7} bis etwa 10^{-4} M, vorzugsweise etwa 10^{-6} bis 10^{-5} M, vor.

[0191] Je nach der verwendeten Art der Polymerase, den Konzentrationen von Target und X-meren und den thermodynamischen Eigenschaften der X-mere in dem Gemisch kann die Reaktionstemperatur im Bereich zwischen etwa 0°C und etwa 95°C liegen. Beispielsweise können bei 40 nM Targetnukleinsäuresequenz, 40 nM 6-mer und 7 nM Bst Polymerase zwischen 20 % und 50 % des 6-mers in zwei Stunden bei 5°C verlängert werden, je nach der Sequenz des 6-mers. Ähnliche Verlängerungseffizienzen werden bei 20°C erreicht, was darauf hinweist, dass die Verlängerungseffizienz nicht allein von der Thermodynamik der Wechselwirkung zwischen X-mer und Target abhängig ist. Wichtig ist, dass es nützlich sein kann, die Inkubationstemperatur zyklisch zu durchlaufen. Ein zirkuläres Durchlaufen könnte helfen, eine strukturierte Region des Targets für eine X-mer-Bindung und anschließende Verlängerung freizulegen sowie den Stoffumsatz der Verlängerungsprodukte zu erleichtern. Somit könnte die Gesamtempfindlichkeit der PEA deutlich erhöht werden, indem ermöglicht wird, dass ein gegebenes Targetmolekül als Schablone für mehrere X-mer-Bindungs- und anschließende Verlängerungsreaktionen agiert. Gemäß diesem Aspekt der Erfindung kann ein Zyklus etwa 0,1 bis 5 Minuten lang, noch üblicher etwa 0,5 bis 2 Minuten, bei einer Temperatur von etwa 75°C bis etwa 95°C durchgeführt werden, und ein weiterer Zyklus kann etwa 1 bis 20 Minuten lang, noch üblicher etwa 5 bis 15 Minuten lang, bei einer Temperatur von etwa 5°C bis etwa 45°C ausgeführt werden. Die Anzahl der Zyklen kann etwa 2 bis etwa 20 oder mehr betragen. Allgemein sind die Zyklustemperaturen und die Zyklusdauer ausgewählt, um eine Optimierung der Verlängerung des hybridisierten X-mers einer gegebenen Länge zu liefern.

[0192] Die Reihenfolge des Kombinierens der verschiedenen Reagenzien, um die Kombination zu bilden, kann variieren. Üblicherweise wird die das Targetpolynukleotid enthaltende Probe mit einer vorab hergestellten Kombination von Kettenabbruch-Nukleosid-Triphosphaten und Nukleotidpolymerase kombiniert. Die X-mere können in der hergestellten Kombination enthalten sein oder können später hinzugegeben werden. Jedoch können eine gleichzeitige Hinzufügung aller obigen Stoffe sowie andere schrittweise oder nacheinander erfolgende Hinzufügungsreihenfolgen eingesetzt werden, vorausgesetzt, dass alle oben beschriebenen Reagenzien vor dem Beginn der Reaktionen kombiniert werden.

[0193] Die PEACA ist ein weiteres generisches Verfahren zum Erzeugen von verschachtelten Sätzen von überlappenden und halbüberlappenden X-meren. Bei diesem Verfahren gibt es vier grundlegende Schritte ([Fig. 3](#)). Bei Schritt 1 darf ein Gemisch von X-meren, die entweder alle möglichen X-mer-Sequenzen oder Teilsätze derselben darstellen, an zufälligen Positionen entlang der Targetnukleinsäuresequenz gemäß Wat-

son-Crick-Basenpaarungsregeln hybridisieren. Bei Schritt 2 werden die hybridisierten X-mer unter Verwendung einer Nukleotidpolymerase und eines Gemisches von einem oder mehr natürlichen Nukleotid-Triphosphaten (dNTPs) und natürlichen oder massenmarkierten Dideoxynukleotid-Triphosphaten (ddNTPs) verlängert. Bei Schritt 3 wird ein Abschnitt des 5'-Terminus der X-mer-Verlängerungsprodukte mittels enzymatischer oder chemischer Spaltung beseitigt. Bei Schritt 4 werden die resultierenden X-mer-Verlängerungsprodukte mittels Massenspektroskopie analysiert.

[0194] Die PEACA ist dazu entworfen, X-mer-Verlängerungsprodukte zu erzeugen, wo die Produktlänge durch die Targetsequenz und die Identität der dNTPs und Kettenabbruch-Nukleotide (d. h. ddNTPs), die in dem Reaktionsgemisch vorliegen, definiert wird. Das Ausmaß der potentiellen Überlappung zwischen den X-mer-Produkten hängt nicht nur von der Länge ab, sondern auch von der Bedeckungskomplexität des X-mer-Vorläufer-Gemisches. Eine Verwendung von X-mer-Vorläufer-Gemischen, die eine höhere Bedeckungskomplexität aufweisen, führt zu einer größeren Überlappung zwischen den Verlängerungsprodukten. Wenn alle vier dNTPs und ein einzelnes Kettenabbruch-Nukleotid (d. h. ddNTP) in dem Reaktionsgemisch vorliegen, so erhöht ein Erhöhen des dNTP/ddNTP-Molarverhältnisses nicht nur die Länge der Verlängerungsprodukte, sondern führt auch zu einer größeren potentiellen Überlappung zwischen den Verlängerungsprodukten. Umgekehrt ergeben kleinere dNTP/ddNTP-Verhältnisse kürzere Verlängerungsprodukte und eine geringere potentielle Überlappung.

[0195] Bei der PEACA ([Fig. 3](#)) wird eine Spaltung eines Teils des 5'-Endes durch ein spezifiziertes Nukleotid in dem X-mer-Vorläufer definiert. Die PEACA weist den zusätzlichen Vorteil auf, dass sie die Beseitigung eines bestimmten Teils des 5'-Endes der X-mer-Produkte unter Verwendung eines enzymatischen oder chemischen Prozesses ermöglicht. Eine derartige Eigenschaft könnte genutzt werden, um PEACA-Produkte zu erzeugen, die die gleiche Länge aufweisen, wenn mit X-mer-Gemischen gemischter Längen begonnen wird. Dies könnte ermöglichen, dass die an A/T reicherden X-mer-Vorläufer länger als diejenigen sind, die reich an C/G sind (um Unterschiede bezüglich der Thermostabilität zu kompensieren), auf eine Spaltung hin jedoch PEACA-Produkte mit identischen oder ähnlichen Längen erzeugen. Überdies ist der 5'-Terminus des X-mers einer Fehl-Hybridisierung mit dem Target ausgesetzt und weniger anfällig für eine Prüfung durch die Polymerase. Somit könnte die PEACA-Spaltungsreaktion verwendet werden, um fehlerhafte Informationen vor einer Massenanalyse zu beseitigen.

[0196] Bei der zweiten Version der PEACA (PEACA II, [Fig. 4](#)) wird eine Spaltung durch ein festgelegtes Nukleotid definiert, das während des Verlängerungsprozesses integriert wird. Bei dieser Version wird kein Teil des X-mer-Vorläufers in dem abschließenden X-mer-Produkt zurückgehalten. Zusätzlich dazu, dass sie davon abhängig ist, welche der dNTPs und ddNTPs in dem Reaktionsgemisch vorliegen, kann die Produktlänge auch durch Variieren des molaren Verhältnisses von dNTP/ddNTP definiert werden, wenn beide Formen eines gegebenen Nukleotids in einem Reaktionsgemisch vorliegen. Beispielsweise ergeben Werte eines größeren dNTP/ddNTP-Verhältnisses im Durchschnitt längere Verlängerungsprodukte. Umgekehrt ergeben kleinere dNTP/ddNTP-Verhältnisse kürzere Verlängerungsprodukte.

[0197] Die Bedingungen zum Durchführen der PEACA-Verlängerungsreaktionen sind ähnlich denen, die oben für die PEA beschrieben wurden. Der Puffer-pH-Wert, die Ionenstärke, die Hinzufügung von Tensiden, die Temperatur (und ein zirkuläres Durchlaufen derselben), die Polymerasekonzentration, die X-mer-Konzentration, die Targetkonzentration und die dNTP/ddNTP-Konzentrationen und -Verhältnisse werden allesamt optimiert, um eine maximale Spezifität, Verlängerungseffizienz und einen maximalen Informationsgehalt zu erreichen.

[0198] Die folgenden Beispiele des Bewirkens einer Spaltung eines bestimmten Teils des 5'-Endes werden als Nächstes veranschaulichungshalber und nicht einschränkungshalber beschrieben. Bei einem Lösungssatz kann ein eine Spaltung definierendes Nukleotid verwendet werden. Das eine Spaltung definierende Nukleotid kann ein Ribonukleotid sein, das anfällig für eine Spaltung mittels einer Endoribonuklease, z. B. Ribonuklease A, oder durch eine chemische Base wie z. B. Ammonium oder Hydroxid und dergleichen, ist. Das die Spaltung definierende Nukleotid kann ein Nukleotid sein, das eine 5'-N-Phosphoamidit-Internukleotidbindung bildet, die mittels einer Säure, z. B. Trichloressigsäure oder verdünnte HCl, gespalten werden kann. Wenn das X-mer ausschließlich aus Ribonukleotiden besteht, kann eine Desoxyribonuklease verwendet werden, z. B. DNaseI.

[0199] Der Spaltungspunkt kann auch durch ein Nukleotid oder einen Satz von Nukleotiden definiert sein, die eine Spaltung mittels eines definierten Reagens oder Enzyms blockieren. Beispielsweise kann das X-mer aus Triphosphatbindungen an dem 3'-Terminus und natürlichen Phosphodiesterbindungen an dem 5'-Terminus be-

stehen. Eine Spaltung des X-mers mit einer 5'- bis 3'-Exonuklease, die empfindlich gegenüber Thiophosphaten ist, z. B. T7-Gen-6-Protein, verschlechtert das X-mer bis zu dem Punkt der Triphosphatbindung.

[0200] Die Bedingungen zum Durchführen der Spaltung sind diejenigen, die in der Technik für die obigen Enzyme allgemein bekannt sind. Kurz gesagt sind diese Bedingungen eine Inkubation des Enzyms in einem entsprechenden Puffer, der ein divalentes Metallion enthalten kann (falls erforderlich). Geeignete 5'- bis 3'-Exonukleasen, d. h. diejenigen Enzyme, die Nukleotide jeweils immer einzeln vom Ende eines Polynukleotids von dem 5'- zu dem 3'-Ende spalten, umfassen z. B. DNA-Polymerase und T7-Gen 6. Geeignete Endonukleasen, d. h. diejenigen Enzyme, die Bindungen in einer Nukleinsäure spalten, umfassen z. B. Desoxyribonuklease I, Ribonuklease A oder dergleichen. Geeignete chemische Reaktionen für einen chemischen Aufschluss umfassen z. B. diejenigen Reagenzien zum Durchführen einer basen- oder säurekatalysierten Hydrolyse von Phosphatestern und dergleichen wie z. B. Chlorwasserstoffsäure, Trifluoressigsäure und dergleichen. Die Bedingungen zum Durchführen der obigen enzymatischen oder chemischen Reaktionen sind Fachleuten hinreichend bekannt und werden hier nicht wiederholt.

[0201] XLA ist ein weiteres generisches Verfahren zum Erzeugen verschachtelter Sätze von überlappenden und halbüberlappenden X-meren. Bei diesem Verfahren gibt es drei grundlegende Schritte ([Fig. 5](#)). Bei Schritt 1 darf ein Gemisch von X-meren, die entweder alle möglichen X-mer-Sequenzen oder Teilsätze derselben darstellen, an zufälligen Positionen entlang der Targetnukleinsäuresequenz gemäß Watson-Crick-Basenpaarungsregeln hybridisieren. Bei Schritt 2 werden die X-meren, die nebeneinander hybridisieren, unter Verwendung einer Ligase, z. B. einer DNA-Ligase, die die Bildung einer Phosphodiesterbindung unterstützt, um zwei benachbarte Basen in separaten Oligonukleotiden zu verbinden, enzymatisch miteinander ligiert. Derartige Ligasen umfassen z. B. T4-DNA-Ligase, Taq-DNA-Ligase, E.coli-DNA-Ligase und dergleichen. Alternativ dazu können benachbarte X-mer-Vorläufer unter Verwendung eines Kondensationsmittels chemisch ligiert werden. Geeignete Kondensationsmittel umfassen z. B. Carbodiimide, Cyanbromidderivate und dergleichen. Bei Schritt 3 werden die resultierenden ligierten nX-mer-Produkte mittels Massenspektroskopie analysiert.

[0202] Das Ausmaß der Überlappung zwischen den nX-mer-Produkten hängt von der Sequenzvollständigkeit des abfragenden X-mer-Gemisches ab. Wenn beispielsweise alle 4.096 6-mere in dem abfragenden Gemisch vorliegen, so ist die maximale Überlappung zwischen den resultierenden nX-mer-Produkten möglich. Das Bereitstellen eines Teilsatzes der 4.096 möglichen 6-mere führt zu einer geringeren Überlappung zwischen den nX-mer-Produkten.

[0203] Die Bedingungen zum Durchführen der Reaktionen sind bei diesem Lösungsansatz ähnlich den oben beschriebenen. Der pH-Wert für das Medium liegt üblicherweise im Bereich zwischen etwa 4,5 und 9,5, noch üblicher im Bereich zwischen etwa 5,5 und 8,5 und vorzugsweise im Bereich von etwa 6 bis 8.

[0204] Die Reaktion wird ausreichend lange durchgeführt, um das gewünschte ligierte Produkt zu erzeugen. Allgemein beträgt der Zeitraum zum Durchführen des gesamten Verfahrens zwischen etwa 10 und 200 Minuten. Es ist üblicherweise wünschenswert, den Zeitraum zu minimieren.

[0205] Je nach der Art der verwendeten Ligase, den Konzentrationen von Target und X-meren und den thermodynamischen Eigenschaften der X-meren in dem Gemisch kann die Reaktionstemperatur von 0°C bis 95°C variieren. Wie im Fall von PEA und PEACA kann es vorteilhaft sein, die Inkubationstemperatur zyklisch zu durchlaufen, um die strukturierte Region des Targets für eine X-mer-Bindung und anschließende Ligation freilegen zu helfen sowie den Stoffumsatz der ligierten Produkte zu erleichtern.

[0206] Die Konzentration der Ligase wird üblicherweise empirisch bestimmt. Vorzugsweise wird eine Konzentration verwendet, die ausreichend ist, um die meisten, wenn nicht alle, Vorläufer-X-meren, die spezifisch zu der Targetnukleotidsequenz hybridisieren, zu ligieren. Die wichtigsten einschränkenden Faktoren sind allgemein die Reaktionszeit und die Kosten des Reagens.

[0207] Die Konzentration jedes X-mer-Vorläufers ist allgemein so, wie dies oben für die PEA beschrieben wurde, und kann gemäß seiner Thermostabilität angepasst werden, wie oben erörtert wurde. Das absolute Verhältnis von Target zu X-mer-Vorläufer soll empirisch bestimmt werden.

[0208] Der Umfang der Phosphorylierung des 5'-Terminus des X-mer-Gemisches kann das Ausmaß der Ligation (Gesamtzahl ligierter Produkte) und die Länge von Ligationsprodukten (Wert von n) beeinflussen. Das Ausmaß und die Länge der Ligation kann auch dadurch gesteuert werden, dass eine Modifizierung an dem 3'-Terminus des X-mer-Gemisches eingeführt wird, die die Ligation blockiert. Bei einem Lösungsansatz werden

zwei Sätze von X-mer-Gemischen zusammen in einem einzigen Ligationsreaktionsgemisch verwendet. Die X-mere in dem ersten X-mer-Gemisch besitzen einen 5'-phosphorylierten Terminus und einen 3'-blockierten Terminus (p--y), wohingegen die X-mere in dem zweiten X-mer-Gemisch sowohl 5'- als auch 3'-Hydroxyltermini (o--o) aufweisen. Dies führt dazu, dass lediglich 2X-mer-Ligationsprodukte die Form o--o/p--y aufweisen. Ein Blockieren des 3'-Terminus kann beispielsweise dadurch bewerkstelligt werden, dass eine Gruppe verwendet wird, die keiner Kondensation unterzogen werden kann, z. B. eine unnatürliche Gruppe wie z. B. 3'-Phosphat, ein 3'-endständiges Dideoxy, ein Polymer oder eine Oberfläche oder ein anderes Mittel zum Hemmen einer Ligation. Dieser Lösungsansatz weist große informationstechnische Vorteile auf, da die zwei Sätze gemeinsam optimiert werden können.

[0209] Die Verwendung von modifizierten X-meren, um den Umfang und die Länge der Ligation zu steuern, kann mit der Integration einer Ionisierungsmarke zum Zweck einer erhöhten Ionisierungseffizienz kombiniert werden. Die X-mere in dem ersten X-mer-Gemisch besitzen einen 5'-phosphorylierten Terminus und einen 3'-blockierten und -markierten Terminus (p--z). Die X-mere in dem zweiten X-mer-Gemisch weisen sowohl 5'- als auch 3'-Hydroxyltermini auf (o--o). Dies führt dazu, dass lediglich 2X-mer-Ligationsprodukte die Form o--o/p--z aufweisen. Die durch z dargestellte Gruppe kann aus einer einzelnen Funktionalität bestehen, die den doppelten Zweck sowohl des Blockierens als auch des Ionisierungsmarkierens erfüllt, z. B. eine Quaternäres-Ammonium-Gruppe, die über das 3'-Hydroxyl verbunden ist, oder sie kann getrennte Funktionalitäten darstellen, z. B. ein 3'-endständiges Dideoxy zum Blockieren mit einer separaten Ionisierungsmarke, die an die Nukleosidbase angelagert ist.

[0210] PEA, PEACA und XLA besitzen eine Anzahl wünschenswerter Attribute. Erstens sind sie alle lösungsbasierte Systeme und werden durch standardmäßige Lösungsmassenwirkungs- und Diffusionsprozesse geregelt. Dies steht im Gegensatz zu keine Unterstützung erfahrenden oberflächenbasierten Arrayhybridisierungssystemen, wo die Sonde physisch an der Oberfläche angebracht ist und nicht in der Lage ist, zu diffundieren, wodurch die Kinetik der Hybridisierung verlangsamt wird. Im Gegensatz zu oberflächengebundenen Arrays besteht ein Kennzeichen der vorliegenden Erfindung darin, dass eine hohe Vielzahl von Oligonukleotiden entlang der Targetsequenz bindet. Dies erhöht wahrscheinlich die Gesamteffizienz der X-mer-Bindungs- und der anschließenden enzymatischen Reaktion. Da die X-mer-Vorläufer kurz sind, ist es außerdem weniger wahrscheinlich, dass sie intramolekulare Strukturen bilden.

[0211] Zweitens nutzen PEA, PEACA und XLA hochspezifische enzymatische Prozesse. Im Fall von PEA und PEACA dient der hohe Grad an Spezifität der Polymerase für perfekte Duplexe im Wesentlichen dazu, den Hybridisierungsprozess „Korrektur zu lesen“, indem lediglich diejenigen Primer verlängert (und somit für eine Erfassung markiert) werden, die zu der richtigen Targetsequenz hybridisierten. Es ist wahrscheinlich, dass dieses „Korrekturlesen“ die Gesamtspezifität der Untersuchung über diejenige, die mittels Hybridisierungsverfahren ohne Unterstützung erzielt werden kann, hinaus erhöht. Es ist wahrscheinlich, dass bei der XLA durch das Ligaseenzym auch sowohl die Hybridisierungseffizienz als auch die Hybridisierungsspezifität erhöht werden.

[0212] Drittens können PEA, PEACA und XLA im Gegensatz zu Oberflächenbasis-Arrayhybridisierungssystemen, die sich auf die Erfassung des Hybridisierungssereignisses selbst stützen, sogar übergangsweise stabile Primer-Target-Wechselwirkungen bezüglich einer Erfassung markieren. Die Lebensdauer der Wechselwirkung zwischen den X-mer-Vorläufern und dem Target muss nur lang genug sein, dass sie durch die Polymerase oder Ligase erkannt wird und dass die Polymerase oder Ligase ansprechend darauf agiert. Dies ermöglicht, dass eine gegebene Targetsequenz als Schablone für mehrere Vorläufer-Bindungs- und anschließende -Verlängerungs- oder -Ligationsreaktionen fungiert. Dieses zyklische Durchlaufen und die Fähigkeit, Übergangssereignisse zu erfassen, kann die Gesamterfassungssensibilität der Verfahren über die, die unter Verwendung von oberflächenbasierten Hybridisierungsuntersuchungen ohne Unterstützung erzielt werden kann, hinaus erhöhen. Wie oben erörtert wurde, könnte diese Art des zyklisch Durchlaufens einer Reaktion extern erleichtert werden, indem die Temperatur während der Verlängerungs- oder Ligationsreaktion künstlich zyklisch durchlaufen wird.

[0213] Schließlich weisen die Verlängerungs- oder Ligationsprodukte, die sich aus Verfahren ergeben, eine Massenbandbreite auf, die größer ist als die der Vorläufer-X-mere. Somit sollten die Spektral spitzen, die sich aus nicht zur Reaktion gebrachten Vorläufern ergeben, nicht die Massenspektralsignatur gewünschter Verlängerungs- oder ligierter Produkte stören. Im Fall von PEA und PEACA wird ferner in Betracht gezogen, dass massenmarkierte ddNTPs verwendet werden können. Dies ermöglicht eine größere Untersuchungsflexibilität und ermöglicht ein Multiplexieren des Massenanalyseschrittes. Ferner wird in Betracht gezogen, dass Ionisierungsmarken durch einen 5'- oder 3'-Linker in die X-mer-Vorläufer integriert werden, direkt in die Base oder den Zucker oder in die Kettenabbruch-Nukleotide. Diese Attribute sollten die Gesamtempfindlichkeit der Un-

tersuchungen erhöhen helfen und Trennschritte vereinfachen und möglicherweise eliminieren helfen, was eine Automatisierung der Untersuchung und einen Probendurchsatz erleichtert.

[0214] Die ersten drei hierin beschriebenen Verfahren sind auf ein Abfragen von Targets frei in Lösung gerichtet. Es wird jedoch auch in Betracht gezogen, dass die XLA-Methodologie in Verbindung mit oberflächengebundenen Oligonukleotiden, z. B. Arrays von Oligonukleotiden, verwendet werden kann, um die Gesamttauflösungsleistung von Arraysystemen zu erhöhen. Die Arrays beinhalten allgemein eine Oberfläche, die ein Mosaik von verschiedenen Oligonukleotiden enthält, die einzeln zu diskreten, bekannten Bereichen der Oberfläche lokalisiert sind. Derartige geordnete Arrays, die eine große Anzahl von Oligonukleotiden enthalten, wurden als Hilfsmittel für einen hohen Durchsatz umfassende Analysen von Genotyp und Genexpression entwickelt. Oligonukleotide, die auf einem Feststoffträger synthetisiert werden, erkennen eindeutig komplementäre Nukleinsäuren mittels Hybridisierung, und Arrays können entworfen werden, um spezifische Targetsequenzen zu definieren, Genexpressionsmuster zu analysieren oder spezifische allelische Variationen zu identifizieren.

[0215] Die vorliegende Erfindung kann unter Verwendung von an einen Träger angelagerten Oligonukleotiden praktiziert werden. Unter Bezugnahme auf [Fig. 6](#) können bei der vorliegenden Erfindung Arrays von Oligonukleotiden wie z. B. DNA-Arrays erzeugt werden, so dass die DNA-Sonden durch eine Art photo- oder chemisch spaltbaren Linker an ihrem 3'-Terminus an die Oberfläche angelagert werden. Der Linker kann mittels Licht-, chemischer, Oxidations-, Reduktions-, säurelabiler, basenlabiler und enzymatischer Verfahren spaltbar sein. Diese oberflächengebundenen Sonden weisen auch 5'-endständiges Phosphat auf. Beispiele für photospaltbare Linker sind diejenigen, die auf der o-Nitrobenzylgruppe beruhen, z. B. die in der WO 96/04160 beschriebenen, usw. Beispiele von Linkern, die mittels Reduktion spaltbar sind, sind diejenigen, die eine Dithioat-Funktionalität aufweisen und die mittels milder Reduktionsmittel wie z. B. Dithiothreitol oder β -Mercaptoethanol gespalten werden können. Beispiele von säurelabilen spaltbaren Linkern sind diejenigen, die eine 5'-N-Phosphoamidit-Internukleotidbindung oder ein abasisches Nukleotid als Komponente des Linkers enthalten, usw. Beispiele von basenlabilen spaltbaren Linkern sind diejenigen, die ein Ribonukleotid als Komponente des Linkers enthalten, usw.

[0216] Unter Bezugnahme auf [Fig. 6](#) beinhaltet die arraybasierte X-mer-Ligationsuntersuchung (AXLA) vier Schritte. Bei Schritt 1 wird die Targetprobe unter Bedingungen, die mit der oben beschriebenen Ligationsreaktion kompatibel sind, zu den oberflächengebundenen Sonden des Arrays hybridisiert. Die Targetnukleinsäure kann entweder nicht gekennzeichnet sein oder mit z. B. einer fluoreszierenden Kennzeichnung (an diesen Stellen ist jeweils das Band oder Gerät defekt) gekennzeichnet sein, und so weiter. Bei Schritt 2 werden die Gemisch-X-mere (nicht-phosphoryliert) zu dem Array hinzugegeben, und man lässt sie gemäß Watson-Crick-Basenpaarungsregeln entlang dem Target willkürlich hybridisieren. Bei Schritt 3 werden diejenigen X-mere, die benachbart zu den oberflächengebundenen Oligonukleotiden hybridisieren, unter Verwendung einer DNA-Ligase ligiert, wie oben beschrieben. Da lediglich die oberflächengebundenen Oligonukleotide 5'-endständige Phosphate aufweisen, erfolgt eine Ligation lediglich zwischen denjenigen 6-mer-X-meren, die neben einer DNA-Sonde hybridisiert werden, und nicht zwischen 6-mer-X-meren, die an anderen Positionen entlang dem Target nebeneinander hybridisiert werden. Bei Schritt 4 werden die ligierten X-mere anschließend von der Oberfläche abgespalten und unter Verwendung von Massenspektroskopie Merkmal um Merkmal (oder Merkmalsatz) analysiert. Die Bedingungen zum Durchführen der Ligationsreaktionen sind bei diesem Lösungssatz ähnlich den oben beschriebenen.

[0217] Der Oligonukleotid-Spaltungsschritt und der matrixunterstützte Ionisierungsschritt können gleichzeitig durchgeführt werden. Obwohl dies eine Systemkompatibilität zwischen den Bedingungen erfordert, die zur Hybridisierung, zur Oligonukleotid-Photospaltung und zur Matrixbildung und -ionisierung erforderlich sind, wird dadurch das Erfordernis, das gespaltene Produkt aus einem Reaktionsgefäß wie z. B. einem Mikrotitergefäß in ein anderes zu transferieren, eliminiert. Überdies wird dadurch das Potential, dass ein ligiertes Produkt nach dem Photospaltungsschritt, jedoch vor der Analyse, von einem Merkmal in ein anderes diffundiert, minimiert.

[0218] Bei dem obigen Modus wird die Massenzahlkomplexität (MNC, siehe nachfolgende Erörterung) durch die MNC des X-mer-Vorläufer-Gemisches (siehe nachfolgende Erörterung), multipliziert mit dem tatsächlichen Merkmalszählwert des Arrays, gegeben. In dieser Situation führt ein Anordnen aller 6-mer-Oligonukleotide (4.096 Merkmale) zu einem Array und ein Durchführen einer XLA mit einem X-mer-Gemisch zu einem effektiven Gesamtmerkmalszählwert einer ausreichenden Größe zu einer De-Novo-Sequenz von mehr als 1 kB DNA-Fragmenten. Dieses Verfahren kann auch mit nicht-einheitlichen Arrayarchitekturen kombiniert werden, um den Informationsgehalt der Untersuchung weiter zu erhöhen. Derartige nichteinheitliche Arrayarchitekturen umfassen veranschaulichungshalber und nicht einschränkungshalber die Integration, an definierten Positionen in den X-meren, entweder von universellen Basen, die mit allen vier natürlichen Basen ein Paar bilden, oder

degenerierten Basen, die mit einem definierten Teilsatz von natürlichen Basen ein Paar bilden. Die Attribute dieser Architekturtypen bezüglich oberflächenbasierter Hybridisierungsarraysysteme wurden bereits beschrieben (Pevzner P.A. u. a., J. Biomolecular Structure Dynamics 9, 399 (1991)).

Entwurf von Vorläufer-X-mer-Gemischen

[0219] Die Leistungsfähigkeit jeder oben beschriebenen Art von Untersuchung hängt von Charakteristika der zum Abfragen der Targetnukleinsäure verwendeten X-mer-Gemische ab. Wie oben erörtert wurde, ist ein hoher Massenüberlappungsgrad zwischen X-meren, die eine unterschiedliche Sequenz aufweisen, eine unausweichliche Konsequenz dessen, dass X-mera aus lediglich vier Bausteinen bestehen (siehe Histogramm der [Fig. 7a](#)). Die Reagenzien der vorliegenden Erfindung sind dazu entworfen, die Mehrdeutigkeiten zu verringern und somit die Leistungsfähigkeit bei allen Anwendungen, die Massenspektrometrie verwenden, um die Sequenz der Targetnukleinsäure zu analysieren, zu erhöhen. Diese Verringerung wird bewerkstelligt, indem ein Gemisch (Ω) von natürlichen und massenmodifizierten X-mer-Vorläufer-X-meren verwendet wird, die eine hohe Massenzahlkomplexität ($MNC(\Omega)$) und Bedeckungskomplexität ($CC_M(\Omega)$) aufweisen. Ferner sind die Gemische der Erfindung insofern generisch oder universell, als sie bei jeglicher Anwendung verwendet werden können, deren Ziel darin besteht, Sequenzinformationen einer Targetnukleinsäure zu bestimmen. Ferner können die Gemische ohne Bezugnahme auf etwaige Vorab-Informationen über die Targetnukleinsäuresequenz, einschließlich z. B. des Vorliegens, der Position oder der Identität einer Mutation, entworfen werden. Jedoch soll dies nicht implizieren, dass die Gemische nicht nützlich darin wären, Targetnukleinsäuresequenzen zu analysieren, bei denen manche Informationen über die Sequenz vorab bekannt waren, oder dass Vorab-Informationen bei der Interpretation der Massenspektren nicht hilfreich wären.

[0220] Für spezifische Anwendungen (z. B. Mutationserfassung) kann die Leistungsfähigkeit einer Untersuchung in der Länge einer Targetnukleinsäure, innerhalb derer das Problem (bei diesem Beispiel Erfassung einer Mutation) mit einer gegebenen Erfolgsrate von z. B. 95 % gelöst werden kann, gemessen werden. Mit zunehmender Leistungsfähigkeit der Untersuchung nimmt auch die Länge zu, die mit einer gegebenen Erfolgsrate analysiert werden kann. Dasselbe gilt für die Erfolgsrate, mit der gegebene Längen analysiert werden können. Ein guter Maßstab der Nützlichkeit ist die DNA-Länge, die an einem automatisierten DNA-Gel-Elektrophorese-Sequenzer analysiert werden kann, üblicherweise etwa 500 Basen. Somit ist ein vernünftiges Ziel die Analyse von 500 Basentargets mit einer Erfolgsrate > 95 %.

[0221] Um die theoretische Leistungsfähigkeit der Untersuchung zu bestimmen, genügt es, das X-mer-Massenspektrum für typische Targets, entweder in einer willkürlich gezogenen DNA-Sequenz oder einer genomischen DNA, die aus genomischen Datenbanken gezogen ist, zu simulieren und zu testen, ob die gewünschten Informationen aus dem Ergebnis extrahiert werden können. Dies ist jedoch ein zeitaufwendiger Vorgang, insbesondere wenn viele alternative Untersuchungsentwürfe betrachtet werden. Somit wird als Maß der Untersuchungsleistungsfähigkeit eine Vertretungsmaßzahl erstellt, die nachfolgend beschrieben wird. Die Untersuchung wird anschließend unter Verwendung dieser Vertretungsmaßzahl optimiert. Eine endgültige Analyse wird mittels Untersuchungssimulationen durchgeführt.

[0222] Die zur Optimierung verwendete Maßzahl ist die oben definierte durchschnittliche Mehrdeutigkeit ($A(\Omega)$). ($A(\Omega)$) wird gegeben, indem das Histogramm der Massen der möglichen Produkt-X-mera in jedem der Reaktionsgemische berechnet wird und indem anschließend die durchschnittliche Anzahl von Produkten, die mit jedem einzelnen Produkt verwechselt werden können, berechnet wird. Wenn die Produkte eine Länge N aufweisen und das Histogramm für ein Gemisch i $h_i(m)$ Zählwerte bei einer Masse m aufweist (d. h. $h_i(m)$ mögliche Produkte in diesem Gemisch weisen eine Masse m auf), so lautet die durchschnittliche Mehrdeutigkeit $\sum_i (sum(h_i(m)^2/(4^N)))$. Der Optimierungsprozess kann als Versuch angesehen werden, die Histogramme so stark wie möglich abzuflachen und zu verbreitern (siehe [Fig. 7b](#) und [Fig. 7c](#)).

[0223] Hier sollte erwähnt werden, dass, wenn eine vollständige Bestimmung des Kurzwortinhalts gewünscht wird, die Histogramme keine Spitzen aufweisen dürfen, die größer sind als Eins. In diesem Fall sollte das Optimierungsprotokoll darauf ausgerichtet sein, die maximale Spitzenhöhe des Histogramms zu verringern, statt ($A(\Omega)$) zu verringern. Die für eine vollständige Kurzwortauflösung benötigten Gemische werden anschließend erzeugt, indem aus jeder Masse ein einzelnes Oligonukleotid ausgewählt wird und diese in separate Gemische platziert werden. Dies ermöglicht, dass die Erfassung einer bestimmten Masse bei der Analyse eines Gemisches eindeutig dem gewählten X-mer zugeordnet wird. Die Anzahl von Gemischen wird somit durch die maximale Spitzenhöhe dieses abschließenden Histogramms gegeben.

[0224] Die Massenzahlkomplexität (MNC), die eine Angabe der Leistungsfähigkeit der Untersuchung ist, wird

bestimmt, indem $(A(\Omega))$ in die Gesamtanzahl möglicher Produkte aufgeteilt wird. Dies ist eine sinnvolle Maßnahme, wenn Untersuchungen verglichen werden, bei denen sich die Anzahl der Produkte unterscheidet. Unsere Simulationen haben gezeigt, dass eine auf diese Weise berechnete MNC monoton auf die theoretische Leistungsfähigkeit in Bezug darauf, Mutationen zu erfassen, sowohl den Typ als auch die Position von Mutationen zu identifizieren und eine Genotypisierung einzelner Nukleotid-Polymorphien durchzuführen, bezogen ist. Somit ist es angemessen, mit dieser Vertretungsmaßzahl zu optimieren. Als Hinweis weisen oberflächengebundene Sondenarrays aller n-mer-Oligonukleotide immer eine Mehrdeutigkeit von 1 und einen effektiven Merkmalszählwert auf, der gleich der Anzahl von Oligonukleotidsonden ist.

[0225] Wie oben angegeben wurde, besteht das Hauptziel des Entwurfsprozesses darin, die durchschnittliche Mehrdeutigkeit $(A(\Omega))$, die oben definiert wurde, zu verringern. Wenn bei einem einzelnen Reaktionsgemisch von 6-meren nur natürliche Basen verwendet werden, so werden die Mehrdeutigkeit und die MNC unter Verwendung der oben gegebenen Formel festgelegt $(MNC = 53)$, und sie sind äquivalent zu den Informationen, die mittels einer einheitlichen Fragmentierung des Targets erhalten werden könnten. Modifizierte Basen, z. B. 5-Iodcytidin, 5-Fluoruridin und 2'-O-Methylguanosin, können verwendet werden, um Oligonukleotide von der Mehrdeutigkeit zu befreien, indem bestimmte Basen selektiv durch ihre Modifikationen ersetzt werden. Wenn wir beispielsweise natürliche Basen als ACGT und einen Satz von Modifikationen als $A^*G^*C^*T^*$ bezeichnen, so kann das Oligonukleotid A*TATATT anhand von Masse von dem Oligonukleotid TATATAT unterschieden werden. Angesichts der Einschränkungen der verfügbaren Phosphoramiditbasen, der synthetischen Strategie, des Ausmaßes der Ligation oder Polymeraseverlängerung, der Länge der Vorläufer-Oligonukleotide und der Anzahl von Reaktionsgemischen ist die Optimierung somit ein Prozess des Auswählens der entsprechenden Substitutionen.

[0226] [Fig. 8](#) zeigt ein Diagramm eines Optimierungsprozesses für PEA, bei dem die X-mer-Vorläufer-X-mere einzeln synthetisierte 6-mer sind. Bei dem Prozess werden die X-mere unter Verwendung eines Kettenabbruch-Nukleosid-Triphosphats, das ein natürliches Dideoxynukleotid ist, um ein einziges Nukleotid verlängert, um den 7-mer-Kurzwortinhalt des Targets unter Verwendung eines einzigen Reaktionsgemisches zu rekapitulieren. Der Optimierungsvorgang ist darauf ausgerichtet, einen kompletten Satz von 6-mer-Vorläufern zu erzeugen, um das Massenhistogramm der resultierenden 7-mer-Produkte zu verbreitern und abzuflachen.

[0227] Zu Veranschaulichungszwecken folgt man dem Algorithmus der [Fig. 8](#) unter Verwendung von zwei Adenylatderivaten (A), nämlich 2'-Desoxyadenosin und 2'-O-Methyl-2,6-diaminopurin, die eine Masse von 330 bzw. 375 aufweisen, zwei Cytidylatderivaten (C), nämlich 2'-Desoxycytidin und 2-Desoxy-5-iodcytidin, die eine Masse 306 bzw. 434 aufweisen, zwei Guanylatderivaten (G), nämlich 2'-Desoxyguanosin und 2'-O-Methylguanosin, die eine Masse 346 bzw. 376 aufweisen, und zwei Thymidylat- (T-) Derivaten 2'-Desoxythymidin und 2'-Desoxy-5-ioduridin, die eine Masse 321 bzw. 433 aufweisen.

[0228] Man beginne, indem man ein Gemisch aller natürlichen 6-mer betrachtet. Das Massenhistogramm des Gemischs wird anschließend berechnet, und die Massen werden nach Spitzengröße (Frequenz) geordnet. Aus der Spitze, die die höchste Frequenz aufweist (größte Anzahl von 6-meren bei einer gegebenen Masse) wird ein X-mer ausgewählt, und für dieses X-mer werden unter Verwendung des oben beschriebenen Satzes von Reagenzien alternative Massen berechnet. Das X-mer wird anschließend der alternativen Masse, die die geringste Frequenz in dem Massenhistogramm aufweist, neu zugewiesen, vorausgesetzt, dass diese Neuzuweisung zu einem flacheren Histogramm führt. Wenn kein X-mer aus der höchsten Spitze neu zugewiesen werden kann, dann werden X-mere aus fortschreitend kleineren Spitzen ausgewählt und getestet. Wenn ein X-mer erfolgreich lokalisiert und neu zugewiesen wurde, wird das gesamte Massenhistogramm neu berechnet, und der Vorgang wird wiederholt, bis gemäß den beschriebenen Kriterien keine X-mere mehr neu zugewiesen werden können.

[0229] Infolge des obigen Optimierungsprozesses wird bei den verwendeten Molekülen eine durchschnittliche Mehrdeutigkeit von weniger als 30 beobachtet. Dies wird berechnet, indem die Formel für die oben beschriebene durchschnittliche Mehrdeutigkeit auf das in [Fig. 7b](#) und c angegebene abschließende Massenhistogramm angewendet wird. In diesem Fall enthält jede Spitze in dem Massenhistogramm etwa 30 von den insgesamt möglichen 16.384 7-meren (siehe [Fig. 7b](#) und c). Dies führt zu einer MNC von etwa 559 (siehe Tabelle I), die unter Verwendung der oben angegebenen und beschriebenen Gleichungen berechnet wird.

[0230] Der Wert für MNC entspricht gemäß Simulationen der Fähigkeit, eine Einzelpunktmutation in etwa 0,9 kB eines einsträngigen DNA-Targets mit einer Erfolgsrate von 95 % zu erfassen (siehe Tabelle I). In diesem Sinn kann man davon sprechen, dass ein derartiges Gemisch Informationen liefert, die zu einem oberflächengebundenen Hybridisierungsarray, das 559 Merkmale aufweist, äquivalent sind. Der Vorteil der vorliegenden

Erfindung besteht darin, dass die PEA-Untersuchung generisch ist, wohingegen Arrays in ihrer praktischen Verwendung nicht generisch sind. Das bedeutet, dass jegliche 0,9kB-Targetsequenz für Punktmutationen unter Verwendung einer PEA und des oben beschriebenen 6-mer-Vorläufer-Gemischs mit einer durchschnittlichen Erfolgsrate von 95 % abgefragt werden kann.

[0231] Für die PEA führt jegliche Aufteilung von Produkten in unterschiedliche Reaktionsgemische ferner zu einer Abnahme der Mehrdeutigkeit um etwa die Anzahl verwendeter Gemische. Ein besonders einfacher Lösungsansatz ist der folgende: 4 verschiedene Reaktionsgemische werden hergestellt, von denen jedes alle X-mer-Vorläufer und lediglich einen der 4 natürlichen Dideoxynukleotid-Terminatoren enthält. Die Spitzen in den vier Massenspektren können anschließend 7-meren, die in dem jeweiligen Extender enden, absolut zugewiesen werden. Bei diesem Lösungsansatz wird die MNC um den Faktor 4 auf den Wert 2.075 erhöht. Eine MNC dieses Wertes ermöglicht eine Mutationserfassung bei 1,6 kB eines doppelsträngigen heterozygoten PCR-Produkts (siehe Tabelle I). Derselbe Effekt kann erzielt werden, indem vier massenmarkierte Dideoxynukleotid-Terminatoren in einer einzigen Reaktion verwendet werden (siehe [Fig. 9](#)). In diesem Fall sind die Massenmarken dazu entworfen, Massenspektren der erweiterten 7-mer-Produkte gemäß der Identität des integrierten 3'-endständigen Dideoxynukleotids zu trennen.

[0232] Die Massenmehrdeutigkeit kann durch „Kategorisieren“, d. h. Abfragen des Targets mit einzelnen Teilsätzen der Gemische und anschließendes Kombinieren des Ergebnisses jeder Abfrage während der Analyse, weiter verringert werden. Das Gemisch kann mittels Partitionieren kategorisiert werden, so dass jedes einzelne X-mer in einem einzigen Teilgemisch vorliegt. Alternativ dazu kann das Gemisch derart kategorisiert werden, dass ein gegebenes X-mer eventuell in mehr als einem Teilsatzgemisch vorliegt.

[0233] Die Gemische oder Teilsatzgemische der Erfindung können entworfen sein, um die Mehrdeutigkeit des Massenspektrums der Targetnukleinsequenz auf ein beliebiges gewünschtes Maß zu verringern. Das Maß der Mehrdeutigkeit kann auf ein minimales Maß verringert werden, so dass der vollständige Kurzwortinhalt jeglicher Targetnukleinsäuresequenz ermittelt werden kann.

[0234] Zusätzlich zur Massenzahlkomplexität hängt die Auflösungsfähigkeit und Generizität der PEA von der Bedeckungskomplexität ($CC_M(\Omega)$) des X-mer-Gemisches ab. Die Simulationen zeigen, dass X-mer-Gemische, die beträchtlich weniger aufweisen als die theoretische Maximale $CC_M(\Omega)$, eine beträchtliche Auflösungsfähigkeit und Generizität beibehalten. Wie in [Fig. 10](#) gezeigt ist, können aus dem optimierten 6-mer-Gemisch bis zu 50 % der insgesamt 4.096 6-mere entfernt werden, ohne dass dies eine bedeutende Auswirkung auf die Targetlänge hat, bei der eine einzelne Mutation mit einer gegebenen Erfolgsrate erfasst werden kann. Während zusätzliche X-mere aus dem Gemisch entfernt werden, fällt die Leistungsfähigkeit der Untersuchung stark ab. Dieser Effekt ist auf das Gleichgewicht zurückzuführen, das zwischen zwei Aspekten erreicht werden muss: es muss eine ausreichend große Anzahl der gesamten X-mere vorliegen, um eine Bedeckungsgenerizität aufrechtzuerhalten, und gleichzeitig muss die Gesamtanzahl der vorliegenden X-mere ausreichend klein gehalten werden, um eine Massenüberlappung zu verringern. Wie erwartet, ist für höhere Erfolgsraten ein größerer Prozentsatz der gesamten X-mere erforderlich. Es wird erwartet, dass sehr hohe Erfolgsraten (> 99 %) erzielt werden können, indem das Target mit zwei separaten Gemischen abgefragt wird, die gleichzeitig optimiert wurden, um orthogonale Informationen zu liefern.

[0235] Für XLA hängt die Leistungsfähigkeit der Untersuchung auch von der MNC und $CC_M(\Omega)$ des X-mer-Vorläufer-Gemisches ab. Wie in Tabelle I gezeigt ist, liefert die XLA, die ein sequenzvollständiges 6-mer-Gemisch verwendet, das aus natürlichen Nukleotiden besteht, eine MNC von etwa 200. Wenn jedoch eine XLA unter Verwendung des modifizierten 6-mer-Gemisches durchgeführt wird, das für die PEA optimiert wurde, ist die MNC um etwa das Zweifache auf etwa 400 erhöht. Man sollte beachten, dass die optimalen (MNC-) Histogrammcharakteristika der X-mer-Vorläufer für PEA und XLA unterschiedlich sind. Allgemein ist das optimale Histogramm des X-mer-Vorläufers für PEA flach, wohingegen das optimale X-mer-Histogramm für XLA eher bimodal ist und eine Neigung hin zu den Enden der Massenbandbreite aufweist. Dies ist darauf zurückzuführen, dass der Effekt der Ligation der X-mer-Vorläufer darin besteht, zu bewirken, dass sich das Histogramm der Produkte sich zur Mitte der Massenbandbreite hin konzentriert, und somit, dass sich die MNC verringert. Um dieser Tendenz entgegen zu wirken und somit die MNC zu erhöhen, ist es hilfreich, zu bewirken, dass sich das X-mer-Vorläufer-Histogramm sich zu der unteren und der oberen Massengrenze hin neigt.

[0236] Wie im Fall von PEA kann die zusammengesetzte MNC für XLA erhöht werden, indem das Target mit einer Vielzahl von definierten Teilsätzen von X-mer-Gemischen abgefragt wird. Für XLA ist jedoch die optimale Kategorisierung von X-mer-Vorläufern problematischer, da jeder Vorläufer die Gelegenheit haben muss, mit jedem anderen Vorläufer zu ligieren, um die Generizität der Untersuchung aufrechtzuerhalten. Wenn also S

Sätze von X-mer-Vorläufern gemacht werden und die Untersuchungsreaktion X-mer-Paare ligiert, so müssen S(S-1)/2 Reaktionsgemische erzeugt werden.

[0237] Wie aus der obigen Erörterung hervorgeht, variiert der genaue Optimierungsprozess je nach der Art der Untersuchung und den Einzelheiten der gesuchten Informationen. Unter bestimmten Umständen kann eine erschöpfende Liste aller X-mer-Vorläufer-Sätze erstellt werden, und die Besten können ausgewählt werden. Beispielsweise kann diese Strategie realistisch verwendet werden, wenn die nachstehend beschriebene kombinatorische Synthesestrategie für 6-mer-Vorläufer verwendet wird, da die Gesamtanzahl möglicher Vorläufersätze $2^{24}/6!$ beträgt, was eine Zahl ist, die innerhalb einer überschaubaren Zeit umfassend analysiert werden kann. In komplizierteren Fällen wird ein Optimierungsalgorithmus wie der für die PEA beschriebene verwendet.

[0238] Eine potentielle Schwierigkeit, die sich ergeben kann, ist das Vorliegen von Isotopspitzen in dem Spektrum, die sich aus der Isotopenzusammensetzung des Moleküls ergeben, was die Menge an bestimmten Isotopen für einzelne Atome in dem Molekül bedeutet. Beispielsweise besteht natürlicher Kohlenstoff aus 99 % Kohlenstoff-12 und 1 % Kohlenstoff-13. Ein typisches 7-mer-Oligonukleotid weist etwa 70 Kohlenstoffatome auf und erzeugt somit zwei nahezu gleichstarke Spitzen, die eine atomare Masseneinheit auseinanderliegen. Eine Art und Weise, dieses Problem anzugehen, besteht darin, sich auf hochentwickelte Datenanalyseverfahren zu stützen, um die Isotopspitzen zu entfalten. Dies würde ein Berechnen und/oder empirisches Bestimmen der wahren Massenverteilung möglicher X-mer-Verlängerungsprodukte unter Verwendung der natürlichen Isotopenhäufigkeit beinhalten. Diese Informationen werden anschließend verwendet, um die Überlappung zwischen verschiedenen X-mer-Isotopspitzen zu entfalten. Eine zusätzliche Art und Weise, diesem Problem zu begegnen, besteht darin, die X-mer-Vorläufer derart in Teilsatzgemische zu partitionieren, dass in einem Teilsatz nur geradzahlige reine Verlängerungs- oder Ligationsprimärmassen entstehen können und in dem anderen Teilsatz lediglich ungeradzahlige Primärmassen entstehen können. Somit können die vorherrschenden N+1-Isotopspitzen (die hauptsächlich auf Kohlenstoff-13 zurückzuführen sind) nicht eine andere Primärspitze stören, da in dem gegebenen Teilsatzgemisch keine Primärmasse erzeugt werden kann. Alternativ dazu können die X-mer-Vorläufer unter Verwendung von gereinigtem Kohlenstoff-12 hergestellt werden. Bei diesem Lösungsansatz könnten die grundlegenden Nukleosidbausteine aus Bakterien oder Hefe isoliert werden, die in Medien kultiviert werden bzw. wird, die einen angereicherten oder reinen C-12-haltigen Nukleosid-Vorläufer oder eine eben solche Kohlenstoffquelle enthalten (siehe z. B. Crain, P.F., Methods Enzymol. 193, 782 (1990), Nikonowica, D. P u. a., Nucleic Acids Res. 20, 4.501 (1992), Hall, K.B., Methods Enzymol. 261, 542 (1995), Macallan u. a., Proc. Natl. Acad. Sci USA, 95, 708 (1998)). Unter Verwendung standardmäßiger Protokolle könnte das Nukleosid anschließend in geeignete Phosphoramidite zum Zweck einer automatisierten Synthese umgewandelt werden (siehe z. B. Oligonucleotides and Analogues, A Practical Approach, F. Eckstein (Herausgeber), IRL Press, Oxford (1991)). Ein ähnlicher Lösungsansatz könnte mit den Isotopspitzen verwendet werden, die durch die Verwendung von massenmodifizierten oder ionisierungsmarkierten Nukleotiden eingeführt werden. Die Massen- oder Ionisierungsmarken könnten aus isotopisch angereicherten Vorläufern hergestellt werden oder könnten alternativ dazu teilweise oder gänzlich aus Elementen bestehen, die in der Natur zum großen Teil als ein Isotop vorkommen, z. B. Fluor, Phosphor oder Iod.

Analyseschritt

[0239] Die Reaktionsgemische bzw. gereinigten X-mer-Produkte werden anschließend mittels Massenspektrometrie analysiert. Die Einzelheiten der Analyse sind in der Technik bekannt und werden hier nicht wiederholt. Geeignete Massenspektrometer sind bei Methods in Enzymology, B. Karger & W. Hancock (Herausgeber), Academic Press, San Diego, V270 (1996) und Methods in Enzymology, J. McCloskey (Herausgeber), Academic Press, San Diego, V193 (1990) beschrieben. Diese umfassen matrixunterstützte Laserdesorption/-ionisierung („MALDI“), Elektrospray- („ESI“-), Ionenzyklotronresonanz- („ICR“-), Fourier-Transformations-Typen und verzögerte Ionenextraktion und Kombinationen oder Variationen der obigen. Geeignete Massenanalysatoren umfassen magnetische Sektorfeldinstrumente/Magnetische-Ablenkung-Instrumente in/bei Einzel-Quadrupol, Tripel-Quadrupol („MS/MS“-Quadrupol), Fourier-Transformations- und Flugzeit- („TOF“-) Konfigurationen und dergleichen.

[0240] Es wird in Betracht gezogen, dass die Reaktionsprodukte vor einer Massenspektralanalyse unter Verwendung von Techniken wie z. B. Hochdruckflüssigchromatographie (HPLC), Kapillarelektrophorese und dergleichen gereinigt werden können (siehe [Fig. 5](#)). Umkehrphasen-HPLC kann verwendet werden, um die verlängerten oder ligierten X-mer-Produkte gemäß der Hydrophobie zu trennen. Die resultierenden HPLC-Fraktionen können anschließend mittels Massenspektrometrie analysiert werden. Derartige Techniken können die Auflösungsleistungsfähigkeit der beanspruchten Verfahren beträchtlich erhöhen.

[0241] Für eine Analyse mittels MALDI oder dergleichen ist es oft wünschenswert, die X-mer-Vorläufer oder X-mere zu modifizieren, um der Analyse wünschenswerte Charakteristika zu verleihen. Beispiele derartiger Modifikationen umfassen diejenigen, die durchgeführt werden, um die Laserenergie zu verringern, die erforderlich ist, um das X-mer zu verflüchtigen, um die Fragmentierung zu minimieren, um vorwiegend einfach geladener Ionen zu erzeugen, um das Ansprechverhalten der gewünschten Oligonukleotide ungeachtet der Zusammensetzung oder Sequenz zu normieren, um die Spitzenbreite zu verringern und um die Empfindlichkeit und/oder Selektivität des gewünschten Analyseprodukts zu erhöhen. Beispielsweise kann ein Modifizieren der Phosphodiester-Hauptkette der X-mere mittels Kationenaustausch aufgrund einer Heterogenität der Kationen, die pro Nukleotideinheit gebunden sind, nützlich beim Eliminieren einer Spitzenverbreiterung sein. Alternativ kann die geladene Phosphodiester-Hauptkette der X-mere neutralisiert werden, indem Phosphorthioat-Internukleotid-Brücken eingeführt werden und das Phosphorthioat mit Alkyliodid, Iodacetamid, β -Iodethanol oder 2,3-Epoxy-1-propanol alkyliert wird, um eine neutrale alkylierte Phosphorthioat-Hauptkette zu bilden. Derartige Alkylierungsprozeduren können in Kombination mit einem Ionisierungsmarkierungsschema verwendet werden, wie es beispielsweise in der WO 96/27681 ausführlich beschrieben ist.

[0242] Ferner kann es nützlich sein, Nukleotidbasen zu integrieren, die die Empfindlichkeit in Bezug auf Depurination (Fragmentierung während der Massenspektrometrie) verringern, z. B. N7- oder N9-Deazapurinnukleotide, RNA-Bausteine, Oligonukleotidtriester und Nukleotidbasen, die Phosphorthioatfunktionen aufweisen, die wie oben beschrieben alkyliert werden können, und dergleichen.

Datenanalyse

[0243] Nachdem ein Massenspektrum erhalten wird, wird eine Analyse durchgeführt, um die durch die jeweilige Anwendung definierten Informationen zu liefern. Beispielsweise erfordert eine Mutationserfassung lediglich eine qualitative Analyse der Daten, da diese Arten von Anwendungen allgemein ein Vergleichen der Massenspektren zwischen einer Referenzsequenz und einer unbekannten Variante derselben beinhalten. Wenn Unterschiede bezüglich Massenspitzen vorliegen, dann liegt bei der unbekannten Variante irgendein Mutationstyp (oder Sequenzdifferenztyp) vor.

[0244] Eine Mutationsidentifikation erfordert eine anspruchsvollere Analyse. Wie dies auch bei der Mutationserfassung der Fall ist, beinhaltet eine Mutationsidentifikation allgemein einen Vergleich zwischen einer Referenzsequenz und einer unbekannten Variante. Um jedoch die genaue Position und Identität einer heterozygoten Mutation in der Variantensequenz zu identifizieren, wird der folgende Prozess angewendet. Zuerst werden Spitzen identifiziert, die in dem Probenmassenspektrum auftreten, jedoch nicht im Wildtyp-Spektrum. Als Nächstes werden aus der Liste aller möglichen Produktgemisch-X-mere diejenigen identifiziert, deren Massen mit den neuen Spitzen übereinstimmen. Dann werden mögliche Mutationsstellen identifiziert, die dazu führen würden, dass jedes der identifizierten Produktgemisch-X-mere vorliegt. Wenn die Art der Mutation bekannt ist (z. B. Substitution), so können viele mögliche Mutationsstellen zurückgewiesen werden, und somit können viele X-mere zurückgewiesen werden. Schließlich wird das theoretische Spektrum jeder Mutation auf Übereinstimmung mit dem beobachteten Spektrum getestet.

[0245] Anspruchsvollere Prozesse können verwendet werden, um Mehrdeutigkeiten aufzulösen, die auf Unterschiede bezüglich Verlängerungs- oder Ligationseffizienzen, Ionisierungseffizienzen und Isotopeneffekten zurückzuführen sind. Je nach der CC_M und der MNC des X-mer-Vorläufer-Gemisches können unter Verwendung von Algorithmen, die denjenigen ähneln, die für das Sequenzieren unter Verwendung von Oligonukleotidarrays entwickelt wurden, De-Novo-Sequenzinformationen erhalten werden (siehe z. B.: Pevzner, P.A., J. Biomolecular Structure Dynamics 7, 63 (1989), Pevzner P.A. u. a., J. Biomolecular Structure Dynamics 9, 399 (1991), Ukkonen, E., Theoretical Computer Science 92, 191 (1992)).

Bausätze der Erfindung

[0246] Ein weiterer Aspekt der vorliegenden Erfindung bezieht sich auf Bausätze, die nützlich dafür sind, ein Verfahren gemäß der Erfindung auf zweckmäßige Weise durchzuführen. Um die Vielseitigkeit der vorliegenden Erfindung zu verbessern, können die Reagenzien in einer verpackten Kombination, in demselben oder separaten Behältern vorgesehen sein, so dass das Verhältnis der Reagenzien eine beträchtliche Optimierung des Verfahrens vorsieht. Die Reagenzien können jeweils in separaten Behältern vorliegen, oder es können verschiedene Reagenzien in einem oder mehr Behältern kombiniert sein, je nach der Stabilität der Reagenzien und der Reaktionsfreudigkeit der Reagenzien untereinander.

[0247] Bei einem Ausführungsbeispiel umfasst ein Bausatz ein Gemisch aus natürlichen und massenmodifi-

zierten X-mer-Vorläufern, die eine minimale Länge von 3 Nukleotiden und eine minimale Gemischbedeckungskomplexität von 56 geteilt durch die Anzahl gesonderter X-mere in dem Gemisch aufweisen. Die Länge der X-mer-Vorläufer kann für jeden X-mer-Vorläufer unabhängig ausgewählt werden. Jeder der X-mer-Vorläufer in dem Gemisch wird durch eine einzige chemische Spezies dargestellt. Das Gemisch weist eine Massenzahlkomplexität (MNC) auf, die größer ist als die jedes natürlichen Äquivalents.

[0248] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, und eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten besteht.

[0249] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, und eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus massenmodifizierten Kettenabbruch-Triphosphaten besteht.

[0250] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und Verlängerungs-Nukleotid-Triphosphaten besteht.

[0251] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus massenmodifizierten Kettenabbruch-Triphosphaten und Verlängerungs-Nukleotid-Triphosphaten besteht.

[0252] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten, Verlängerungs-Nukleotid-Triphosphaten und einer Nuklease besteht.

[0253] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus massenmodifizierten Kettenabbruch-Triphosphaten, Verlängerungs-Nukleotid-Triphosphaten und einer Nuklease besteht.

[0254] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten, Alpha-Thiophosphat-Nukleotid-Triphosphaten und einer 5'-Exonuklease besteht.

[0255] Bei einem anderen Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde, ein Enzym, das eine Nukleotidpolymeraseaktivität aufweist, eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus massenmodifizierten Kettenabbruch-Triphosphaten, Alpha-Thiophosphat-Nukleotid-Triphosphaten und einer 5'-Exonuklease besteht.

[0256] Bei einem weiteren Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde und eine DNA-Ligase.

[0257] Bei einem weiteren Ausführungsbeispiel umfasst ein Bausatz ein Gemisch wie es oben beschrieben wurde und ein Kondensationsmittel.

[0258] Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung ist ein Bausatz zum Durchführen eines Verfahrens, wie es oben beschrieben wurde. Der Bausatz umfasst ein Gemisch, wie es oben beschrieben wurde, eine DNA-Ligase und ein Array, das eine Oberfläche und eine Vielzahl von Nukleinsäuresequenzsonden umfasst, die einen an die Oberfläche angelagerten spaltbaren Linker und eine Nukleinsäuresequenz umfassen, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist.

[0259] Bei einem Aspekt umfasst ein Bausatz ein Kondensationsmittel, ein Array, das eine Oberfläche und eine Vielzahl von Nukleinsäuresequenzsonden umfasst, die einen an die Oberfläche angelagerten spaltbaren Linker und eine Nukleinsäuresequenz umfassen, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist,

wobei das 3'-Ende der Nukleinsäuresequenz an den spaltbaren Linker angelagert ist.

[0260] Der Bausatz kann ferner andere separat gepackte Reagenzien zum Durchführen des Verfahrens sowie Hilfsreagenzien usw. umfassen. Die relativen Mengen der verschiedenen Reagenzien in den Bausätzen können stark variiert werden, um Konzentrationen der Reagenzien zu liefern, die die Reaktionen, die während des vorliegenden Verfahrens eintreten müssen, im Wesentlichen optimieren. Unter entsprechenden Umständen kann bzw. können ein oder mehr der Reagenzien in dem Bausatz als Trockenpulver, das üblicherweise lyophilisiert ist, einschließlich Bindemitteln, vorgesehen sein, die auf eine Auflösung hin eine Reagenslösung liefern, die die geeigneten Konzentrationen zum Durchführen eines Verfahrens gemäß der vorliegenden Erfindung aufweist. Der Bausatz kann ferner eine schriftliche Beschreibung eines Verfahrens gemäß der vorliegenden Erfindung, wie es oben beschrieben wurde, umfassen.

[0261] Die Reagenzien, Verfahren und Bausätze der Erfindung sind unter anderem nützlich bei der Mutationserfassung, Mutationsidentifikation, Polymorphieanalyse, Genotypisierung, De-Novo-Sequenzierung, Neu-Sequenzierung, Genexpressionsprofilerstellung, cDNA-Clusterbildung und dergleichen.

[0262] Man sollte verstehen, dass die obige Beschreibung den Schutzmfang der Erfindung veranschaulichen und nicht einschränken soll. Fachleuten auf dem Gebiet, auf das sich die Erfindung bezieht, werden andere Aspekte, Vorteile und Modifikationen innerhalb des Schutzmangs der Erfindung einleuchten. Die folgenden Beispiele sind dargelegt, um Fachleuten Beispiele an die Hand zu geben, wie das Verfahren und die Produkte der Erfindung hergestellt und genutzt werden können, und sollen den Schutzmang dessen, was die Erfinder als ihre Erfindung ansehen, nicht einschränken.

Beispiele

[0263] Die folgenden drei Beispiele beziehen sich auf die oben beschriebenen Verfahren, die eine Region der menschlichen p53-Gen-Sequenz als Targetnukleinsäure verwenden. [Fig. 11](#) zeigt eine 62- und eine 378-Nukleotid-Region des p53-Gens mit bekannten Mutationssstellen, die in Fettdruck angegeben sind. Für alle Analysen wird das Komplement der in [Fig. 11](#) angegebenen Sequenzen verwendet. Alle Beispiele sind Simulationen. Somit sind die Besonderheiten bezüglich der Reaktionsbedingungen (d. h. Puffer, X-mer- und Targetkonzentrationen, Polymerase- oder Ligasotyp usw.) hier nicht relevant. Eine Interpretation dieser Beispiele hängt lediglich von der Massenkomplexität und der Bedeckungskomplexität der X-mer-Vorläufer, der Targetlänge und -sequenz und der Art der verwendeten Untersuchung ab. Alle Beispiele gehen davon aus, dass die Reaktionen so verlaufen, wie sie in dem ganzen Text beschrieben und in den Figuren angegeben sind. Wichtig ist, dass davon ausgegangen wird, dass in der Tat lediglich diejenigen X-mere, die exakte Komplemente der Targetsequenz sind, verlängert (im Fall von PEA & PEACA) oder ligiert (im Fall von XLA) werden. Der Hauptzweck aller Beispiele besteht darin, die theoretische Leistungsfähigkeit jeder Untersuchung bezüglich der Art von Massenspektren und des Informationsgehalts, die bzw. den jede Untersuchung erzeugen würde, zu veranschaulichen.

Beispiel 1 (PEA)

[0264] Unter Verwendung des 62-Nukleotid-p53-Fragments als Target, aller vier natürlichen ddNTPs und eines sequenzvollständigen Satzes (4.096) von 6-meren, die aus den vier natürlichen Nukleotiden bestehen, wird eine PEA durchgeführt. [Fig. 12](#) gibt den Satz von 56 überlappenden X-mer-Verlängerungsprodukten an, die für die Wildtyp-p53-Targetsequenz erwartet werden. Die Massenspektren der 7-mer-Produkte, die dem Wildtyp und der einzigen G2481C-Mutante entsprechen, sind in [Fig. 13](#) angegeben. Die Integrierte-Differenz-Spektren in [Fig. 13](#) offenbaren, welche Massen sich zwischen dem Wildtyp und der Mutante unterscheiden. Positive-Differenz-Spitzen entsprechen Massen, die beim Wildtyp, jedoch nicht bei der Mutante vorliegen, wohingegen Negative-Differenz-Spitzen Massen in der Mutante, jedoch nicht im Wildtyp entsprechen. Diese Spektraldata werden dann wie oben beschrieben interpretiert.

[0265] Jedoch ist es wichtig zu betonen, dass die in den Integrierte-Differenz-Spektren offenbarten Informationen davon ausgehen, dass die Hybridisierungs-, Verlängerungs-, Ionisierungs- und Erfassungsschritte für alle X-mere mit gleicher Effizienz erfolgen. Da es unwahrscheinlich ist, dass dieses Quantisierungsniveau der Fall ist, auch nicht bei einer guten Optimierung, werden die einzelnen Spektraldata auf eine binäre Form reduziert ([Fig. 14](#)). Diese Art der Transformation erfordert dann lediglich, dass die obigen Schritte einen definierten Schwellenpegel erfüllen. Obwohl diese Eliminierung der quantitativen Beschaffenheit der Daten die Gesamtleistungsfähigkeit der Untersuchung schmälert, offenbaren die resultierenden Binärdifferenzspektren trotzdem Unterschiede zwischen dem Wildtyp und der Mutante ([Fig. 14](#)). [Fig. 15](#) gibt die aus der PEA result-

tierenden Massenspektren für alle drei möglichen Mutationen an den Positionen 2481 und 2482 in dem 62-Nukleotid-p53-Fragment an. Wie in [Fig. 16](#) gezeigt ist, sind die Binärdifferenzspektren ausreichend leistungsfähig, um alle sechs Mutationen zu offenbaren.

[0266] Wenn die C2481A-Mutation mittels einer PEA in dem 378-Nukleotid-p53-Fragment abgefragt wird, offenbaren die resultierenden Binärdifferenzspektren lediglich eine einzige Massendifferenz ([Fig. 17](#)). Wenn jedoch dieselbe Abfrage unter Verwendung des oben beschriebenen einzeln optimierten massenmodifizierten 6-mer-Gemisches durchgeführt wird, offenbaren die resultierenden Differenzspektren sechs Massendifferenzen ([Fig. 18](#)). Dies veranschaulicht die erhöhte Leistungsfähigkeit, die dieser Untersuchung zur Verfügung steht, wenn an den Vorläufern Massenmodifikationen vorgenommen werden.

[0267] Es sollte betont werden, dass das obige Ergebnis generisch ist – die Auswahl von X-mer-Vorläufern wurde ohne Verwendung von Informationen über das bestimmte zu analysierende Target getroffen. Wie in Tabelle 1 angegeben ist, offenbaren PEA-Analysen, die das einzeln optimierte 6-mer-Gemisch verwenden, zumindest eine Binärmassendifferenz für eine einzige Mutation in einem 890-Nukleotide langen Target mit einer Erfolgsrate von 95 %. Wenn eine höhere Erfolgsrate gewünscht wird, wird eine kürzere Targetlänge, die abgefragt werden kann, verwendet (siehe [Fig. 10](#)). Längere Targets oder höhere Erfolgsraten sind unter Verwendung einer PEA zugänglich, wenn das Target entweder mit einem orthogonalen Satz von X-mer-Gemischen oder mit jedem ddNTP in einer separaten Reaktion abgefragt wird (siehe Tabelle 1).

Beispiel 2 (PEACA)

[0268] Das 62-Nukleotid-p53-Fragment wird abgefragt, indem vier separate PEACA-Analysen durchgeführt werden, wobei jede einen definierten Satz von X-meren, dNTPs und ddNTPs enthält. Eine Reaktion (A) enthält alle 6-meren, die die Form $X_1X_2X_3rA_4X_5X_6$ (wobei rX ein Ribonukleotid ist), dGTP, dCTP, dTTP und ddATP aufweisen. Desgleichen enthält die (G)-Reaktion alle 6-meren, die die Form $X_1X_2X_3rG_4X_5X_6$ (wobei rG ein Ribonukleotid ist), dATP, dCTP, dTTP und ddGTP aufweisen, und so weiter. Das erwartete Ergebnis dieses Prozesses ist ein Satz von halb-überlappenden X-mer-Produkten, die variable Längen aufweisen ([Fig. 19](#)). Die resultierenden zusammengesetzten Massenspektren (wobei alle vier Reaktionen vermischt sind) sowohl für den Wildtyp als auch die einzige C2481A-Mutante sind in [Fig. 20](#) angegeben. Die Binärdifferenzspektren offenbaren sieben Massenspitzendifferenzen, die auf die gegebene Mutation hinweisen.

[0269] Es ist wichtig zu beachten, dass das PERCA-Verfahren zu 4-mer-Spaltungsprodukten ($X_1X_2X_3A_4$, $X_1X_2X_3G_4$, $X_1X_2X_3C_4$ und $X_1X_2X_3T_4$) führt, die nicht auf die Targetsequenz hinweisen. Somit sind alle informativen Informationen in den Binärdifferenzspektren auf diejenigen Massen beschränkt, die 5-meren und größer (> 1.550 amu) entsprechen. Sogar unter diesem Vorbehalt liegen ausreichend Massendifferenzen vor, um die Mutation zu identifizieren.

[0270] Alle drei Mutationen bei Position 2481 in dem 378-Nukleotid-p53-Fragment werden ebenfalls unter Verwendung dieser bestimmten Version der Analyse offenbart ([Fig. 21](#)).

[0271] Diese Ergebnisse sind generisch; die PEACA kann unter Verwendung von natürlichen X-mer-Vorläufern eine einzelne Mutation in einem 380 Nukleotide langen Target mit einer Erfolgsrate von 95 % erfassen (siehe Tabelle I). Die Auflösungsleistungsfähigkeit der PEACA kann erhöht werden, indem an den Positionen X_5 und X_6 spezifische Massenmodifikationen integriert werden. Eine größere Leistungsfähigkeit kann erzielt werden, indem das definierte rX weiter zu dem 5'-Terminus hin, z. B. bei X_3 , platziert wird, und indem definierte Massenmodifikationen an Positionen X_4 , X_5 und X_6 integriert werden.

Beispiel 3 (XLA)

[0272] Das 62-Nukleotid-p53-Fragment wird abgefragt, indem eine XLA durchgeführt werden, bei der ein 6-mer-Gemisch verwendet wird, das aus natürlichen Nukleotiden gebildet ist. [Fig. 22](#) gibt die einzelnen Massenspektren für die Wildtypsequenz und die C2481A-Mutante an. Die Binärdifferenzspektren offenbaren fünf Massenspitzendifferenzen, die auf die gegebene Mutation hinweisen. Dieselbe Mutation kann mittels XLA in dem 378-Nukleotid-Fragment erfasst werden, jedoch offenbaren die resultierenden Binärdifferenzspektren lediglich drei Massendifferenzen ([Fig. 23](#)). Wie oben erörtert und in Tabelle 1 gezeigt ist, kann die Leistungsfähigkeit der XLA durch Verwendung von massenmodifizierten X-mer-Gemischen erhöht werden.

[0273] Alle in der vorliegenden Spezifikation erwähnten Veröffentlichungen und Patentanmeldungen sind durch Bezugnahme in das vorliegende Dokument aufgenommen, so als ob jede einzelne Veröffentlichung oder

Patentanmeldung spezifisch und einzeln als durch Bezugnahme aufgenommen angegeben wäre.

[0274] Obwohl die vorstehende Erfindung zu Zwecken eines klaren Verständnisses veranschaulichungshafter und beispielhaft einigermaßen ausführlich beschrieben wurde, wird Fachleuten angesichts der Lehren dieser Erfindung ohne weiteres einleuchten, dass an derselben bestimmte Änderungen und Modifikationen vorgenommen werden können, ohne von der Wesensart oder dem Schutzmfang der beigefügten Patentansprüche abzuweichen.

Tabelle I. Auswirkung der X-mer-Zusammensetzung auf den Informationsgehalt für verschiedene Untersuchungen

Unter- suchung	X-mer- Länge	Verfahren der X-mer- Synthese	X-mer- Zusammen- setzung	Anzahl von Reaktionen pro Target	MNC	einsträngige Mutations- erfassung (Länge in Nukleotiden)	doppel- strängige Mutations- erfassung (Länge in Nukleotiden)	einsträngige Mutations- identifi- kation (Länge in Nukleotiden)
PEA	6-mere	kombina- torisch	natürlich	1	53	100*	40*	70*
PEA	6-mere	kombina- torisch	natürlich	4 (1 für jedes ddNTP) †	173	300*	130*	190*
PEA	6-mere	einzeln	natürlich	10 (Teil- sätze von Vorläufer- Gemischen)	530	850*	400*	600*
PEA	5-mere	kombina- torisch	massenmodi- fiziert †	1	275	520	212	360
PEA	5-mere	kombina- torisch	massenmodi- fiziert †	4 (1 für jedes ddNTP) †	835	1.450	540	1.040

PEA	6-mere kombina- torisch	massenmodi- fiziert +	1	348	700	320	500
PEA	6-mere kombina- torisch	massenmodi- fiziert +	4 (1 für jedes ddNTP) ‡	1.198	2.650	950	1.750
PEA	6-mere einzel mit Massen- optimierung	massenmodi- fiziert +	1	559	890	440	630
PEA	6-mere einzel mit Massen- optimierung	massenmodi- fiziert +	4 (1 für jedes ddNTP) ‡	2.075	3.600	1.600	2.400*
PEA	6-mere einzel mit Massen- optimierung	massenmodi- fiziert +	10 (Teil- sätze von Vorläufer- Gemischen)	5.590	8.900*	4.400*	6.300*
XLA	6-mere kombina- torisch	natürlich	1	nicht er- mitt.	~ 400	~ 200	~ 300
XLA	6-mere kombina- torisch	massenmodi- fiziert +	1	nicht er- mitt.	~ 900	~ 450	~ 700

PEACA	6-mere	kombinatorisch	natürlich	4 (gemischt für MS-Analyse)	nicht ermitt.	~ 380	~ 170	nicht ermittelt
-------	--------	----------------	-----------	-----------------------------	---------------	-------	-------	-----------------

[0275] Berichtete Werte sind aus in dem Text beschriebenen Berechnungen oder Simulationen abgeleitet. Die Werte in den letzten drei Spalten beziehen sich auf die Nukleotidlängen, innerhalb derer das Problem mit

einer 95%igen Erfolgsrate gelöst werden kann. Die MNC ist im Text definiert. Mutationsidentifikation bezieht sich auf die Fähigkeit, sowohl die Identität als auch die Position der Mutation in einer ansonsten bekannten Sequenz zu ermitteln. * Die berichteten Werte für diese Untersuchungsbedingungen sind lediglich Schätzungen, da erst noch vollständige Simulationen durchgeführt werden müssen. † Die Sätze von Vorläufern, die massenmodifizierte Nukleotide enthalten, waren aus den vier natürlichen Nukleotiden und den folgenden Derivaten zusammengesetzt: 2'-O-Methyl-2,6-diaminopurin, 5-Iodcytidin, 2'-O-Methylguanosin und 5-Ioduridin. Mutationserfassung bezieht sich auf die Fähigkeit, Unterschiede bei Massensignaturen zwischen einem Wildtyp und einer einzelnen Mutante desselben zu erfassen. ‡ Man sollte beachten, dass alternativ dessen, vier separate Reaktionen (1 für jedes ddNTP) durchzuführen, eine einzige Reaktion durchgeführt werden könnte, wenn massenmarkierte ddNTPs so entworfen wären, dass sie alle 7-mer-Verlängerungsprodukte in vier discrete Regionen der Massenspektren trennen, je nach der endständigen ddNTP, die wie oben und in [Fig. 9](#) beschrieben hinzugefügt wird.

Patentansprüche

1. Ein Gemisch oder Satz von Teilgemischen, das beziehungsweise der natürliche und massenmodifizierte X-mer-Vorläufer umfasst,
wobei die X-mer-Vorläufer eine minimale Länge von 3 Nukleotiden aufweisen,
wobei die Gemischbedeckung eine minimale Gemischbedeckungskomplexität oder eine zusammengesetzte Mischbedeckungskomplexität von 56 geteilt durch die Anzahl gesonderter X-mer-Vorläufer in dem Gemisch aufweist,
wobei die Massenzahlkomplexität des Gemischs höher ist als die Massenzahlkomplexität irgendeines anderen natürlichen Äquivalents des Gemischs,
wobei die Länge unabhängig für jeden X-mer-Vorläufer ausgewählt ist und
wobei jeder der X-mer-Vorläufer in dem Gemisch eine eindeutige Masse aufweist.
2. Ein Gemisch gemäß Anspruch 1, bei dem die X-mer-Vorläufer eine definierte isotope Zusammensetzung aufweisen.
3. Ein Gemisch gemäß Anspruch 1 oder 2, bei dem das Gemisch oder der Satz von Teilgemischen eine Gemischbedeckungskomplexität von zumindest etwa 1/2 aufweist, wenn das Gemisch zumindest 128 gesonderte X-mere enthält, eine Gemischbedeckungskomplexität von zumindest etwa 1/4 aufweist, wenn das Gemisch zumindest 256 gesonderte X-mere enthält, und eine Gemischbedeckungskomplexität von zumindest etwa 1/8 aufweist, wenn das Gemisch zumindest 512 gesonderte X-mere enthält.
4. Ein Gemisch gemäß einem der vorhergehenden Ansprüche, bei dem die X-mer-Vorläufer eine Ionisierungsmarke aufweisen.
5. Ein Gemisch gemäß einem der vorhergehenden Ansprüche, bei dem die Zusammensetzung des Gemischs bekannt ist.
6. Ein Gemisch gemäß einem der vorhergehenden Ansprüche, bei dem zumindest manche der massenmodifizierten X-mer-Vorläufer zumindest eine Massenmarke oder zumindest eine chemische Modifizierung einer Internukleosidbindung, einer Zuckerhauptkette oder einer Nukleosidbase aufweisen.
7. Ein Verfahren zum Analysieren einer anvisierten Nukleinsäuresequenz, das folgende Schritte umfasst:
(1) Hybridisieren eines Gemischs oder Satzes von Teilgemischen, das beziehungsweise der natürliche und massenmodifizierte X-mer-Vorläufer aufweist, zu einer anvisierten Nukleinsäure,
wobei das Gemisch eine minimale Gemischbedeckungskomplexität oder eine zusammengesetzte Mischbedeckungskomplexität von 56 geteilt durch die Anzahl von gesonderten X-mer-Vorläufern in dem Gemisch aufweist,
wobei die Länge für jeden X-mer-Vorläufer unabhängig ausgewählt ist,
wobei jeder der X-mer-Vorläufer in dem Gemisch eine eindeutige Masse aufweist,
wobei die X-mer-Vorläufer ein 3'-Ende und ein 5'-Ende aufweisen,
(2) Verarbeiten der Hybride, um die Masse der X-mer-Vorläufer-Abschnitte der Hybride in einer anvisierten, durch die Sequenz bewirkten Reaktion zu verändern; und
(3) Analysieren des Produkts des Schritts (2) mittels Massenspektrometrie.
8. Ein Verfahren gemäß Anspruch 7, bei dem die anvisierte Nukleinsäuresequenz ein 3'-Ende und ein 5'-Ende aufweist und zu einer Vielzahl von Nukleinsäuresonden in einem Array hybridisiert wird, das folgende

Merkmale aufweist:

- (a) eine Oberfläche; und
- (b) die Vielzahl von Nukleinsäuresonden, die Folgendes umfassen:
 - (i) einen spaltbaren Linker, der an der Oberfläche angelagert ist; und
 - (ii) eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an dem spaltbaren Linker angelagert ist;
und wobei der Verarbeitungsschritt folgende Schritte umfasst:
Ligieren der hybridisierten X-mer-Vorläufer, die neben dem endständigen 5'-Phosphat angeordnet sind, mit der oberflächengebundenen Sonde, um einen Komplex aus hybridisiertem Vorläufer/Sonde zu bilden, wobei die anvisierte Nukleinsäuresequenz an demselben angelagert ist; und
Spalten des Komplexes an dem spaltbaren Linker.

9. Ein Bausatz zum Ausführen eines Verfahrens zum Analysieren einer anvisierten Nukleinsäuresequenz, der folgende Merkmale aufweist:

- a. das Gemisch gemäß Anspruch 1; und
- b. das Enzym, das eine Nukleotidpolymerase-Aktivität aufweist; und
- c. eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht; oder
- b. ein Enzym, das eine DNA-Polymerase-Aktivität aufweist;
- c. eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht;
- d. eine Vielzahl von Verlängerungs-Nukleotid-Triphosphaten; oder
- b. ein Enzym, das eine DNA-Polymerase-Aktivität aufweist;
- c. eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht;
- d. eine Nuklease; und optional
- e. eine Vielzahl von Verlängerungs-Nukleotid-Triphosphaten; oder
- b. eine DNA-Polymerase;
- c. eine Vielzahl von Nukleotiden, die aus der Gruppe ausgewählt sind, die aus natürlichen Kettenabbruch-Triphosphaten und massenmodifizierten Kettenabbruch-Triphosphaten besteht;
- d. eine 5'-Exonuklease; und
- e. eine Vielzahl von Verlängerungs-Nukleotid-Triphosphaten und 5'-(α)-Phosphothioat-Analoga; oder
- b. eine DNA-Ligase; und optional
- c. eine Nuklease; oder
- b. ein Kondensationsmittel.

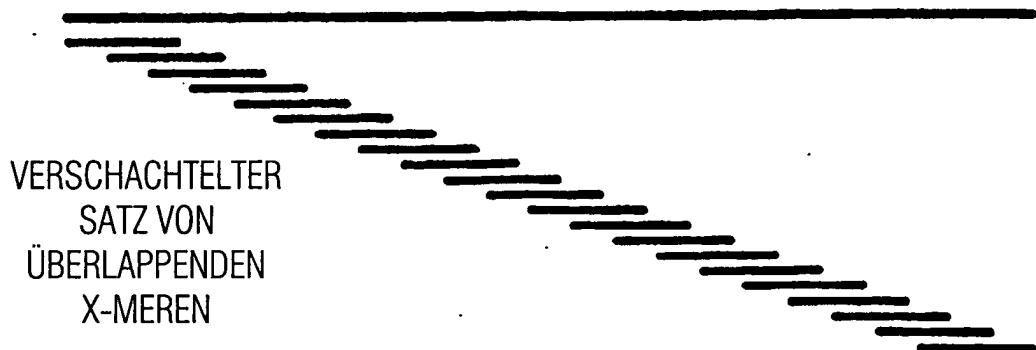
10. Ein Bausatz gemäß Anspruch 9, der ferner folgendes Merkmal aufweist:

ein Array, das folgende Merkmale aufweist:

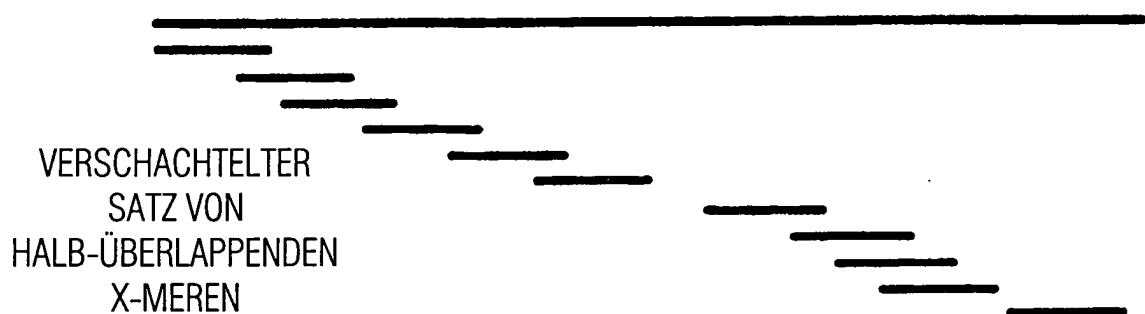
- (a) eine Oberfläche; und
- (b) eine Vielzahl von Nukleinsäuresequenzsonden, die Folgendes umfassen:
 - (i) einen spaltbaren Linker, der an der Oberfläche angelagert ist; und
 - (ii) eine Nukleinsäuresequenz, die ein 3'-Ende und ein endständiges 5'-Phosphat aufweist, wobei das 3'-Ende der Nukleinsäuresequenz an dem spaltbaren Linker angelagert ist.

Es folgen 21 Blatt Zeichnungen

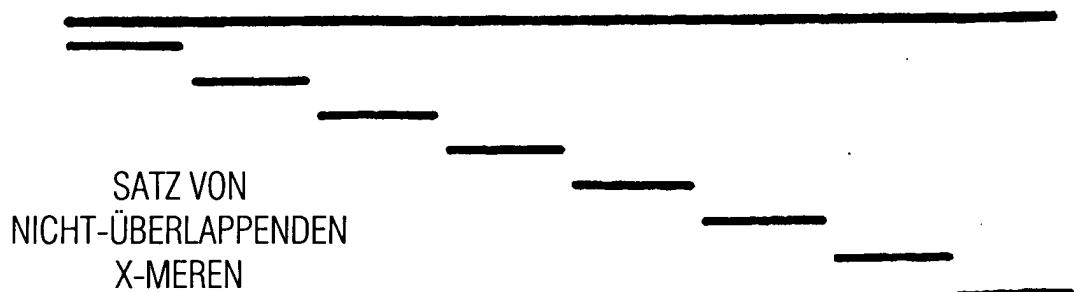
FIGUR 1A



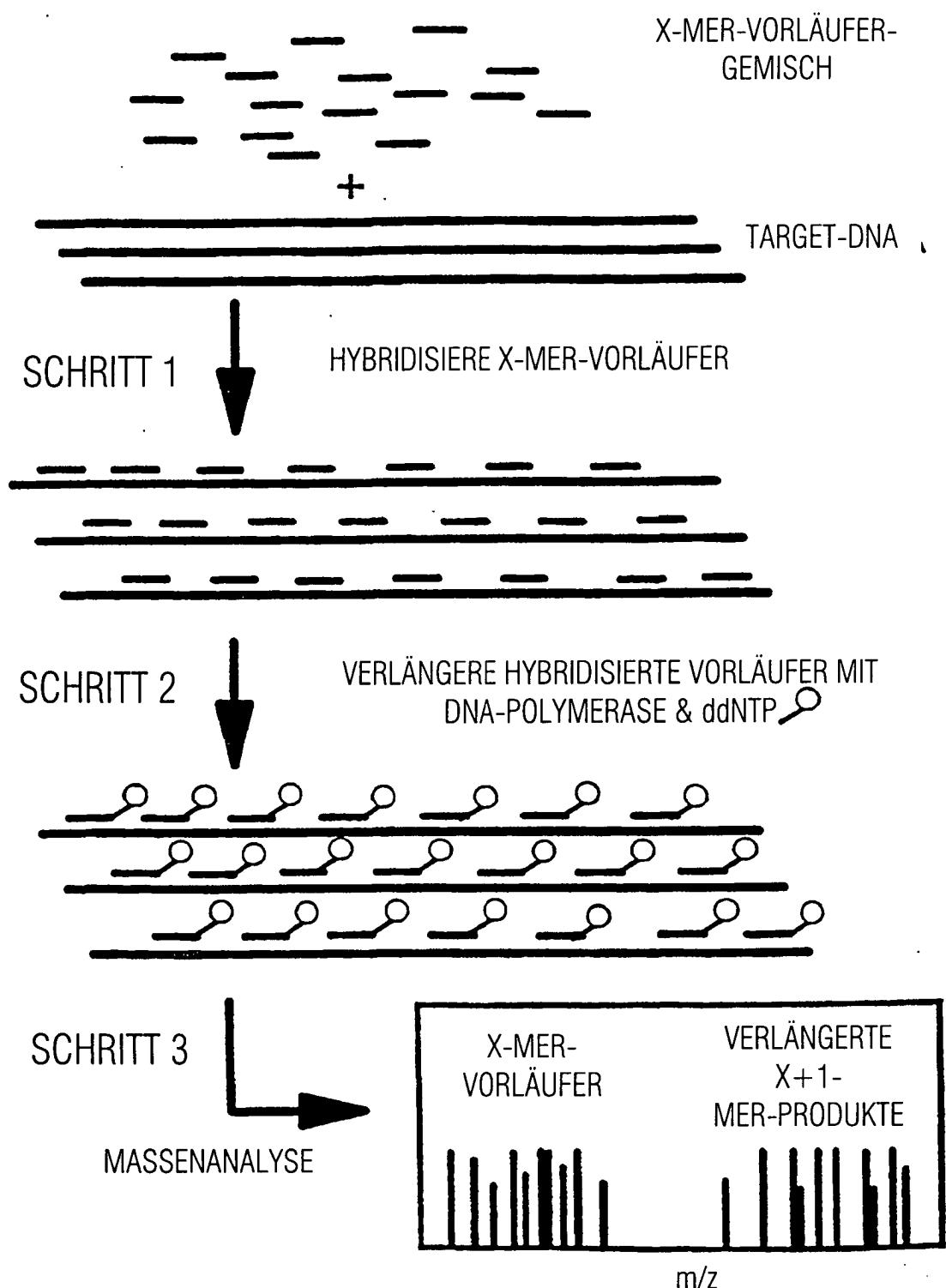
FIGUR 1B



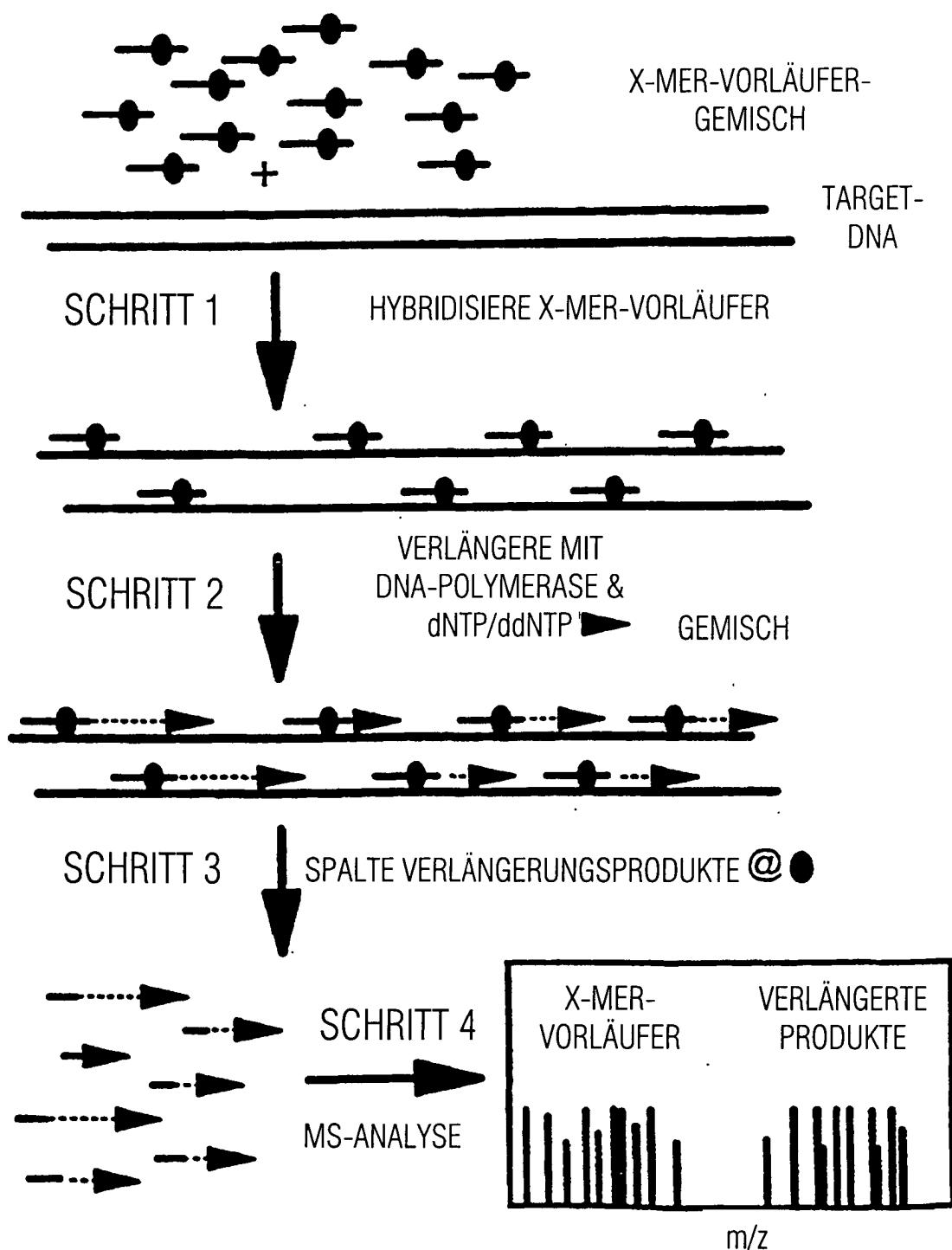
FIGUR 1C



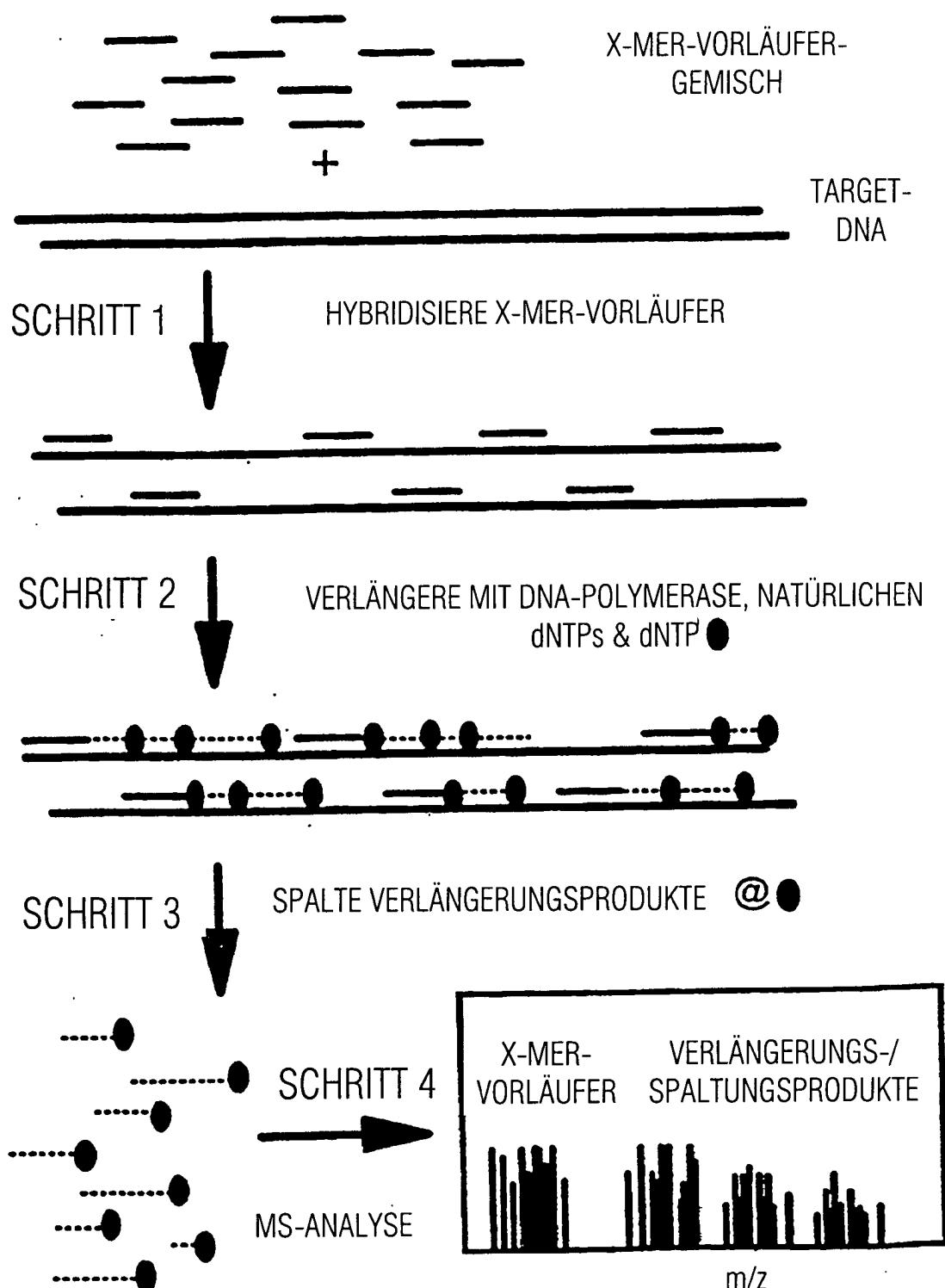
FIGUR 2



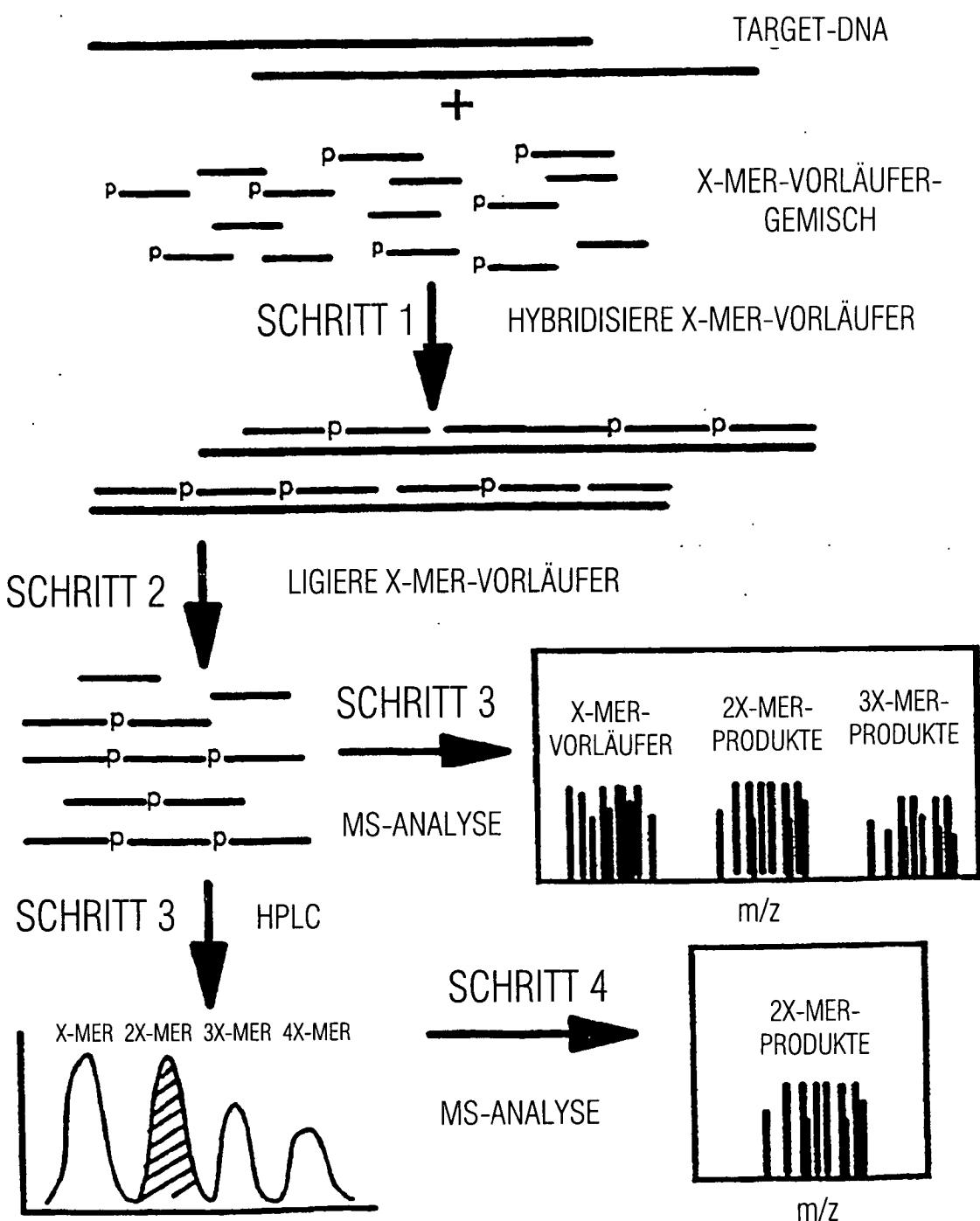
FIGUR 3



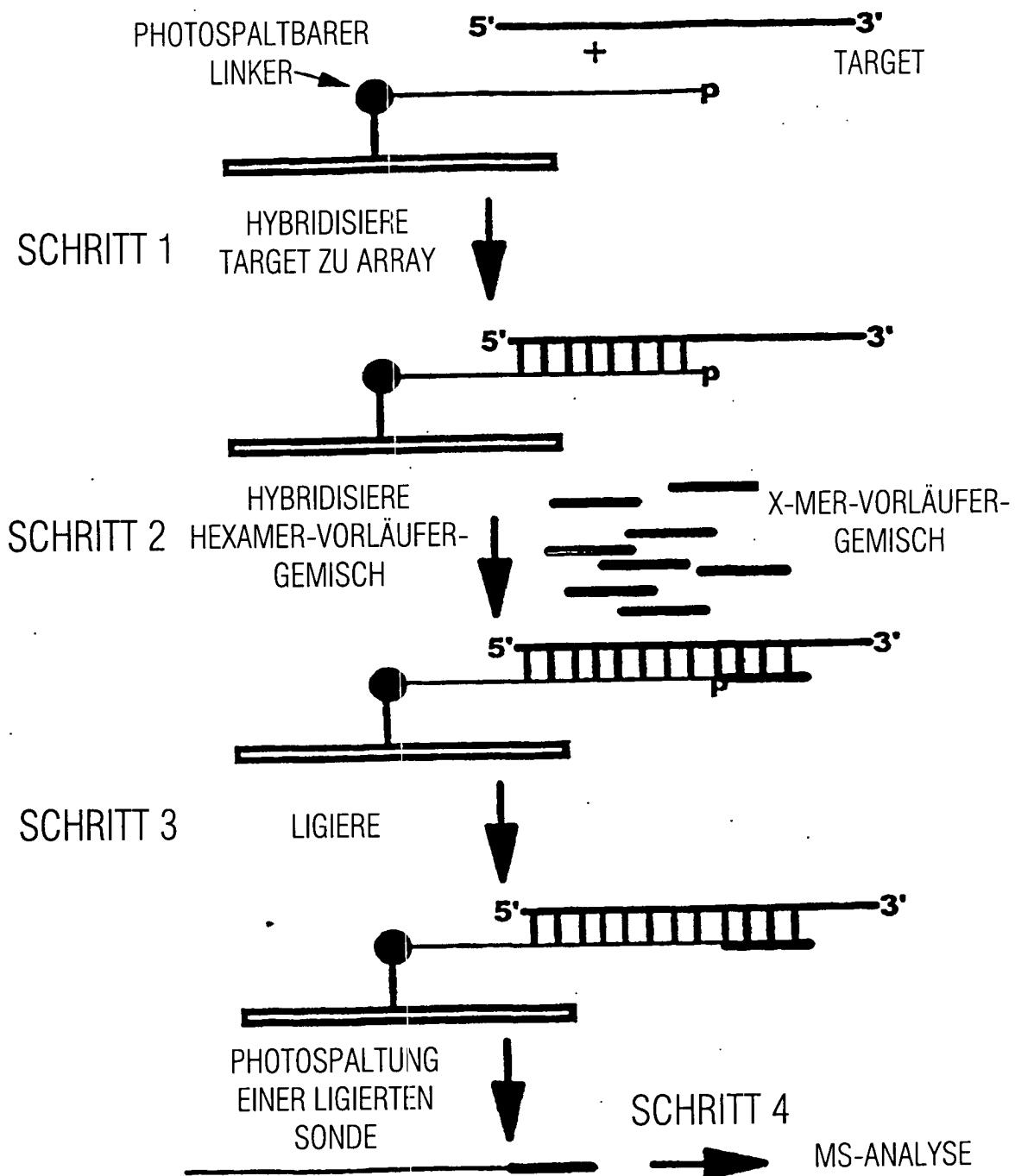
FIGUR 4



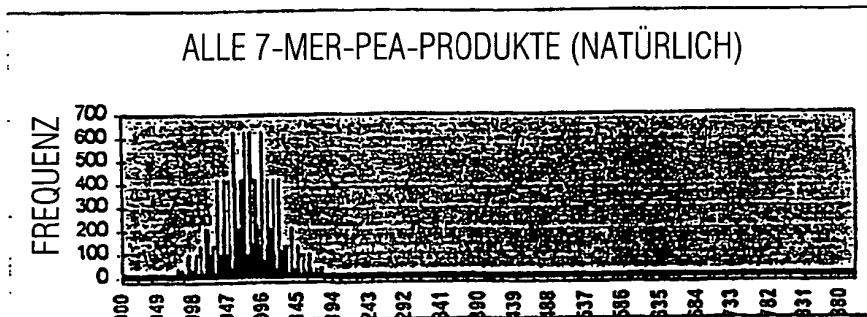
FIGUR 5



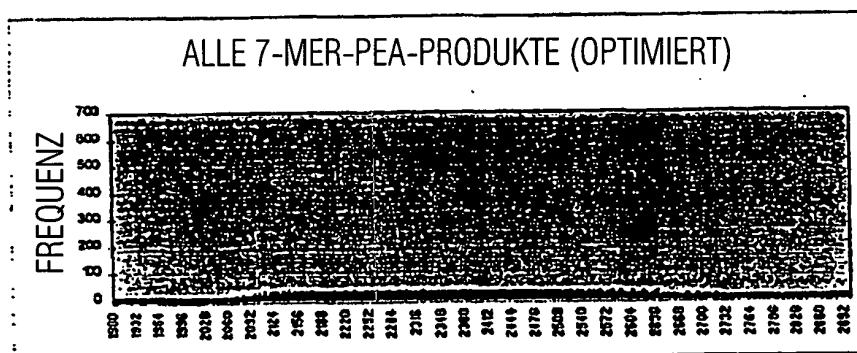
FIGUR 6



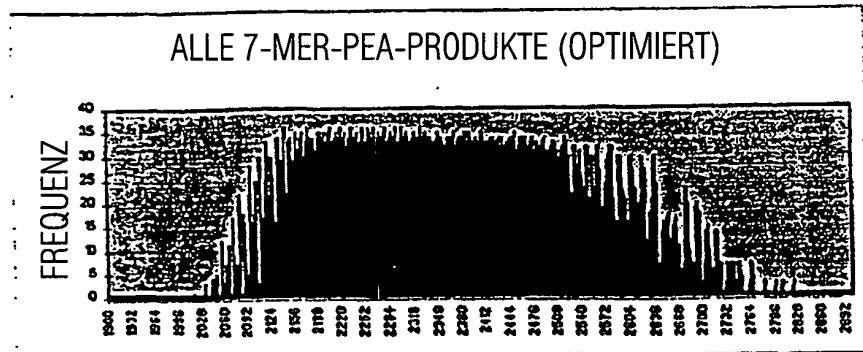
FIGUR 7A



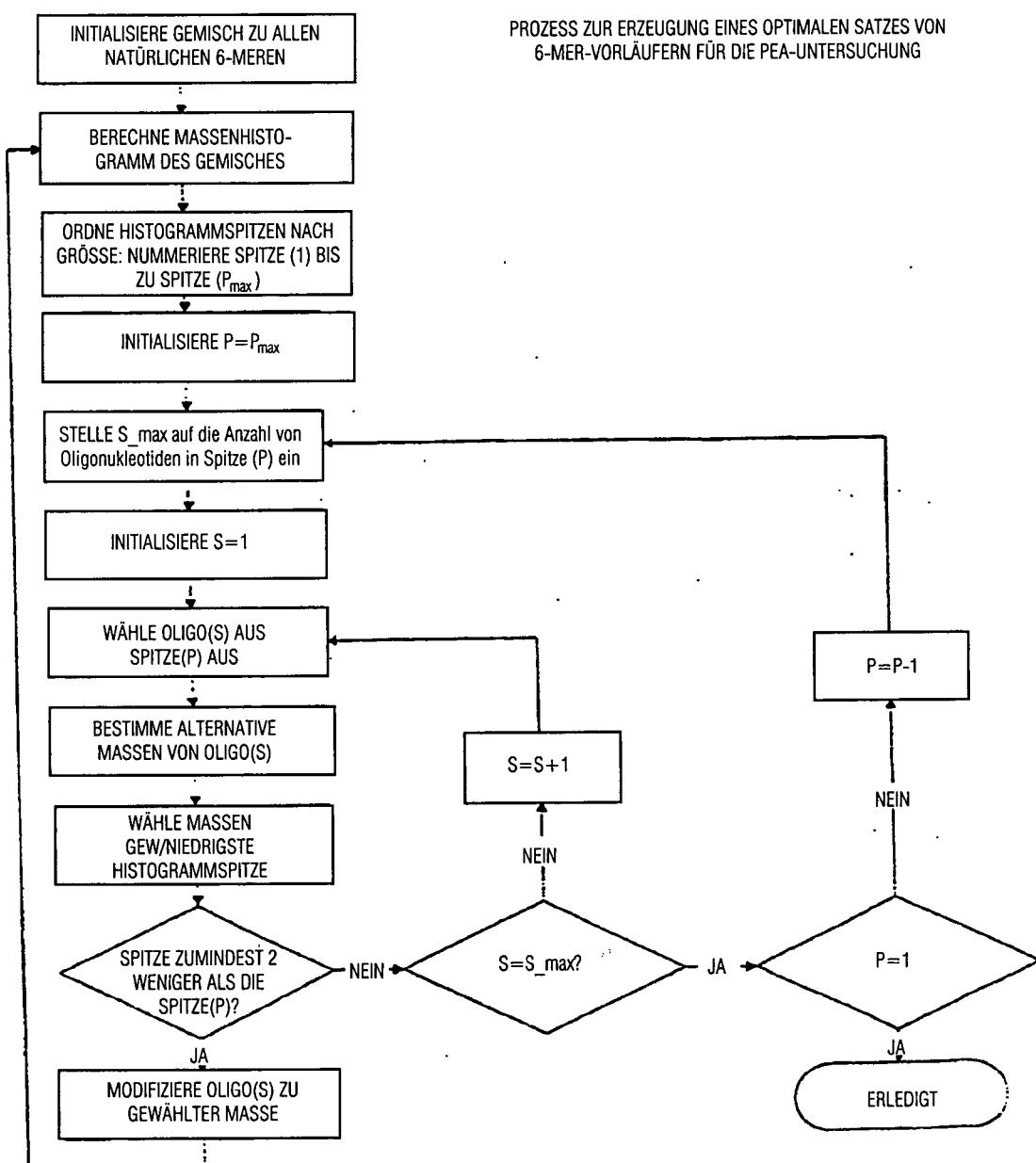
FIGUR 7B



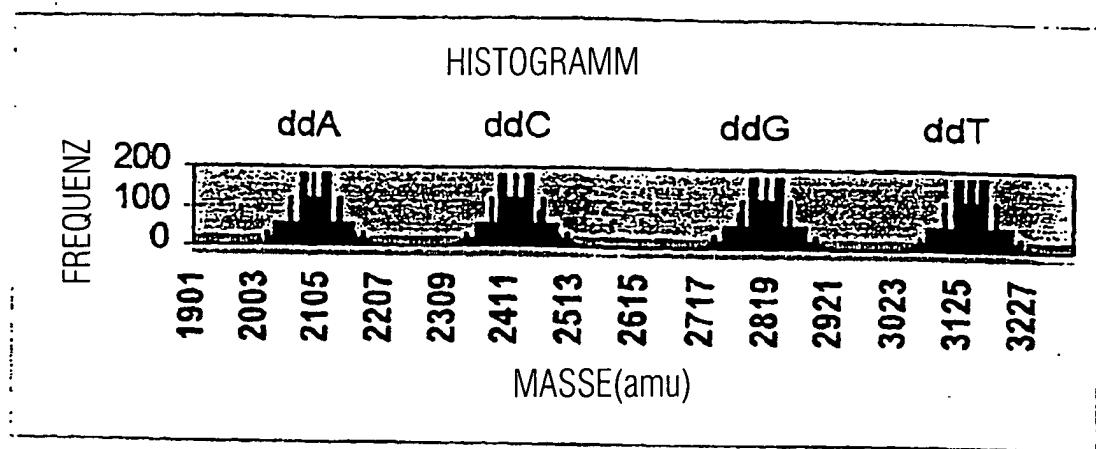
FIGUR 7C



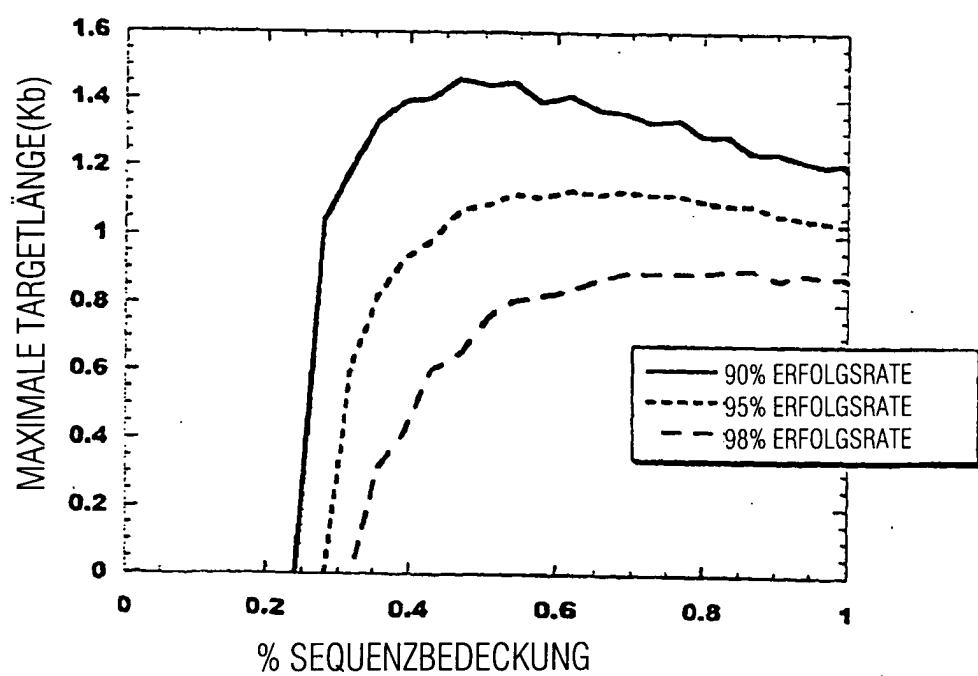
FIGUR 8



FIGUR 9



FIGUR 10



FIGUR 11

62-NUKLEOTID-FRAGMENT DES WILDTYP-P53-GENS UND 6 BEKANNTER MUTATIONEN:

P53_wt
AtgtgttaacagttcctgcatggcGCCatgaacCCGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 1)

P53_2451-6
AtgtgttaacagttcctgcatggcCGCatgaaccGGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 2)

P53_2481-1
AtgtgttaacagttcctgcatggcGCCatgaacTGGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 3)

P53_2481-2
AtgtgttaacagttcctgcatggcGCCatgaacGCGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 4)

P53_2481-5
AtgtgttaacagttcctgcatggcGCCatgaacAGGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 5)

P53_2482-3
AtgtgttaacagttcctgcatggcGCCatgaaccAGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 6)

P53_2482-5
AtgtgttaacagttcctgcatggcGGCatgaaccTGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 7)

P53_2482-6
AtgtgttaacagttcctgcatggcGCCatgaaccCGAGGcccatttcaccatcatcacactg (SEQUENZKENNNR. 8)

378-NUKLEOTID-FRAGMENT DES WILDTYP-P53-GENS

CagtcacacgacatgacggagggttgtgaggCGCgtccccccacatgagcgctgtcgatagcgatggctggcccccttcacgcattctatccgagttggaaaggaaattttcgctgtggatattggatgacagaaaacttttcgacatagtgtgtggctgtggccctatgagccggcttggctgtactgtaccaccatccactacaactacatgttaacagttcctgcatggcGCCatgaacCCGAGGcccatttcaccatccatcatcacactggaaagactccagttgtgtaatctactggacggacagctttgagggtgCCTgtttgtgcctgtccctggagagacccggcgcacagaggaagagaatctcccaagaaaggggagccctcaccacgagctgccccccagggagacactaagcgagcactgccccaaacaaacccagctccctcccccagccaaaggaaaccactggatggagaatatttcacccttcagatccgtgggcgtgagcgcttcgagatgttcccgagagctqaatgaggcccttggaaact (SEQUENZKENNNR. 9)

740-NUKLEOTID-FRAGMENT DES WILDTYP-P53-GENS

cagggcagctacggttccgtctgggttttgcatttcgtggacagccaaatgtgtgacttgcacgtactccccctgcctcaacaagatgtttgcacactggccaaagacactgcctgtcgacgtgtgggtgatccacaccccccgcacccgcgtccgcgcacatggccatctacaaggcgtcacagcacaatgacggagggttgtgaggCGCgtccccccacatgagcgctgtcgatagcgatggctggcccccttcacgcattttatccgagttggaaaggaaatttcgctgtggatattggatgacagaaaacttttcgacatagtgtgtggctgtggccctatgagccgcctgaggttgcgtactgtaccaccatccactacaactacatgttaacagttcctgcatggcGCCatgaaccCGAGGcccatttcaccatcatcacactggaaagactccagttgtgtaatctactggacggacagctttgagggtgCCTgtttgtgcctgtccctggagagacccggcgcacagaggaagagaatctcccaagaaaggggagccctcaccacgagctgccccccagggagacactaagcgagcactgccccaaacaaacccagctccctcccccagccaaaggaaaccactggatggagaatatttcacccttcagatccgtgggcgtgagcgcttcgagatgttcccgagagctqaatgaggcccttggaaact (SEQUENZKENNNR. 10)

FIGUR 12

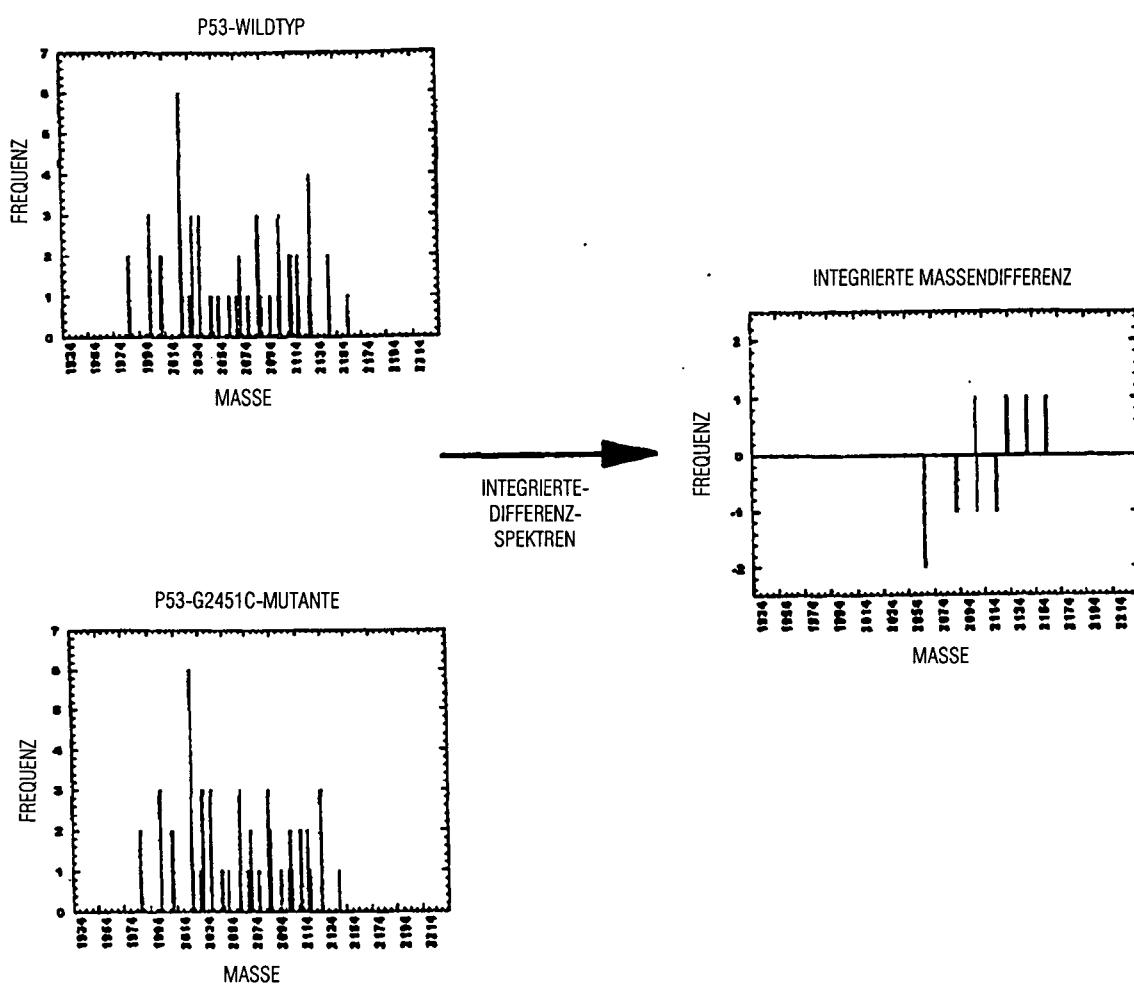
TACACATTGTCAGGACGTACCCGCGTACTTGGCTCCGGCTGGAGCTTGAGTAC (SEQUENZKENNNR. 11)

```

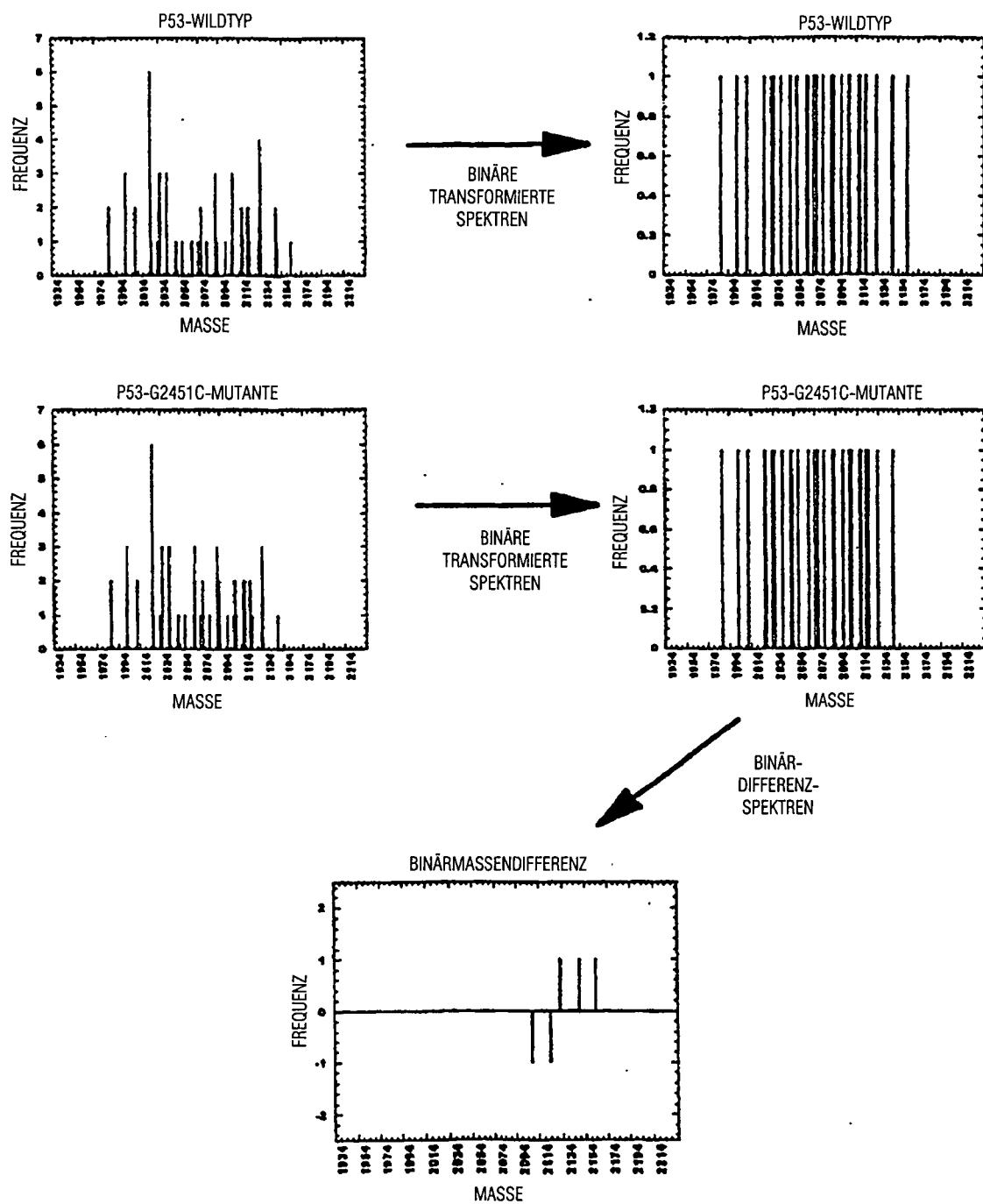
ATGTGTA
TGTGTA
GTGTAAC
TGTACA
GTAACAG
TAACAGT
AACAGTT
ACAGTTC
CAGTTCC
AGTTCCT
GTTCCCTG
TTCCTGC
TCCTGCA
CCTGCAT
CTGGATG
TGCATGG
GCATGGG
CATGGGC
ATGGGGG
TGGGGGG
GGGGGGG
GGCGGCA
GGCGGAT
CGCGATG
GGCGATG
GCATGAA
CATGAA
ATGAACC
TGAACCG
GAACCGG
AACCGGA
ACCGGAG
CCGGAGG
CGGAGG
GGAGG
GAGGGCC
AGGGCCA
GGCCCAT
GGCCATC
GGCATCC
GCATCC
CATCCG
ATCCGA
TCCTGAC
CCCTGAC
CTGACCA
TCGACAT
GCGACAT
AGCGACG
GGCGCAT
CATGATG
ATCATOR
TCATCAG
CATGACA
ATCAGAC
TCAGACT
CATGACTG

```

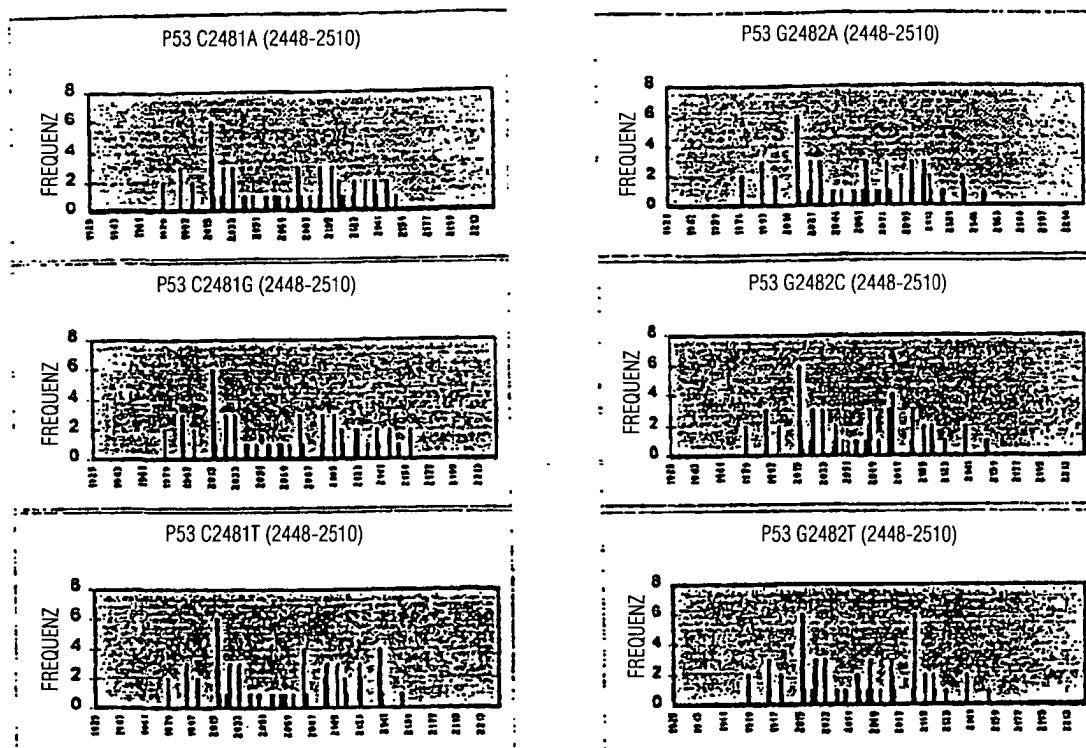
FIGUR 13



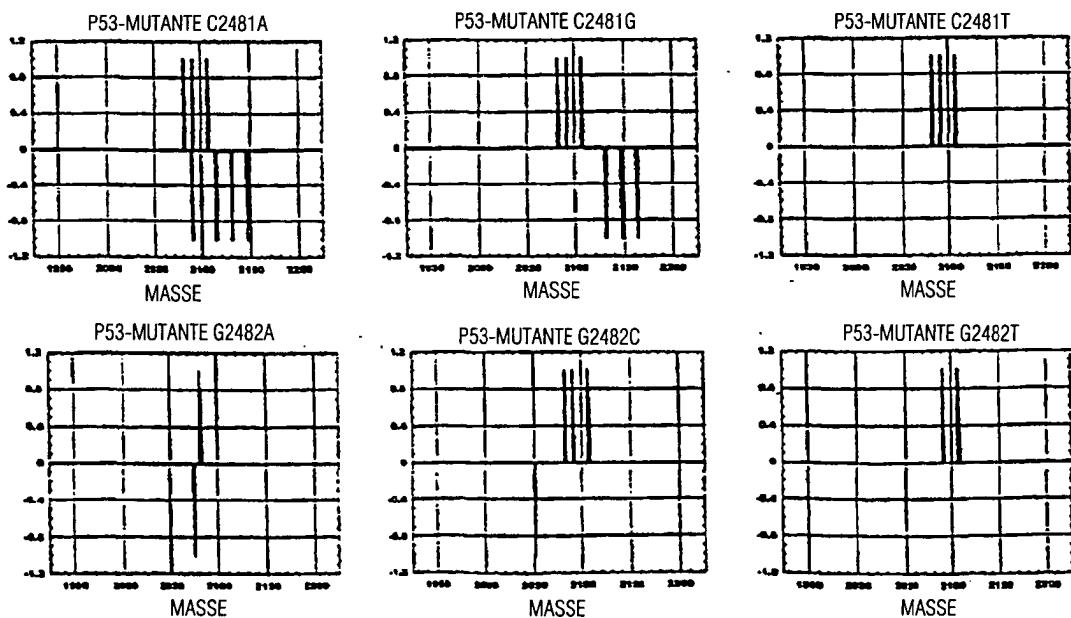
FIGUR 14



FIGUR 15

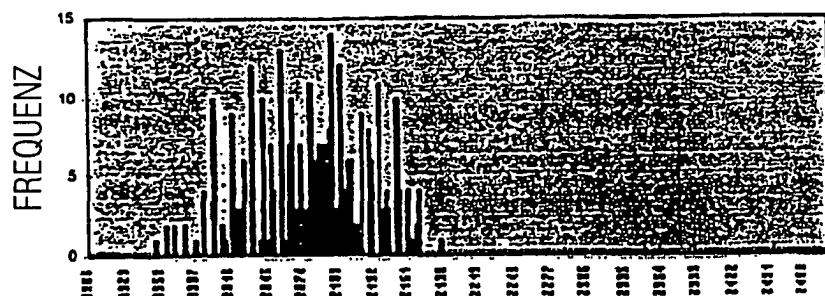


FIGUR 16

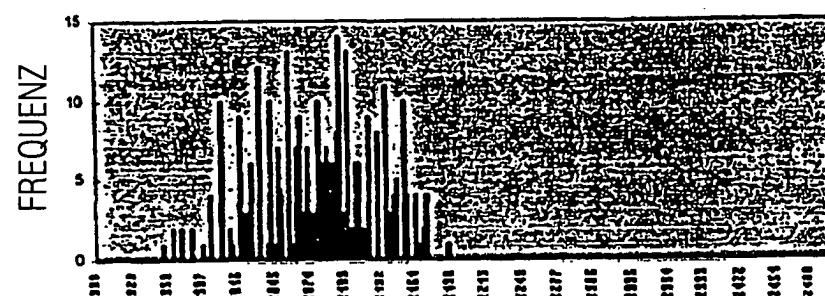


FIGUR 17

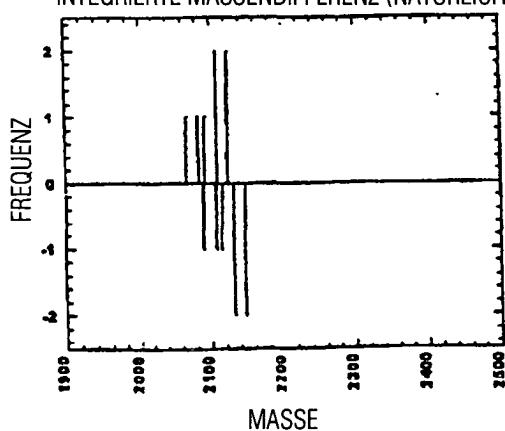
P53-WILDTYP (NATÜRLICH)



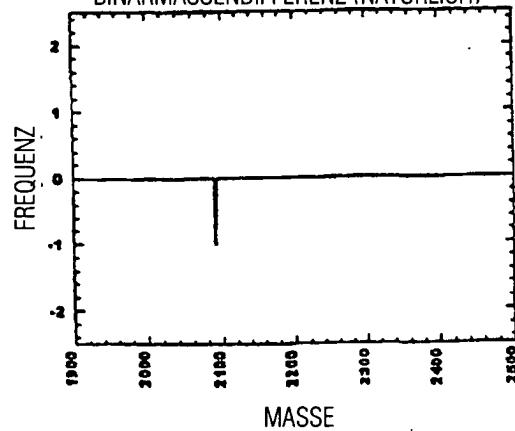
P53C2481A-MUTANTE (NATÜRLICH)



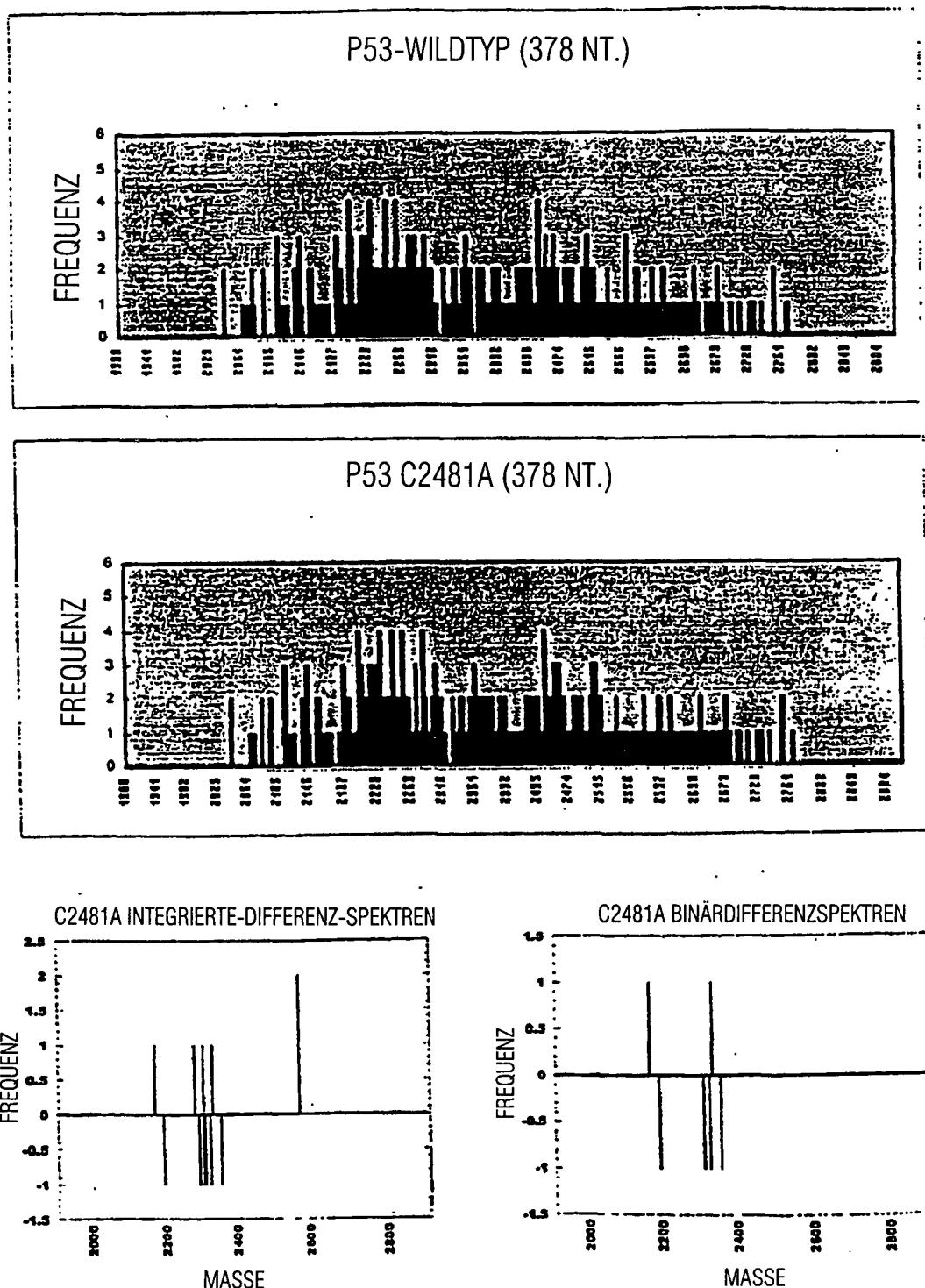
INTEGRIERTE MASSENDIFFERENZ (NATÜRLICH)



BINÄRMASSENDIFFERENZ (NATÜRLICH)



FIGUR 18



FIGUR 19

TACACATTGTCAGGGACGTACCCCGCCGACTGGCTCCGGGTAGGAGTGCTAGTACTGTGAC (SEQUENZKENNNR. 11)

```

A
TGTGTA
A
CA
CTTCCCTGCA
TGGGCGGCA
TGA
A
CCGGA
GGCGCA
TCTCTA
CCA
TCA
TCA
CA
CTG

```

```

ATG
TG
TAACAG
TTCCCTG
CATG
G
G
cG
G
CATG
AACCG
G
AG
G

```

CCCATCTCAACATCATCACACTG (SEQUENZKENNNR. 12)

```

ATGTGTAAC
AGTTC
C
TGC
ATGGGC
Ggc
ATGAAC
C
GGAGGC
C
ATC
C
TC
AC
C
ATC
ATC
AC
AC
TG

```

```

AT
GT
GT
AACACT
T
CCT
GCAT
GGCGGCGAT

```

GAACCGGAGGGCCAT (SEQUENZKENNNR. 13)

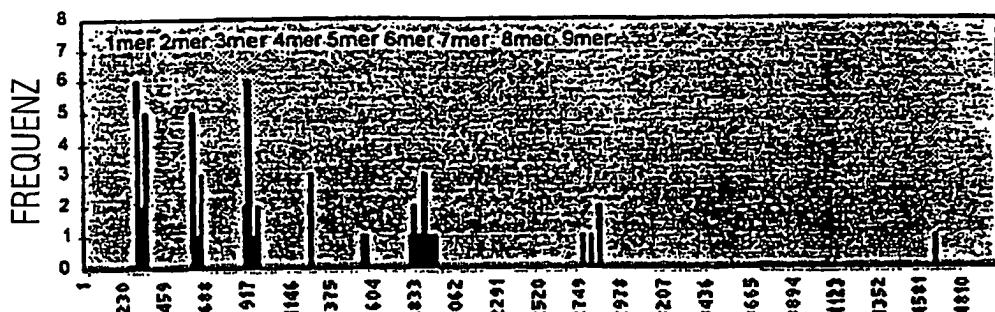
```

CCT
CACCAT
CAT
CACACT
G

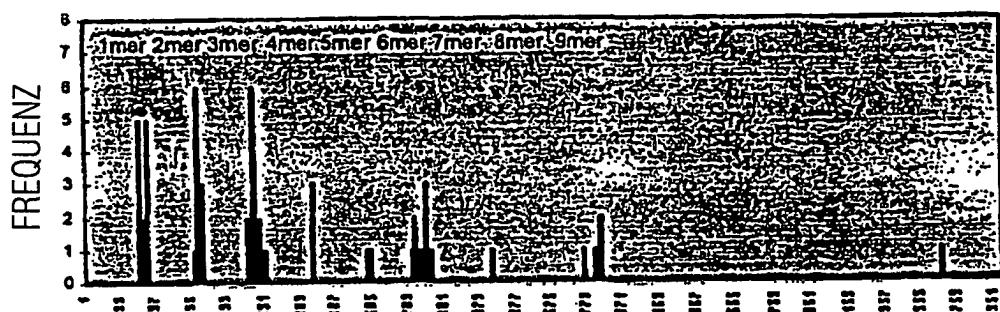
```

FIGUR 20

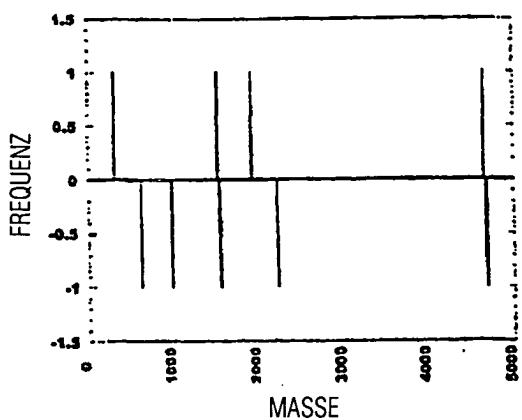
PEACA VON WILDTYP P53



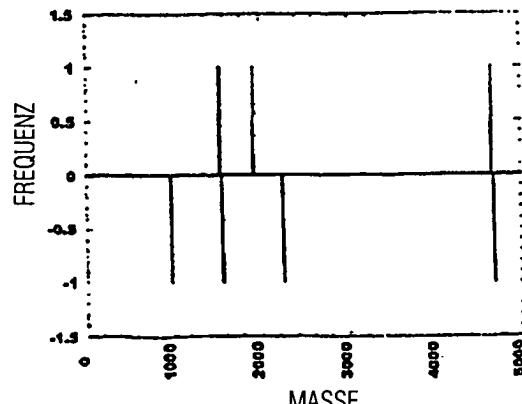
PEACA VON P53 C2481A



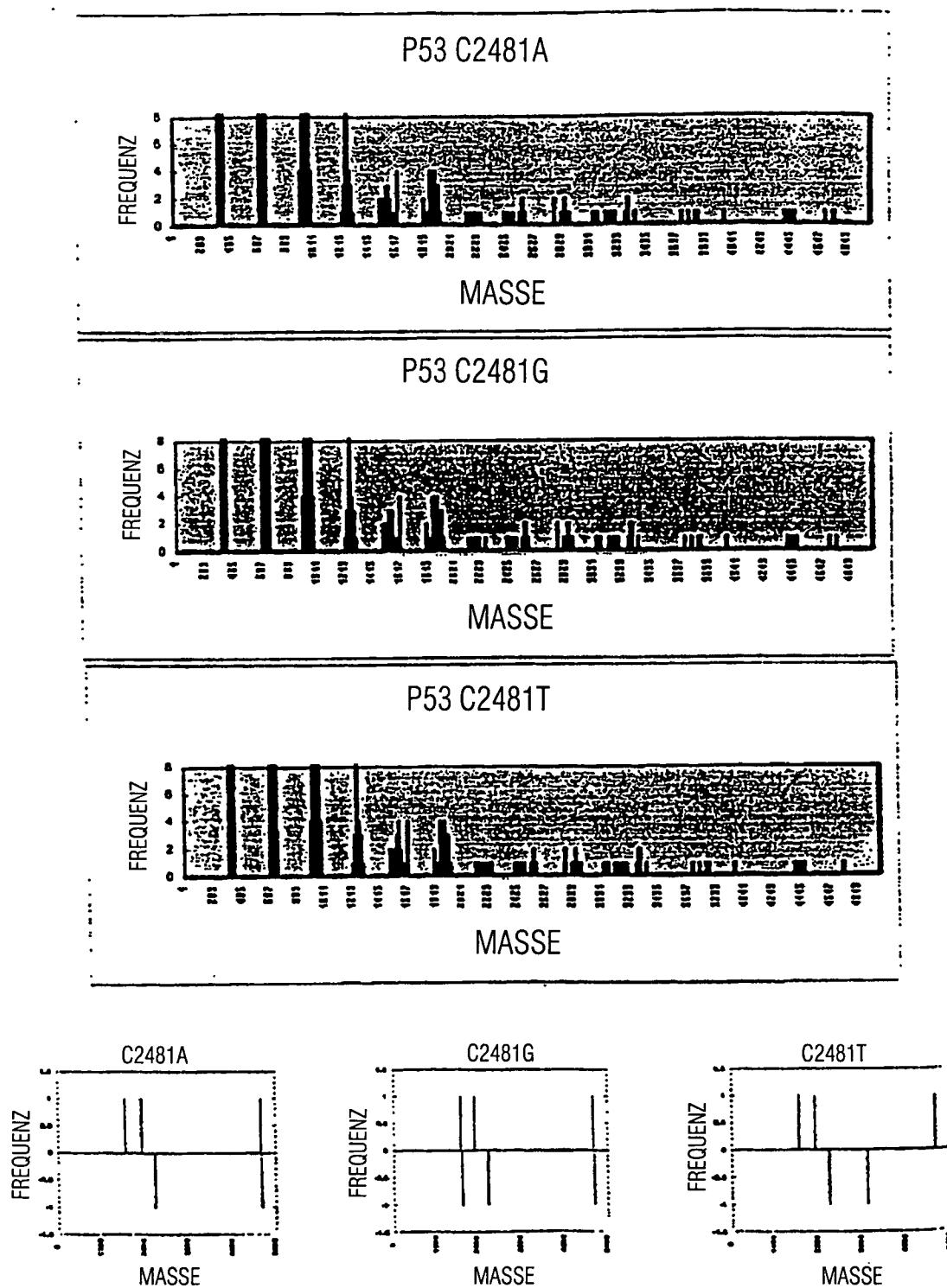
INTEGRIERTE-DIFFERENZ-SPEKTREN



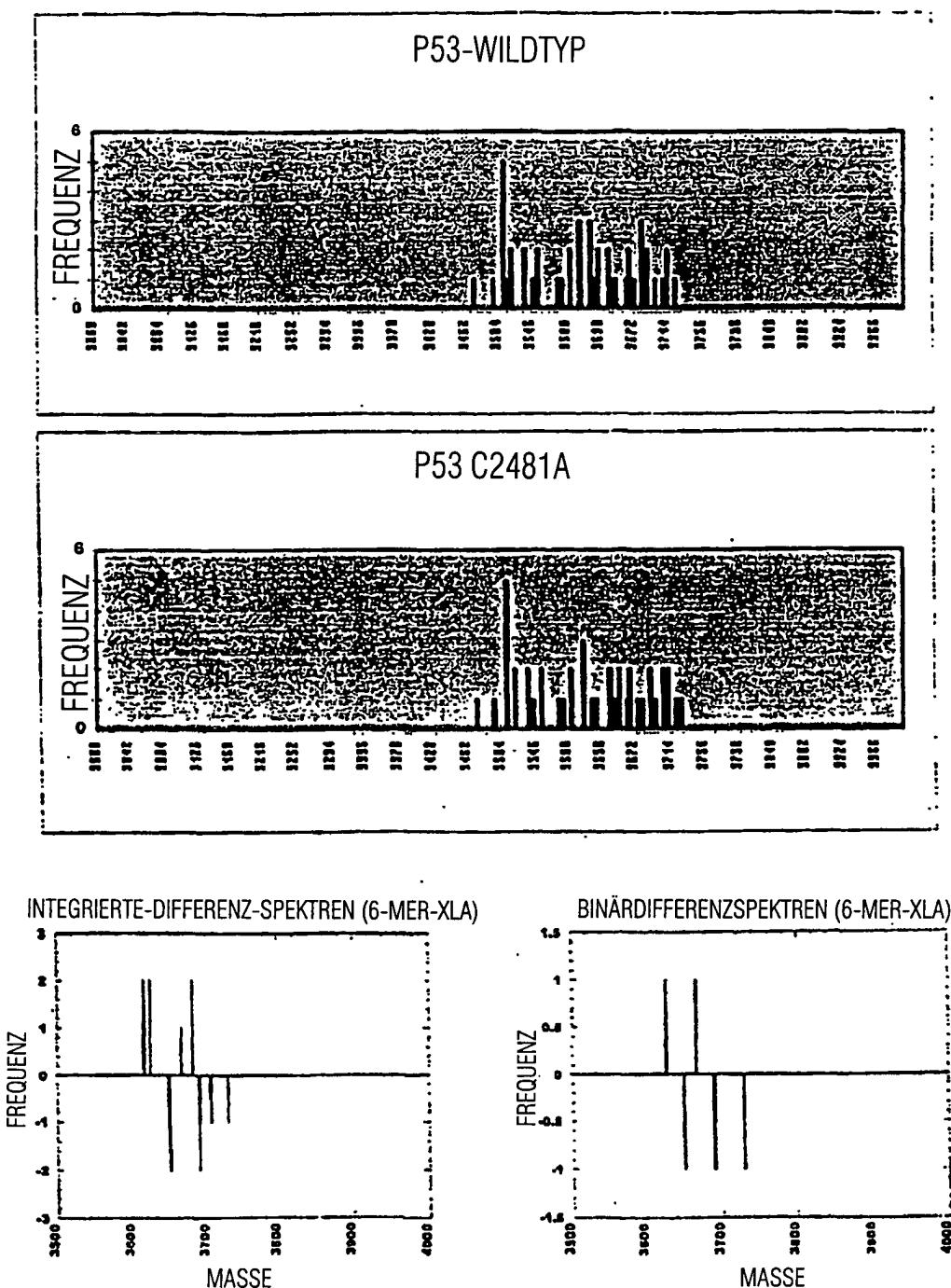
BINÄRDIFFERENZSPEKTREN



FIGUR 21



FIGUR 22



FIGUR 23

