



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 695 34 930 T2** 2007.01.11

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 793 718 B1**

(21) Deutsches Aktenzeichen: **695 34 930.9**

(86) PCT-Aktenzeichen: **PCT/US95/12791**

(96) Europäisches Aktenzeichen: **95 941 325.3**

(87) PCT-Veröffentlichungs-Nr.: **WO 1996/012014**

(86) PCT-Anmeldetag: **12.10.1995**

(87) Veröffentlichungstag

der PCT-Anmeldung: **25.04.1996**

(97) Erstveröffentlichung durch das EPA: **10.09.1997**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **12.04.2006**

(47) Veröffentlichungstag im Patentblatt: **11.01.2007**

(51) Int Cl.⁸: **C12N 15/10** (2006.01)
C12Q 1/68 (2006.01)

(30) Unionspriorität:

322348 **13.10.1994** **US**

358810 **19.12.1994** **US**

(73) Patentinhaber:

Solexa, Inc., Hayward, Calif., US

(74) Vertreter:

Vossius & Partner, 81675 München

(84) Benannte Vertragsstaaten:

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

(72) Erfinder:

**BRENNER, Univ. of Cambridge School of Med., S.,
Cambridge CB 2QQ, GB**

(54) Bezeichnung: **MOLEKULARES MARKIERUNGSSYSTEM**

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 eingelegt, 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] Die Erfindung betrifft allgemein Verfahren zur Identifizierung, Sortierung und/oder Verfolgung von Molekülen, besonders Polynucleotiden, mit Oligonucleotidmarkierungen und genauer ein Verfahren zur Sortierung von Polynucleotiden durch spezifische Hybridisierung an Oligonucleotidtags.

HINTERGRUND

[0002] Die spezifische Hybridisierung von Oligonucleotiden und ihren Analoga ist ein fundamentaler Prozess, welcher in einer breiten Vielfalt von Forschung, medizinischen und industriellen Anwendungen, einschließlich der Identifizierung von krankheitsbezogenen Polynucleotiden in diagnostischen Tests, der Durchmusterung nach Clonen neuer Zielpolynucleotide, der Identifizierung von spezifischen Polynucleotiden in Blots von Gemischen von Polynucleotiden, der Amplifikation von spezifischen Zielpolynucleotiden, der therapeutischen Blockierung von unangemessen exprimierten Genen, der DNA-Sequenzierung und ähnlichem angewendet wird, zum Beispiel Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2. Ausgabe (Cold Spring Harbor Laboratory, New York, 1989); Keller und Manak, *DNA Probes*, 2. Ausgabe (Stockton Press, New York, 1993); Milligan et al., *J. Med. Chem.* 36 (1993): 1923–1937; Drmanac et al., *Science* 260 (1993): 1649–1652; Bains J., *DNA Sequencing and Mapping*, 4 (1993): 143–150.

[0003] Die spezifische Hybridisierung wurde auch als Verfahren zur Verfolgung, Gewinnung und Identifizierung von mit Oligonucleotidtags markierten Verbindungen vorgeschlagen. Zum Beispiel werden bei der Multiplex-DNA-Sequenzierung Oligonucleotidtags verwendet, um in einem Gel, welches aus in der gleichen Sequenzierungsreaktion erzeugten DNA-Fragmenten besteht, elektrophoretisch aufgetrennte Banden zu identifizieren. In dieser Art werden DNA-Fragmente aus vielen Sequenzierungsreaktionen in der gleichen Spur eines Gels aufgetrennt, mit welchem dann mit einzelnen Festphasenmaterialien ein Blot durchgeführt wird, auf welchem die Fragmentbanden aus den einzelnen Sequenzierungsreaktionen mit Oligonucleotidsonden sichtbar gemacht werden, welche spezifisch an Tag-Komplementäre hybridisieren, Church et al., *Science* 240 (1988): 185–188. Ähnliche Verwendungen von Oligonucleotidtags wurden auch zur Identifizierung von Sprengstoffen, möglichen umweltverschmutzenden Schadstoffen, wie Rohöl, und von Geld zur Vorbeugung und dem Nachweis von Fälschungen vorgeschlagen, zum Beispiel übersichtsartig besprochen von Dollinger, Seiten 265–274 in Mullis et al., Herausgeber, *The Polymerase Chain Reaction* (Birkhauser, Boston, 1994). In jüngerer Zeit wurden Oligonucleotidtags verwendende Systeme auch als Mittel zur Manipulation und Identifizierung von Molekülen in komplexen kombinatorischen chemischen Bibliotheken, zum Beispiel als ein Hilfsmittel zur Durchmusterung solcher Bibliotheken nach Arzneimittelkandidaten vorgeschlagen, Brenner und Lerner, *Proc. Natl. Acad. Sci.* 89 (1992): 5381–5383; Alper, *Science* 264 (1994): 1399–1401; und Needels et al., *Proc. Natl. Acad. Sci.* 90 (1993): 10700–10704.

[0004] Church (EP 0303459) beschreibt ein Verfahren zur Multiplex-Sequenzierung von Polynucleotiden. In diesem Verfahren werden Oligonucleotidtags an Polynucleotide angehängt und aliquote Teile von jedem mit einem Tag versehenen Polynucleotid werden in einer konventionellen Art sequenziert, was eine Vielfalt von Fragmenten verschiedener Länge erzeugt, wobei jedes einen Tag enthält. Die mit einem Tag versehenen Fragmente werden dann typischer Weise auf einem Sequenzgel der Größe nach aufgetrennt und die aufgetrennten Fragmente auf dem Gel werden durch Hybridisierung mit komplementären, markierten Sonden untersucht. In dieser Weise kann das Polynucleotid, von welchem jedes Fragment abgeleitet ist, identifiziert werden.

[0005] Dower (WO 93/06121) beschreibt die Verwendung von aus verbundenen Nucleotiden aufgebauten, identifizierenden Tags zur Identifizierung der Sequenz von Monomeren in einem synthetisierten Oligomer, typischer Weise einem Polypeptid. Der Tag ist so entworfen, dass er in seiner Oligonucleotidsequenz die einem Monomer des synthetisierten Oligomers, welches mit der Hinzufügung des bestimmten Tags assoziiert ist, entsprechende Information enthält. Der identifizierende Oligonucleotidtag kann an das gleiche Trägerpartikel gebunden sein wie das synthetisierte Oligomer.

[0006] Sibson (WO 95/20053) beschreibt Populationen von Kügelchen, auf welchen Oligonucleotide in einem Misch- und Aufteilverfahren synthetisiert werden und zum Fangen von Polynucleotidfragmenten zur Sequenzierung verwendet werden. Die Oligonucleotide werden so synthetisiert, dass nur eine Art von Oligonucleotid auf einem gegebenen Kügelchen gefunden wird und von der Zahl der Permutationen von Basen auf den Kügelchen erwartet wird, dass sie die Sets von zu sequenzierenden Fragmenten in großem Maße zahlenmäßig übertrifft.

[0007] Die erfolgreiche Implementierung solcher Schemata zur Anwendung von Tags hängt in großen Teilen vom Erfolg beim Erreichen einer spezifischen Hybridisierung zwischen einem Tag und seiner komplementären Sonde ab. Das heißt, die Zahl von falsch positiven und falsch negativen Signalen muss minimiert werden, damit ein Oligonucleotidtag eine Substanz erfolgreich identifiziert. Unglücklicherweise sind solche scheinbaren Signale nicht selten, da die freien Energien der Basenpaarung und Basenstapelung zwischen Nucleotiden in einer Duplex- oder Triplexstruktur weitgehend variieren. Zum Beispiel kann ein aus einer wiederholten Sequenz von Desoxyadenin (A) und Thymidin (T) bestehender, an sein Komplementär gebundener Duplexstrang weniger Stabilität besitzen als ein aus einer wiederholten Sequenz von Desoxyguanidin (G) und Desoxycytidin (C) bestehender Duplexstrang gleicher Länge, welcher an ein teilweise komplementäres, eine Fehlpaarung enthaltendes Ziel gebunden ist. Wenn eine gewünschte Verbindung aus einer großen kombinatorischen chemischen Bibliothek mit dem vorgenannten Oligonucleotid markiert wäre, würde somit eine signifikante Möglichkeit bestehen, dass zusammen mit den perfekt gepaarten, aus dem AT-reichen Tag bestehenden Duplexen, ungewünschte, mit dem GC-reichen Oligonucleotid markierte Verbindungen, selbst in einem fehlgepaarten Duplex, unter zum Nachweis perfekt gepaarter AT-reicher Duplexe entworfener Hybridisierungsbedingungen nachgewiesen würden. Im durch Brenner et al. (vorstehend zitiert) vorgeschlagenen molekularen Tag-Markierungssystem wurde das verwandte Problem der Fehlhybridisierung nahe verwandter Tags durch die Verwendung eines sogenannten "kommalosen" Codes angegangen, was sicherstellt, dass eine in Bezug auf ihren komplementären Tag außerhalb des Registers befindliche (oder rasterverschobene) Sonde im Bezug auf ihren komplementären Tag einen Duplexstrang mit einer oder mehreren Fehlpaarungen für jedes seiner fünf oder mehr Dreibasenwörter oder "Codons" ergeben würde.

[0008] Obwohl Reagenzien wie Tetramethylammoniumchlorid zur Aufhebung der basenspezifischen Stabilitätsunterschiede von Oligonucleotidduplexsträngen erhältlich sind, ist die Wirkung solcher Reagenzien oft begrenzt und ihre Anwesenheit kann mit weiteren Manipulationen der ausgewählten Verbindungen, zum Beispiel der Amplifikation durch die Polymerasekettenreaktion (PCR) oder ähnliches, unverträglich sein oder diese schwieriger machen.

[0009] Solche Probleme haben die gleichzeitige Verwendung von multiplen Hybridisierungssonden in der Untersuchung von multiplen oder komplexen genetischen Loci, zum Beispiel über eine Multiplex-PCR, reversen Dot-Blot oder ähnliches, sehr schwierig gemacht. Als ein Ergebnis wurde die direkte Sequenzierung von bestimmten Loci, zum Beispiel HLA-Genen, als verlässliche Alternative zu indirekten Verfahren unter Verwendung einer spezifischen Hybridisierung zur Identifizierung von Genotypen, zum Beispiel Gyllensten et al., Proc. Natl. Acad. Sci. 85 (1988): 7652–7656, gefördert.

[0010] Die Fähigkeit clonierter und mit einem identischen Tag versehene DNA-Fragmente auf abgegrenzten Festphasenträgern zu sortieren würde solche Sequenzierungen vereinfachen, insbesondere wenn dies mit einer nicht auf einem Gel basierten Sequenzierungsverfahrensweise verbunden wäre, welche gleichzeitig auf viele Proben parallel anwendbar ist.

[0011] Angesichts des vorstehenden wäre es nützlich, wenn ein auf Oligonucleotiden basierendes Tag-Markierungssystem verfügbar wäre, welches ein großes Repertoire von Tags bereitstellte, welches aber auch ohne die Notwendigkeit der Anwendung spezieller Reagenzien zur Veränderung der Unterschiede der freien Energie der natürlichen Basenpaarung und der Basenstapelung das Auftreten von falsch positiven und falsch negativen Signalen minimierte. Solch ein Markierungssystem würde Anwendungen in vielen Bereichen, einschließlich der Erzeugung und Verwendung von kombinatorischen chemischen Bibliotheken, der Kartierung und Sequenzierung von DNA in großem Maßstab, genetischer Identifizierung, medizinischer Diagnostik und ähnlichem, finden.

Zusammenfassung der Erfindung

[0012] Ein erfindungsgemäßes Ziel ist es, ein molekulares Tag-Markierungssystem zur Verfolgung, Gewinnung und Identifizierung von Verbindungen bereitzustellen.

[0013] Ein anderes erfindungsgemäßes Ziel ist es, ein Verfahren zur Sortierung identischer Moleküle oder Subklassen von Molekülen, insbesondere Polynucleotiden, auf Oberflächen von Festphasenmaterialien durch die spezifische Hybridisierung von Oligonucleotidtags und ihren Komplementären bereitzustellen.

[0014] Ein weiteres erfindungsgemäßes Ziel ist es, eine kombinatorische chemische Bibliothek bereitzustellen, deren dazugehörigen Verbindungen durch die spezifische Hybridisierung von Oligonucleotidtags und ihren Komplementären identifiziert werden.

[0015] Noch ein weiteres erfindungsgemäßes Ziel ist es, ein System zur Markierung mit Tags und Sortierung von vielen tausend Fragmenten, insbesondere zufällig überlappenden Fragmenten, eines Zielpolynucleotids zur gleichzeitigen Untersuchung und/oder Sequenzierung bereitzustellen.

[0016] Ein anderes erfindungsgemäßes Ziel ist es, ein schnelles und verlässliches Verfahren zur Sequenzierung von Zielpolynucleotiden bereitzustellen, welche in Länge im Bereich von wenigen hundert Basenpaaren bis zu einigen Zehntausenden von Basenpaaren besitzen.

[0017] Meine Erfindung erreicht diese und andere Ziele durch Bereitstellung eines Verfahrens und von Materialien zur Verfolgung, Identifizierung und/oder Sortierung von Klassen oder Subpopulationen von Molekülen durch die Verwendung von Oligonucleotidtags. Ein erfindungsgemäßer Oligonucleotidtag besteht aus einer Vielzahl von Untereinheiten, wobei jede Untereinheit aus einem Oligonucleotid mit einer Länge von 3 bis 6 Nucleotiden besteht. Untereinheiten eines Oligonucleotidtags werden aus einem minimal kreuzhybridisierenden Set ausgewählt. In solch einem Set enthält ein aus einer Untereinheit des Sets und dem Komplementär einer jeglichen anderen Untereinheit des Sets bestehender Duplexstrang oder Triplexstrang wenigstens zwei Fehlpaarungen. Anders gesagt bildet die Untereinheit eines minimal kreuzhybridisierenden Sets im besten Fall einen Duplexstrang oder Triplexstrang, welcher zwei Fehlpaarungen mit dem Komplementär einer jeden anderen Untereinheit des gleichen Sets hat. Die Zahl der in einer bestimmten Ausführungsform verfügbaren Oligonucleotidtags hängt von der Zahl der Untereinheiten pro Tag und von der Länge der Untereinheit ab. Die Zahl ist im allgemeinen deutlich niedriger als die Zahl aller möglichen Sequenzen in der Länge des Tags, welche für einen n Nucleotide langen Tag 4^n wäre. Stärker bevorzugt sind die Untereinheiten Oligonucleotide mit einer Länge von 4 bis 5 Nucleotiden.

[0018] In einem erfindungsgemäßen Aspekt werden an einen Festphasenträger angehängte Komplementäre von Oligonucleotidtags verwendet, um Polynucleotide aus einem Gemisch von Polynucleotiden, wobei jedes einen Tag enthält, zu sortieren. In dieser Ausführungsform werden Komplementäre der Oligonucleotidtags auf der Oberfläche eines Festphasenträgers wie einem mikroskopischen Kügelchen oder einem spezifischen Ort auf einer Anordnung von Syntheseorten auf einem einzelnen Träger synthetisieren, so dass Populationen von identischen Sequenzen in spezifischen Bereichen erzeugt werden. Das heißt, die Oberfläche von jedem Träger im Fall eines Kügelchens oder jeden Bereichs im Fall einer Anordnung wird durch nur eine Art von Komplementär derivatisiert, welches eine bestimmte Sequenz hat. Die Population solcher Kügelchen oder Bereiche enthält ein Repertoire von Komplementären mit unterschiedlichen Sequenzen, wobei die Größe des Repertoires von der Zahl der Untereinheiten pro Oligonucleotidtag und der Länge der verwendeten Untereinheiten abhängt. Ähnlich umfasst jedes der zu sortierenden Polynucleotide einen Oligonucleotidtag im Repertoire, so dass identische Polynucleotide den gleichen Tag und verschiedene Polynucleotide verschiedene Tags haben. Wie nachstehend mehr im vollem Umfang erklärt wird, wird dieser Zustand durch Erhebung einer ausreichend kleinen Auswahl von mit einem Tag versehenen Polynucleotiden aus der vollen Gesamtheit von mit einem Tag versehenen Polynucleotiden erreicht. So dass, wenn Populationen von Trägern und Polynucleotiden unter Bedingungen, welche eine spezifische Hybridisierung der Oligonucleotidtags mit ihren entsprechenden Komplementären erlauben, gemischt werden, Subpopulationen von identischen Polynucleotiden auf bestimmte Kügelchen oder Bereiche sortiert werden. Die Subpopulationen von Polynucleotiden können dann auf dem Festphasenträger durch mikrobiokemische Verfahren manipuliert werden.

[0019] Im Allgemeinen umfasst das erfindungsgemäße Verfahren die folgenden Schritte: (a) Binden eines Oligonucleotidtags aus einem Repertoire von Tags an jedes Molekül in einer Population von Molekülen (i), so dass im Wesentlichen all die gleichen Moleküle oder Subpopulationen von Molekülen in der Population den gleichen Oligonucleotidtag gebunden haben und im Wesentlichen alle unterschiedlichen Moleküle oder unterschiedlichen Subpopulationen von Molekülen in der Population unterschiedliche Oligonucleotidtags gebunden haben und (ii), so dass jeder Oligonucleotidtag aus dem Repertoire eine Vielzahl von Untereinheiten umfasst und dass jede Untereinheit aus der Vielzahl aus einem Oligonucleotid besteht, welches eine Länge von drei bis sechs Nucleotiden oder drei bis sechs Basenpaaren hat, wobei die Untereinheiten aus einem minimal kreuzhybridisierenden Set ausgewählt sind; und (b) Sortieren der Moleküle oder Subpopulationen von Molekülen aus der Population durch spezifisches Hybridisieren der Oligonucleotidtags mit ihren entsprechenden Komplementären.

[0020] Ein wichtiger erfindungsgemäßer Aspekt ist die Verwendung der Oligonucleotidtags zur Sortierung von Polynucleotiden für die parallele Sequenzbestimmung. Bevorzugt wird eine solche Sequenzierung durch die folgenden Schritte ausgeführt: (a) Erzeugung einer Vielzahl von Fragmenten aus dem Zielpolynucleotid, welche das Zielpolynucleotid abdecken; (b) Binden eines Oligonucleotidtags aus einem Repertoire von Tags an jedes Fragment aus der Vielzahl (i), so dass im Wesentlichen all die gleichen Fragmente den gleichen Oligo-

nucleotidtag gebunden haben und im Wesentlichen alle verschiedenen Fragmente verschiedene Oligonucleotidtags gebunden haben und (ii) so dass jeder Oligonucleotidtag aus dem Repertoire eine Vielzahl von Untereinheiten umfasst und jede Untereinheit der Vielzahl aus einem Oligonucleotid besteht, welches eine Länge von drei bis sechs Nucleotiden oder von drei bis sechs Basenpaaren hat, wobei die Untereinheiten aus einem minimal kreuzhybridisierenden Set ausgewählt sind;

Sortieren der Fragmente durch spezifische Hybridisierung der Oligonucleotidtags mit ihren entsprechenden Komplementären; (c) Bestimmen der Nucleotidsequenz eines Teils von jedem der Fragmente aus der Vielzahl, bevorzugt durch eine wie nachstehend beschriebene Verfahrensweise der Einzelbasensequenzierung; und (d) Bestimmung der Nucleotidsequenz des Zielpolynucleotids durch Zuordnen der Sequenzen der Fragmente.

[0021] Meine Erfindung überwindet eine entscheidende Unzulänglichkeit derzeitiger Verfahren zur Tag-Markierung oder Markierung von Molekülen mit Oligonucleotiden: Durch Codierung der Sequenzen der Tags in Übereinstimmung mit der Erfindung ist die Stabilität von jedem fehlgepaarten Duplexstrang oder Triplexstrang zwischen einem Tag und einem Komplementär zu einem anderen Tag viel niedriger als jene eines jeden perfekt zusammenpassenden Duplexstranges zwischen dem Tag und seinem eigenen Komplementär. Somit ist das Problem des falschen Sortierens aufgrund der Tatsache, dass fehlgepaarte Duplexstränge von GC-reichen Tags stabiler als perfekt zusammenpassende AT-reiche Tags sind, eliminiert.

[0022] Wenn sie in Kombination mit Festphasenträgern wie mikroskopischen Kügelchen verwendet wird, stellt meine Erfindung ein einfach automatisierbares System zur Manipulation und Sortierung von Polynucleotiden bereit, welches besonders nützlich ist bei parallelen Verfahren in großem Maßstab wie DNA-Sequenzierung im großen Maßstab, worin viele Zielpolynucleotide oder viele Abschnitte eines einzelnen Zielpolynucleotids gleichzeitig sequenziert und/oder untersucht werden.

Kurze Beschreibung der Zeichnungen

[0023] Die [Fig. 1a-Fig. 1c](#) stellen Strukturen von markierten Sonden dar, welche in einem bevorzugten Verfahren der "Einzelbasen"-Sequenzierung verwendet werden, welches mit der Erfindung verwendet werden kann.

[0024] Die [Fig. 2](#) stellt die relative Position der Nucleaseerkennungsstelle, der Ligierungsstelle und der Spaltstelle in einem ligierten Komplex, welcher zwischen einem Zielpolynucleotid und einer in einem bevorzugten "Einzelbasen"-Sequenzierungsverfahren verwendeten Sonde gebildet wird.

[0025] Die [Fig. 3](#) ist ein Flussdiagramm, welches den allgemeinen Algorithmus zur Erzeugung minimal kreuzhybridisierender Sets darstellt.

[0026] Die [Fig. 4](#) stellt ein Schema für die Synthese und die Verwendung einer kombinatorischen chemischen Bibliothek dar, in welcher Mitglied-Verbindungen mit Oligonucleotidtags in Übereinstimmung mit der Erfindung markiert sind.

[0027] Die [Fig. 5](#) stellt in einem Diagramm einen Apparat zur Durchführung paralleler Verfahren wie der Polynucleotidsequenzierung in Übereinstimmung mit der Erfindung dar.

Definitionen

[0028] "Komplementär" oder "Tag-Komplementär" bezieht sich wie hier im Bezug auf Oligonucleotidtags verwendet auf ein Oligonucleotid, an welches ein Oligonucleotidtag spezifisch hybridisiert, um einen perfekt zusammenpassenden Duplexstrang oder Triplexstrang zu bilden. In Ausführungsformen, wo die spezifische Hybridisierung einen Triplexstrang ergibt, kann der Oligonucleotidtag so ausgewählt werden, dass er entweder doppelsträngig oder einzelsträngig ist. Somit soll der Begriff "Komplementär" wo Triplexstränge gebildet werden entweder ein doppelsträngiges Komplementär eines einzelsträngigen Oligonucleotidtags oder ein einzelsträngiges Komplementär eines doppelsträngigen Oligonucleotidtags umfassen.

[0029] Der Begriff "Oligonucleotid" schließt wie hier verwendet lineare Oligomere von natürlichen oder modifizierten Monomeren oder Verknüpfungen, einschließlich Desoxyribonucleosiden, Ribonucleosiden, α -anomeren Formen davon, Peptid-Nucleinsäuren (PNAs) und ähnliches ein, welche fähig sind, an ein Zielpolynucleotid auf dem Wege eines regelmäßigen Musters von Monomer-zu-Monomer-Interaktionen wie dem Watson-Crick-Typ der Basenpaarung, Basenstapelung, Hoogsten- oder revers-Hoogsten-Typen der Basenpaarung und ähnliches spezifisch zu binden. Für gewöhnlich sind Monomere über Phosphodiesterbindungen oder

Analoge davon verknüpft, um Oligonucleotide zu bilden, welche in der Größe im Bereich von wenigen Monomereinheiten zum Beispiel 3–4 bis zu einigen zehn Monomereinheiten liegen. Wann immer ein Oligonucleotid durch eine Sequenz von Buchstaben wie "ATGCCTG" dargestellt wird, wird es verstanden werden, dass die Nucleotide in einer 5' → 3'-Anordnung von links nach rechts sind und dass "A" Desoxyadenosin bedeutet, "C" Desoxycytidin bedeutet, "G" Desoxyguanosin bedeutet und "T" Thymidin bedeutet, sofern es nicht anders angegeben ist. Analoga von Phosphodiesterbindungen schließen Phosphorothioate, Phosphorodithioate, Phosphoroanilidate, Phosphoramidate und ähnliches ein. Für gewöhnlich umfassen erfindungsgemäße Oligonucleotide die vier natürlichen Nucleotide, sie können jedoch auch nicht-natürliche Nucleotidanaloge umfassen. Dem Fachmann ist klar, wann Oligonucleotide mit natürlichen oder nicht-natürlichen Nucleotiden verwendet werden können, zum Beispiel werden, wo eine Umsetzung durch Enzyme benötigt wird, für gewöhnlich Oligonucleotide, welche aus natürlichen Nucleotiden bestehen, benötigt.

[0030] "Perfekt zusammenpassend" bedeutet im Bezug auf einen Duplexstrang, dass die Poly- oder Oligonucleotidstränge, welche den Duplexstrang bilden, miteinander eine doppelsträngige Struktur bilden, so dass jedes Nucleotid in jedem Strang eine Watson-Crick-Basenpaarung mit einem Nucleotid in dem anderen Strang eingeht. Der Begriff schließt auch die Paarung von Nucleosidanaloge wie Desoxyinosin, Nucleosiden mit 2-Aminopurinbasen und ähnliches ein, welche verwendet werden können. Im Bezug auf einen Triplexstrang bedeutet der Begriff, dass der Triplexstrang aus einem perfekt zusammenpassenden Duplexstrang und einem dritten Strang besteht, in welchem jedes Nucleotid eine Hoogsteen- oder reverse Hoogsteen-Assoziation mit einem Basenpaar des perfekt zusammenpassenden Duplexstranges eingeht. Umgekehrt bedeutet eine "Fehlpaarung" in einem Duplexstrang zwischen einem Tag und einem Oligonucleotid, dass ein Paar oder Triplet von Nucleotiden in dem Duplexstrang oder Triplexstrang keine Watson-Crick- und/oder Hoogsteen- und/oder reverse Hoogsteen-Bindung eingeht.

[0031] Wie hier verwendet schließt "Nucleosid" die natürlichen Nucleoside, einschließlich 2'-Desoxy- und 2'-Hydroxyl-Formen, wie zum Beispiel in Kornberg und Baker, DNA Replication, 2. Ausgabe (Freeman, San Francisco, 1992) beschrieben ein. "Analoga" schließt im Bezug auf Nucleoside synthetische Nucleoside, welche modifizierte Baseneinheiten und/oder modifizierte Zuckereinheiten besitzen, zum Beispiel durch Scheit, Nucleotide Analogs (John Wiley, New York, 1980); Uhlman und Peyman, Chemical Reviews, 90 (1990): 543–584; oder ähnliches beschrieben, nur mit der Voraussetzung ein, dass sie zur spezifischen Hybridisierung fähig sind. Solche Analoga schließen synthetische Nucleoside ein, welche entworfen sind, um Bindungseigenschaften zu verbessern, die Degeneriertheit zu reduzieren, die Spezifität zu erhöhen und ähnliches.

Detaillierte Beschreibung der Erfindung

[0032] Die Erfindung stellt ein Verfahren zur Markierung und Sortierung von Molekülen, insbesondere Polynucleotiden, durch die Verwendung von Oligonucleotidtags bereit. Die erfindungsgemäßen Oligonucleotidtags umfassen eine Vielzahl von "Wörtern" oder Untereinheiten, welche aus minimal kreuzhybridisierenden Sets von Untereinheiten ausgewählt sind. Untereinheiten aus solchen Sets können keine Duplexstränge oder Triplexstränge mit dem Komplementär einer anderen Untereinheit des gleichen Sets mit weniger als zwei fehlgepaarten Nucleotiden bilden. Somit werden die Sequenzen von jeglichen zwei Oligonucleotidtags eines Repertoires, welche Duplexstränge bilden, niemals "näher" sein als ein Unterschied von zwei Nucleotiden. In bestimmten Ausführungsformen könne Sequenzen von jeglichen zwei Oligonucleotidtags eines Repertoires sogar "weiter" entfernt sein, zum Beispiel durch Entwerfen eines minimal kreuzhybridisierenden Sets, so dass Untereinheiten keinen Duplexstrang mit dem Komplementär einer anderen Untereinheit des gleichen Sets mit weniger als drei fehlgepaarten Nucleotiden bilden können und so weiter. In solchen Ausführungsformen wird eine größere Spezifität erreicht, aber das gesamte Repertoire von Tags ist kleiner. Somit muss für Tags einer gegebenen Länge und Wortgröße ein Kompromiss zwischen dem gewünschten Grad der Spezifität und der gewünschten Größe des Repertoires gemacht werden. Die Erfindung ist besonders nützlich zur Markierung und Sortierung von Polynucleotiden für parallele Verfahren wie Sequenzierung, Fingerprinting und andere Arten von Untersuchungen.

Konstruktion von Oligonucleotidtags aus minimal kreuzhybridisierenden Sets von Untereinheiten

[0033] Die Nucleotidsequenzen der Untereinheiten für jegliches minimal kreuzhybridisierende Set werden geeigneter Weise durch einfache Computerprogramme spezifiziert, welche dem in [Fig. 3](#) dargestellten, allgemeinen Algorithmus folgen und durch das Programm Minhx veranschaulicht werden, dessen Quellcode in Anhang I aufgeführt ist. Minhx berechnet alle minimal kreuzhybridisierenden Sets, die Untereinheiten haben, die aus drei Arten von Nucleotiden zusammengesetzt sind und eine Länge von vier haben.

[0034] Der Algorithmus der [Fig. 3](#) wird implementiert indem zuerst die Charakteristik der Untereinheiten des minimal kreuzhybridisierenden Sets, das heißt, Länge, Zahl der Basenunterschiede zwischen den Mitgliedern und Zusammensetzung, zum Beispiel bestehen sie aus zwei, drei oder vier Arten von Basen, definiert wird. Eine Tabelle M_n , $n = 1$, wird erzeugt (**100**), welche aus allen möglichen Sequenzen einer gegebenen Länge und Zusammensetzung besteht. Eine anfängliche Untereinheit S_1 wird ausgewählt und mit nachfolgenden Untereinheiten S_i für $i = n + 1$ bis zum Ende der Tabelle verglichen (**120**). Wann immer eine nachfolgende Untereinheit die geforderte Zahl von Fehlpaarungen besitzt, um ein Mitglied des minimal kreuzhybridisierenden Sets zu sein, wird sie in einer neuen Tabelle M_{n+1} (**125**), welche ebenfalls vorher in früheren Durchgängen durch Schritt **120** ausgewählte Untereinheiten enthält, gespeichert. Zum Beispiel wird im ersten Set von Vergleichen $M_2 S_1$ enthalten, im zweiten Set von Vergleichen wird $M_3 S_1$ und S_2 enthalten; im dritten Set von Vergleichen wird $M_4 S_1, S_2$ und S_3 enthalten und so weiter. Ähnlich werden Vergleiche in Tabelle M_i zwischen S_i und allen nachfolgenden Untereinheiten in M_i stattfinden. Beachte, dass jede nachfolgende Tabelle M_{n+1} kleiner als ihre Vorgänger sein wird, da in nachfolgenden Durchgängen durch Schritt **130** Untereinheiten eliminiert werden. Nachdem jede Untereinheit von Tabelle M_n verglichen worden ist (**140**) wird die alte Tabelle durch die neue Tabelle M_{n+1} ersetzt und die nächste Runde von Vergleichen wird begonnen. Der Vorgang stoppt (**160**) wenn eine Tabelle M_n erreicht wird, welche keine nachfolgenden Untereinheiten zum Vergleich mit der ausgewählten Untereinheit S_i enthält, das heißt $M_n = M_{n+1}$.

[0035] Bevorzugt umfassen minimal kreuzhybridisierende Sets Untereinheiten, welche annähernd gleichwertige Beiträge zu Duplexstrangstabilität leisten wie jede andere Untereinheit in dem Set. In dieser Weise ist die Stabilität von perfekt zusammenpassenden Duplexsträngen zwischen jeder Untereinheit und ihrem Komplementär annähernd gleich. Eine Anleitung zur Auswahl solcher Sets wird durch veröffentlichte Verfahren zur Auswahl von PCR-Primern und der Berechnung der Stabilitäten von Duplexsträngen, zum Beispiel Rychlik et al., Nucleic Acids Research, 17 (1989): 8543–8551 und 18 (1990): 6409–6412; Breslauer et al., Proc. Natl. Acad. Sci. 83 (1986): 3746–3750; Wetmur, Crit. Rev. Biochem. Mol. Biol. 26 (1991): 227–259, und ähnliches bereitgestellt. Für kürzere Tags, zum Beispiel etwa 30 Nucleotide oder weniger, wird der durch Rychlik und Wetmur beschriebene Algorithmus bevorzugt und für längere Tags, zum Beispiel etwa 30–35 Nucleotide oder länger, kann ein durch Suggs et al., Seiten 683–693 in Brown, Herausgeber, ICN-UCLA Symp. Dev. Biol., Band 23 (Academic Press, New York, 1981) offener Algorithmus in geeigneter Weise verwendet werden. Es gibt natürlich viele Ansätze zum Entwurf von minimal kreuzhybridisierenden Sets, welche im Umfang der Erfindung dem Fachmann verfügbar sind. Zum Beispiel können zur Minimierung der Auswirkungen von verschiedenen Energien der Basenstapelung von endständigen Nucleotiden beim Zusammenbau von Untereinheiten Untereinheiten bereitgestellt werden, welche die gleichen endständigen Nucleotide haben. In dieser Weise werden die Summen der Energien der Basenstapelung von allen benachbarten endständigen Nucleotiden beim Verbinden der Untereinheiten gleich sein, wodurch die Variabilität der Schmelztemperaturen der Tags reduziert oder ausgemerzt werden.

[0036] Eine bevorzugte Ausführungsform von minimal kreuzhybridisierenden Sets sind jene, deren Untereinheiten aus drei der vier natürlichen Nucleotiden aufgebaut sind. Wie nachstehend ausführlicher diskutiert werden wird, ermöglicht die Abwesenheit einer Art von Nucleotid in den Oligonucleotidtags Zielpolynucleotide durch die Verwendung der 3'→5'-Exonucleaseaktivität einer DNA-Polymerase auf Festphasenträger zu laden. Das Nachfolgende ist ein beispielhaftes minimal kreuzhybridisierendes Set von Untereinheiten, wobei jede Untereinheit vier Nucleotide ausgewählt aus der Gruppe bestehend aus A, G und T umfasst:

Tabelle I

Wort:	w_1	w_2	w_3	w_4
Sequenz:	GATT	TGAT	TAGA	TTTG
Wort:	w_5	w_6	w_7	w_8
Sequenz:	GTAA	AGTA	ATGT	AAAG

[0037] In diesem Set würde jedes Mitglied mit dem Komplementär jedes anderen Mitglieds einen Duplexstrang bilden, welcher drei fehlgepaarte Basen hätte.

[0038] Weitere beispielhafte minimal kreuzhybridisierende Sets werden nachstehend in Tabelle II aufgeführt. Es können natürlich durch den Austausch verschiedener Gruppen von Nucleotiden oder durch die Verwen-

derung von Subsets bekannter minimal kreuzhybridisierender Sets zusätzliche Sets erzeugt werden.

Tabelle II

Beispielhafte minimal kreuzhybridisierende Sets von 4-mer-Untereinheiten

<u>Set 1</u>	<u>Set 2</u>	<u>Set 3</u>	<u>Set 4</u>	<u>Set 5</u>	<u>Set 6</u>
CATT	ACCC	AAAC	AAAG	AACA	AACG
CTAA	AGGG	ACCA	ACCA	ACAC	ACAA
TCAT	CACG	AGGG	AGGC	AGGG	AGGC
ACTA	CCGA	CACG	CACC	CAAG	CAAC
TACA	CGAC	CCGC	CCGG	CCGC	CCGG
TTTC	GAGC	CGAA	CGAA	CGCA	CGCA
ATCT	GCAG	GAGA	GAGA	GAGA	GAGA
AAAC	GGCA	GCAG	GCAC	GCCG	GCCC
	AAAA	GGCC	GGCG	GGAC	GGAG
<u>Set 7</u>	<u>Set 8</u>	<u>Set 9</u>	<u>Set 10</u>	<u>Set 11</u>	<u>Set 12</u>
AAGA	AAGC	AAGG	ACAG	ACCG	ACGA
ACAC	ACAA	ACAA	AACA	AAAA	AAAC
AGCG	AGCG	AGCC	AGGC	AGGC	AGCG
CAAG	CAAG	CAAC	CAAC	CACC	CACA
CCCA	CCCC	CCCG	CCGA	CCGA	CCAG
CGGC	CGGA	CGGA	CGCG	CGAG	CGGC
GACC	GACA	GACA	GAGG	GAGG	GAGG
GCGG	GCGG	GCGC	GCCC	GCAC	GCCC
GGAA	GGAC	GGAG	GGAA	GGCA	GGAA

[0039] Die erfindungsgemäßen Oligonucleotidtags und ihre Komplementäre werden geeigneterweise durch einen automatisierten DNA-Syntheseapparat, zum Beispiel einen Applied Biosystems, Inc. (Foster City, California) Modell 392 oder 394 DNA/RNA Synthesizer, unter Verwendung von Standardchemien, wie Phosphoramiditchemie, zum Beispiel offenbart in den nachfolgenden Quellen: Beaucage und Iyer, Tetrahedron, 48 (1992): 2223–2311; Molko et al., U.S. Pat. No. 4,980,460; Koster et al., U.S. Pat. No. 4,725,677; Caruthers et al., U.S. Pat. No. 4,415,732; 4,458,066; und 4,973,679; und ähnliches, synthetisiert. Alternative Chemien, welche zum Beispiel nicht-natürliche Grundgerüst-Gruppen wie Phosphorothioate, Phosphoramidate und ähnliches ergeben, können auch verwendet werden, vorausgesetzt dass die sich ergebenden Oligonucleotide zur spezifischen Hybridisierung fähig sind. In einigen Ausführungsformen können Tags natürlich vorkommende Nucleotide umfassen, welche die Umsetzung oder Manipulation durch Enzyme erlauben, während die entsprechenden Tag-Komplementäre nicht-natürliche Nucleotidanaloga, wie Peptidnucleinsäuren oder ähnliche Verbindungen umfassen, welche die Bildung von stabileren Duplexsträngen während des Sortierens fördern.

[0040] Wenn Mikropartikel als Träger verwendet werden, werden Repertoires von Oligonucleotidtags und Tag-Komplementären bevorzugt durch eine auf Untereinheiten bezogenen Synthese über "Misch- und Aufteilverfahren" erzeugt, wie zum Beispiel in Shortle et al., Internationale Patentanmeldung EP 0638089 offenbart. Kurz gesagt ist die grundlegende Einheit der Synthese eine Untereinheit des Oligonucleotidtags. Bevorzugt wird eine Phosphoramiditchemie verwendet und es werden für jede Untereinheit in einem minimal kreuzhybridisierenden Set 3'-Phosphoramidit-Oligonucleotide hergestellt, zum Beispiel gäbe es für das vorstehend als erstes aufgeführte Set acht 4-mer 3'-Phosphoramidite. Die Synthese erfolgt wie durch Shortle et al. offenbart oder in direkter Analogie mit den Verfahren, welche verwendet werden, um diversifizierte Oligonucleotidbibliotheken unter Verwendung von Nucleosidmonomeren zu erzeugen, zum Beispiel wie in Telenius et al., Genomics 13 (1992): 718–725; Welsh et al., Nucleic Acids Research 19 (1991): 5275–5279; Grothues et al., Nucleic Acids Research 21 (1993): 1321–1322; Hartley, Europäische Patentanmeldung 90304496.4; Lam et al., Nature 354 (1991): 82–84; Zuckerman et al., Int. J. Pept. Protein Research 40 (1992): 498–507; und ähnlichem offenbart. Im Allgemeinen erfordern diese Verfahren während der Kopplungsschritte einfach nur die Anwendung von Gemischen der aktivierten Monomere an den wachsenden Oligonucleotiden.

[0041] Doppelsträngige Formen von Tags können durch getrennte Synthese der komplementären Stränge, gefolgt durch Mischen unter Bedingungen, welche die Bildung von Duplexsträngen erlauben, hergestellt wer-

den. Alternativ können doppelsträngige Tags zuerst durch synthetisieren eines einzelsträngigen Repertoires, welches an eine bekannte Oligonucleotidsequenz gebunden ist, die als Primerbindungsstelle dient, erzeugt werden. Der zweite Strang wird dann durch Kombinieren des einzelsträngigen Repertoires mit einem Primer und dem Verlängern mit einer Polymerase synthetisiert. Dieser letzte Ansatz wird in Oliphant et al., *Gene* 44 (1986): 177–183 beschrieben. Solche Duplexstrangtags können dann zum Sortieren und Manipulieren des Zielpolynucleotids in Übereinstimmung mit der Erfindung zusammen mit Zielpolynucleotiden in Clonierungsvektoren eingefügt werden.

[0042] In Ausführungsformen wo die spezifische Hybridisierung über die Bildung eines Triplexstranges erfolgt, folgt die Codierung von Tag-Sequenzen den gleichen Prinzipien wie für die Tags, welche Duplexstränge bilden; es gibt jedoch weitere Einschränkungen für die Auswahl der Sequenzen der Untereinheiten. Im Allgemeinen ist die Assoziation des dritten Stranges über den Hoogsteen-Typ der Bindung am stabilsten bei Homopyrimidin-Homopurin-Abschnitten in einem doppelsträngigen Ziel. Für gewöhnlich bilden sich Basentriplets in T-A*T- oder C-G*C-Motiven (wobei „-“ eine Watson-Crick-Paarung anzeigt und „*“ einen Hoogsteen-Typ der Bindung anzeigt); es sind jedoch auch andere Motive möglich. Zum Beispiel erlaubt die Hoogsteen-Basenpaarung parallele und antiparallele Orientierungen zwischen dem dritten Strang (dem Hoogsteen-Strang) und dem Purin-reichen Strang des Duplexstranges, an welchen der dritte Strang abhängig von den Bedingungen und der Zusammensetzung der Stränge bindet. Es gibt umfassende Anleitung in der Literatur zur Auswahl geeigneter Sequenzen, Orientierung, Bedingungen, Nucleosidart (zum Beispiel ob Ribose- oder Desoxyribosenucleoside verwendet werden), Basenmodifikationen (zum Beispiel methyliertes Cytosin und ähnliches), um die Stabilität des Triplexstranges zu maximieren oder anderweitig zu regulieren, wie dies in bestimmten Ausführungsformen erwünscht ist, zum Beispiel Roberts et al., *Proc. Natl. Acad. Sci.* 88 (1991): 9397–9401; Roberts et al., *Science* 258 (1992): 1463–1466; Distefano et al., *Proc. Natl. Acad. Sci.* 90 (1993): 1179–1183; Mergny et al., *Biochemistry* 30 (1991): 9791–9798; Cheng et al., *J. Am. Chem. Soc.* 114 (1992): 4465–4474; Beal und Dervan, *Nucleic Acids Research* 20 (1992): 2773–2776; Beal und Dervan, *J. Am. Chem. Soc.* 114 (1992): 4976–4982; Giovannangeli et al., *Proc. Natl. Acad. Sci.* 89 (1992): 8631–8635; Moser und Dervan, *Science* 238 (1987): 645–650; McShan et al., *J. Biol. Chem.* 267 (1992): 5712–5721; Yoon et al., *Proc. Natl. Acad. Sci.* 89 (1992): 3840–3844; Blume et al., *Nucleic Acids Research* 20 (1992): 1777–1784; Thuong und Helene, *Angew. Chem. Int. Ed. Engl.* 32 (1993): 666–690; und ähnliches. Bedingungen für das Aneinanderlagern einzelsträngiger oder Duplexstrangtags an ihre einzelsträngigen oder Duplexstrangkomplementäre sind wohl bekannt, zum Beispiel Ji et al., *Anal. Chem.* 65 (1993): 1323–1328.

[0043] Erfindungsgemäße Oligonucleotidtags können bei der Länge in einem Bereich von 12 bis zu 60 Nucleotiden oder Basenpaaren liegen. Bevorzugt liegen Oligonucleotidtags bei der Länge in einem Bereich von 18 bis zu 40 Nucleotiden oder Basenpaaren. Stärker bevorzugt liegt die Länge der Oligonucleotidtags in einem Bereich von 25 bis 40 Nucleotiden oder Basenpaaren. In Begriffen von bevorzugten oder stärker bevorzugten Zahlen von Untereinheiten könne diese Bereiche wie folgt ausgedrückt werden:

Tabelle III

Anzahl von Untereinheiten in Tags in bevorzugten Ausführungsformen

<u>Monomere in</u> <u>Untereinheiten</u>	<u>Nucleotide im Oligonucleotidtag</u>		
	(12 - 60)	(18 - 40)	(25 - 40)
3	4 - 20 Untereinheiten	6 - 13 Untereinheiten	8 - 13 Untereinheiten
4	3 - 15 Untereinheiten	4 - 10 Untereinheiten	6 - 10 Untereinheiten
5	2 - 12 Untereinheiten	3 - 8 Untereinheiten	5 - 8 Untereinheiten
6	2 - 10 Untereinheiten	3 - 6 Untereinheiten	4 - 6 Untereinheiten

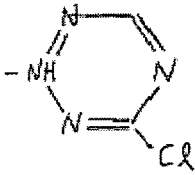
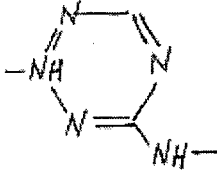
[0044] Am stärksten bevorzugt sind Oligonucleotidtags einzelsträngig und die spezifische Hybridisierung erfolgt durch eine Watson-Crick-Paarung mit einem Tag-Komplementär.

Anhängen von Tags an Moleküle

[0045] Oligonucleotidtags können an viele verschiedene Klassen von Molekülen durch eine Vielfalt von im Fachgebiet wohl bekannten reaktiven Funktionalitäten angehängt werden, zum Beispiel Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Inc., Eugene, 1992); Khanna et al., U.S. Pat. No. 4,318,846; oder ähnliches. Die Tabelle IV stellt beispielhafte Funktionalitäten und reaktive Gruppen als Gegenstück bereit, welche auf den Oligonucleotidtags oder den interessierenden Molekülen sitzen können. Wenn die Funktionalitäten und die Reaktanten als Gegenstück, in einigen Fällen nach Aktivierung, miteinander umgesetzt werden, wird eine verknüpfende Bindungsgruppe gebildet. Mehr noch können, wie nachstehend ausführlicher beschrieben, Tags gleichzeitig synthetisiert werden, während die Moleküle eine Selektion zur Bildung von kombinatorischen chemischen Bibliotheken durchlaufen.

Tabelle IV

Reaktive Funktionalitäten und ihre Reaktanten als Gegenstück und sich ergebende verknüpfende Bindungsgruppen

Reaktive Funktionalität	Funktionalität als Gegenstück	Bindungsgruppe
-NH ₂	-COOH	-CO-NH-
-NH ₂	-NCO	-NHCONH-
-NH ₂	-NCS	-NHCSNH-
-NH ₂		
-SH	-C≡C-CO-	-S-C≡C-CO-
-NH ₂	-CHO	-CH ₂ NH-
-NH ₂	-SO ₂ Cl	-SO ₂ NH-
-OH	-OP(NCH(CH ₃) ₂) ₂	-OP(=O)(O)O-
-OP(=O)(O)S	-NHC(=O)CH ₂ Br	-NHC(=O)CH ₂ SP(=O)(O)O-

[0046] Eine für die Erzeugung von kombinatorischen chemischen Bibliotheken besonders geeignete Klasse von Molekülen schließt lineare polymere Moleküle der Form:



ein, worin L eine Verknüpfungseinheit und M ein Monomer ist, welches aus einer großen Breite von chemischen Strukturen ausgewählt werden kann, um eine Breite von Funktionen bereitzustellen, welche als inerte nicht sterisch behindernde Abstandseinheit dienen können bis zur Bereitstellung einer reaktiven Funktionalität, welche als Verzweigungspunkt zur Anknüpfung anderer Komponenten, eine Stelle zum Anhängen von Markierungen, eine Stelle zum Anhängen von Oligonucleotiden oder anderen bindenden Polymeren zur Hybridisierung oder Bindung an ein therapeutisches Ziel, oder als Bindungsstelle für andere Gruppen zur Beeinflussung der Löslichkeit, Förderung der Duplexstrang- und/oder Triplexstrangbildung, wie interkalierende Substanzen oder alkylierende Substanzen und ähnliches, dienen kann. Die Sequenz und deshalb die Zusammensetzung

zung von solchen linearen polymeren Molekülen kann innerhalb eines an den Tag angehängten Polynucleotids codiert sein, wie durch Brenner und Lerner (vorstehend zitiert) gelehrt. Es kann jedoch nach einem Selektionsvorgang anstelle einer Amplifizierung und dann Sequenzierung des Tags des ausgewählten Moleküls, der Tag selbst oder ein zusätzlicher codierender Abschnitt unter Verwendung eines nachstehend beschriebenen sogenannten „Einzelbasen“-Ansatzes nach Freisetzung des interessierenden Moleküls, zum Beispiel durch Restriktionsspaltung einer in den Tag hineingesetzten Stelle, direkt sequenziert werden. Jedes Molekül, welches durch eine Abfolge von chemischen Reaktionsschritten, welche mit der gleichzeitigen Synthese von Tageinheiten kompatibel ist, kann natürlich zur Erzeugung von kombinatorischen chemischen Bibliotheken verwendet werden.

[0047] Günstiger Weise gibt es eine breite Vielfalt von über Phosphate verknüpften Monomeren, welche für die Erzeugung von kombinatorischen Bibliotheken verfügbar sind. Die nachfolgenden Quellen offenbaren einige Phosphoramidit- und/oder Hydrogenphosphonatmonomere, welche für die Verwendung in der vorliegenden Erfindung geeignet sind und stellen eine Anleitung für deren Synthese und Einschluss in Oligonucleotide bereit: Newton et al., *Nucleic Acids Research* 21 (1993): 1155–1162; Griffin et al., *J. Am. Chem. Soc.* 114 (1992): 7976–7982; Jaschke et al., *Tetrahedron Letters* 34 (1992): 301–304; Ma et al., Internationale Anmeldung WO 9306122; Zon et al., Internationale Anmeldung EP 0500776; Durand et al., *Nucleic Acids Research* 18 (1990): 6353–6359; Salunkhe et al., *J. Am. Chem. Soc.* 114 (1992): 8768–8772; Urdea et al., U.S. Patent 5,093,232; Ruth, U.S. Patent 4,948,882; Cruickshank, U.S. Patent 5,091,519; Haralambidis et al., *Nucleic Acids Research* 15 (1987): 4857–4876 und ähnliches. Insbesondere kann M eine geradkettige, cyclische oder verzweigte organische molekulare Struktur sein, welche 1 bis 20 Kohlenstoff- und 0 bis 10 Heteroatome, ausgewählt aus der Gruppe bestehend aus Sauerstoff, Stickstoff und Schwefel, enthält. Bevorzugt ist M ein Alkyl, Alkoxy, Alkenyl oder Aryl, welches 1 bis 16 Kohlenstoffatome enthält, ein Heterozyklus, welcher 3 bis 8 Kohlenstoffatome und 1 bis 3 Heteroatome, ausgewählt aus der Gruppe bestehend aus Sauerstoff, Stickstoff und Schwefel besitzt, ein Glycosyl oder ein Nucleosidyl. Stärker bevorzugt ist M ein Alkyl, Alkoxy, Alkenyl oder Aryl, welches von 1 bis 8 Kohlenstoffatome enthält, ein Glycosyl oder ein Nucleosidyl.

[0048] Bevorzugt ist L eine phosphorhaltige (V) verknüpfende Gruppe, welche ein Phosphodiester, ein Phosphotriester, ein Methyl- oder Ethylphosphonat, ein Phosphorothioat, ein Phosphorodithioat, ein Phosphoramidat oder ähnliches sein kann. Im Allgemeinen sind Verknüpfungen, welche von Phosphoramidit- oder Hydrogenphosphonatvorläufern abgeleitet sind bevorzugt, so dass die erfindungsgemäßen linearen Polymereinheiten günstig mit kommerziellen, automatisierten DNA-Syntheseautomaten, zum Beispiel Applied Biosystems, Inc. (Foster City, Calif.) Modell 394 oder ähnlichem, synthetisiert werden können.

[0049] n kann abhängig von der Art von M und L wesentlich variieren. Für gewöhnlich variiert n von etwa 3 bis etwa 100. Wenn M ein Nucleosid oder ein Analog davon oder ein Monomer der Größe eines Nucleosids ist und L eine phosphorhaltige M Verknüpfung, dann variiert n von etwa 12 bis etwa 100. Wenn M ein Nucleosid oder ein Analog davon oder ein Monomer der Größe eines Nucleosids ist und L eine phosphorhaltige Verknüpfung ist, dann variiert n bevorzugt von etwa 12 bis etwa 40.

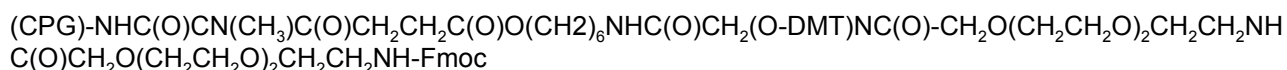
[0050] Peptide sind eine andere bevorzugte Klasse von Molekülen, an welche erfindungsgemäße Tags angehängt werden. Die Synthese von Peptid-Oligonucleotid-Konjugaten, welche in der Erfindung verwendet werden können, wird in Nielsen et al., *J. Am. Chem. Soc.* 115 (1993): 9812–9813; Haralambidis et al., (vorstehend zitiert) und die Internationale Patentanmeldung PCT/AU88/004417; Truffert et al., *Tetrahedron Letters* 35 (1994): 2353–2356; de la Torre et al., *Tetrahedron Letters* 35 (1994): 2733–2736; und ähnlichen Quellen gelehrt. Bevorzugt werden Peptid-Oligonucleotidkonjugate wie nachstehend beschrieben synthetisiert. In Übereinstimmung mit der Erfindung synthetisierte Peptide können aus den natürlichen Aminosäuremonomeren oder nicht-natürlichen Monomeren bestehen, einschließlich der D-Isomere der natürlichen Aminosäuren und ähnlichem.

Kombinatorische chemische Bibliotheken

[0051] Kombinatorische chemische Bibliotheken, welche erfindungsgemäße Tags nutzen, werden bevorzugt durch das in Nielsen et al. (vorstehend zitiert) offenbarte und in [Fig. 4](#) für eine bestimmte Ausführungsform dargestellte Verfahren hergestellt. Kurz gesagt wird ein Festphasenträger wie CPG mit einem spaltbaren Linker derivatisiert, welcher sowohl mit der zur Synthese des Tags verwendeten Chemie als auch der zur Synthese des Moleküls, welches ein Selektionsverfahren durchlaufen wird, verwendeten Chemie kompatibel ist. Bevorzugt werden die Tags unter Verwendung einer Phosphoramiditchemie wie vorstehend beschrieben und mit den durch Nielsen et al. (vorstehend zitiert) empfohlenen Modifikationen synthetisiert, das heißt, dass mit DMT-5'-O geschützte, mit 3'-Phosphoramidit derivatisierte Untereinheiten, welche mit Methyl geschützte Phosphit- und

Phosphateinheiten besitzen, in jedem Synthesesyklus hinzugefügt werden. Verbindungen der Bibliothek sind bevorzugt Monomere, welche Fmoc-, oder gleichwertige, Schutzgruppen haben, welche die Funktionalität maskieren, an welche nachfolgendes Monomer gekoppelt wird. Ein geeigneter Linker für sowohl DMT- als auch Fmoc-Schutzgruppen verwendende Chemien (hier als ein Sarcosin-Linker bezeichnet) wird durch Brown et al., J. Chem. Soc. Chem. Commun. (1989): 891–893 offenbart, was hier durch Bezugnahme als Quelle eingeschlossen wird.

[0052] Die [Fig. 4](#) stellt ein Schema für die Erzeugung einer kombinatorischen chemischen Bibliothek von an Oligonucleotidtags konjugierten Peptiden dar. Der Festphasenträger **200** wird durch einen Sarcosin-Linker **205** (durch die nachstehende Formel veranschaulicht), welcher eine verlängerte Verknüpfungseinheit besitzt um den Zugang für Reagenzien zu erleichtern, wie durch Nielsen et al. (vorstehend zitiert) gelehrt derivatisiert.



[0053] Hier bedeutet „CPG“ einen Controlled-Pore-Glasträger, „DMT“ bedeutet Dimethoxytrityl und „Fmoc“ bedeutet 9-Fluorenylmethoxycarbonyl. In einer bevorzugten Ausführungsform wird anfänglich ein Oligonucleotidabschnitt **214** synthetisiert, so dass in einer doppelsträngigen Form eine Restriktionsendonucleasestelle zur Spaltung der Verbindung der Bibliothek nach dem Sortieren auf einem Mikropartikel oder einem ähnlichem Substrat bereitgestellt wird. Die Synthese verläuft durch aufeinanderfolgendes abwechselndes Anfügen der Untereinheiten S_1 , S_2 , S_3 und ähnliches, um den Tag **212** zu erzeugen und deren entsprechender Monomere der Verbindungen der Bibliothek A_1 , A_2 , A_3 , und ähnliches um die Verbindung der Bibliothek zu erzeugen. Ein „Aufteil- und Mischverfahren“ wird verwendet um eine Vielfältigkeit zu erzeugen.

[0054] Die Untereinheiten in einem minimal kreuzhybridisierenden Set codieren das der Verbindung der Bibliothek hinzugefügte Monomer. Somit kann ein Set von neun Wörtern aus neun Monomeren erzeugte Verbindungen der Bibliothek unzweideutig codieren. Wenn eine Mehrdeutigkeit akzeptabel ist, dann kann eine einzelne Untereinheit mehr als ein Monomer codieren.

[0055] Nachdem die Synthese vollendet ist, wird das Produkt gespalten und die Schutzgruppe entfernt (**220**), um die tag-markierte Verbindung der Bibliothek zu erzeugen, welche dann eine Selektion durchläuft **230**, zum Beispiel Bindung an ein vorherbestimmtes Ziel **235**, wie ein Protein. Das Subset von Verbindungen der Bibliothek, welches aus dem Selektionsprozess **230** erhalten wurde, wird dann auf einem Festphasenträger **245** über deren Tag-Gruppen (dort werden komplementäre Untereinheiten und Nucleotide in kursiver Schrift gezeigt) sortiert (**240**). Nach Ligierung des Oligonucleotidspans **242** an das Tag-Komplementär **250** zur Bildung der Restriktionsstelle **255**, wird das Konjugat mit der entsprechenden Restriktionsendonuclease umgesetzt, um die Verbindung der Bibliothek zu spalten, im Beispiel der [Fig. 4](#) ein Peptid von der Oligonucleotidgruppe. Die Sequenz des Tags und damit die Identität der Verbindung der Bibliothek wird dann durch das nachstehend beschriebene, bevorzugte erfindungsgemäße Einzelbasen-Sequenzierungsverfahren bestimmt.

Festphasenträger

[0056] Festphasenträger zur Verwendung mit der Erfindung können eine breite Vielfalt von Formen, einschließlich Mikropartikeln, Kügelchen und Membranen, Objektträgern, Platten, mikroskopisch bearbeiteten Chips und ähnliches, haben. Ähnlich können erfindungsgemäße Festphasenträger eine breite Vielfalt von Zusammensetzungen, einschließlich Glas, Plastik, Silikon, mit Alkanthiolat derivatisiertes Gold, Cellulose, gering quervernetztes und stark quervernetztes Polystyrol, Kieselgel, Polyamid und ähnliches, umfassen. Bevorzugt wird entweder eine Population von abgegrenzten Partikeln verwendet, so dass jedes eine gleichförmige Beschichtung hat oder eine Population von komplementären Sequenzen des gleichen Tags (und keines anderen) hat, oder ein einzelner oder wenige Träger mit räumlich getrennten Regionen werden verwendet, wobei jede eine gleichförmige Beschichtung oder eine Population von komplementären Sequenzen zu dem gleichen Tag (und keinem anderen) enthält. In der letzten Ausführungsform kann die Fläche der Regionen entsprechend der speziellen Anwendungen variieren; für gewöhnlich liegen die Regionen der Fläche nach im Bereich von einigen μm^2 , zum Beispiel 3–5, bis zu einigen hundert μm^2 , zum Beispiel 100–500. Bevorzugt sind solche Regionen räumlich getrennt, so dass durch Ereignisse erzeugte Signale, zum Beispiel Fluoreszenzemissionen, in benachbarten Regionen durch das verwendete Nachweissystem aufgelöst werden können. In einigen Anwendungen kann es wünschenswert sein Regionen mit gleichförmigen Beschichtungen von mehr als einem Tag-Komplementär zu haben, zum Beispiel für die gleichzeitige Sequenzuntersuchung oder um getrennt tag-markierte Moleküle in nächste Nähe zu bringen.

[0057] Tag-Komplementäre können mit dem Festphasenträger verwendet werden, auf dem sie synthetisiert werden, oder sie können getrennt synthetisiert werden und zur Verwendung an einen Festphasenträger gebunden werden, zum Beispiel wie durch Lund et al., *Nucleic Acids Research*, 16 (1988): 10861–10880; Albretsen et al., *Anal. Biochem.* 189 (1990): 40–50; Wolf et al., *Nucleic Acids Research* 15 (1987): 2911–2926; oder Ghosh et al., *Nucleic Acids Research* 15 (1987): 5353–5372; offenbart. Bevorzugt werden Tag-Komplementäre auf dem gleichen Festphasenträger synthetisiert und verwendet, welcher eine Vielfalt von Formen umfassen und eine Vielfalt von Verknüpfungseinheiten umfassen kann. Solche Träger können Mikropartikel oder Anordnungen oder Matrizen von Regionen umfassen, wo gleichförmige Populationen von Tag-Komplementären synthetisiert werden. Mit der Erfindung kann eine breite Vielfalt von Mikropartikelträgern, einschließlich aus Controlled-Pore-Glas (CPG), stark quervernetztem Polystyrol, Acrylcopolymer, Cellulose, Nylon, Dextran, Latex, Polyacrolein und ähnlichem hergestellten Mikropartikeln, wie in den folgenden beispielhaften Quellen offenbart, verwendet werden: Meth. Enzymol., Abschnitt A, Seiten 11–147, Band 44 (Academic Press, New York, 1976); U.S. Patente 4,678,814; 4,413,070; und 4,046,720; und Pon, Kapitel 19, in Agrawal, Herausgeber, *Methods in Molecular Biology*, Band 20, (Humana Press, Totowa, N.J., 1993). Mikropartikelträger schließen weiterhin kommerziell erhältliche mit Nucleosiden derivatisierte CPG- und Polystyrolkügelchen ein (zum Beispiel erhältlich von Applied Biosystems, Foster City, CA); derivatisierte magnetische Kügelchen; mit Polyethylenglycol beimpftes Polystyrol (zum Beispiel TentaGel™, Rapp Polymere, Tübingen, Deutschland); und ähnliches. Die Auswahl der Charakteristiken des Trägers, wie Material, Porosität, Größe, Form und ähnliches und die Art der verwendeten Verknüpfungseinheit hängt von den Bedingungen ab, unter welchen die Tags verwendet werden. Zum Beispiel sind in Anwendungen, welche die aufeinanderfolgende Umsetzung mit Enzymen mit sich bringen, Träger und Linker, welche die sterische Hinderung der Enzyme minimieren und welche den Zugang zum Substrat erleichtern, bevorzugt. Andere wichtige bei der Auswahl des am besten geeigneten Mikropartikelträgers zu berücksichtigende Faktoren schließen die Gleichmäßigkeit der Größe, Effizienz als Syntheseträger, Grad zu dem die Größe der Oberfläche bekannt ist und optische Eigenschaften ein, zum Beispiel stellen wie nachstehend in vollem Umfang erklärt klare glatte Kügelchen instrumentelle Vorteile bei der Handhabung großer Mengen von Kügelchen auf einer Oberfläche bereit.

[0058] Beispielhafte Verknüpfungseinheiten zur Anhängung und/oder Synthese von Tags auf Mikropartikeloberflächen sind in Pon et al., *Biotechniques* 6 (1988): 768–775; Webb, U.S. Patent 4,659,774; Barany et al., Internationale Patentanmeldung EP 0546055; Brown et al., *J. Chem. Soc. Commun.* 1989: 891–893; Damha et al., *Nucleic Acids Research* 18 (1990): 3813–3821; Beattie et al., *Clinical Chemistry* 39 (1993): 719–722; Maskos und Southern, *Nucleic Acids Research* 20 (1992): 1679–1684; und ähnlichem offenbart.

[0059] Wie vorstehend erwähnt, können Tag-Komplementäre auch auf einem einzelnen (oder wenigen) Festphasenträger/n synthetisiert werden, um eine Anordnung von Regionen zu bilden, welche gleichförmig mit Tag-Komplementären beschichtet sind. Das heißt, dass in jeder Region in solch einer Anordnung das gleiche Tag-Komplementär synthetisiert wird. Verfahren zur Synthese solcher Anordnungen werden in McGall et al., Internationale Anmeldung WO 9322680; Pease et al., *Proc. Natl. Acad. Sci.* 91 (1994): 5022–5026; Southern und Maskos, Internationale Anmeldung EP 0386229; Maskos und Southern (vorstehend zitiert); Southern et al., *Genomics* 13 (1992): 1008–1017; und Maskos und Southern, *Nucleic Acids Research* 21 (1993): 4663–4669; offenbart.

[0060] Bevorzugt wird die Erfindung mit Mikropartikeln oder Kügelchen durchgeführt, welche gleichförmig mit Komplementären der gleichen Tag-Sequenz beschichtet sind. Mikropartikelträger und Verfahren zur kovalenten oder nicht-kovalenten Verknüpfung von Oligonucleotiden an ihre Oberflächen sind wohl bekannt, wie dies durch folgende Quellen veranschaulicht wird: Beaucage und Iyer (vorstehend zitiert); Gait, Herausgeber, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); und die vorstehend zitierten Quellen. Im Allgemeinen ist die Größe und Form eines Mikropartikels nicht kritisch; es sind jedoch Mikropartikel deren Größe im Bereich von einigen wenigen zum Beispiel 1–2 bis zu einigen hundert zum Beispiel 200–1000 µm Durchmesser reicht bevorzugt, da sie die Erzeugung und Manipulation von großen Repertoires von Oligonucleotidtags mit minimalem Reagenz- und Probenverbrauch ermöglichen.

[0061] In einigen bevorzugten Anwendungen werden Controlled-Pore-Glas (CPG)- oder Polystyrolträger als Festphasenträger in der Erfindung verwendet. Solche Träger sind mit basen-labilen Linkern und angehängten Startnucleosiden erhältlich, zum Beispiel Applied Biosystems (Foster City, CA). Bevorzugt werden Mikropartikel, welche Porengrößen zwischen 500 und 1000 Angström haben verwendet.

[0062] In anderen bevorzugten Anwendungen werden nicht-poröse Mikropartikel wegen ihrer optischen Eigenschaften verwendet, welche vorteilhaft verwendet werden können, wenn große Mengen von Mikropartikeln auf planaren Trägern wie einem Mikroskopierobjektträger verfolgt werden sollen. Besonders bevorzugte nicht-

poröse Mikropartikel sind Glycidal-Methacrylat (GMA)-Kügelchen, welche von Bangs Laboratories (Carmel, IN) erhältlich sind. Solche Mikropartikel sind in einer Vielfalt von Größen und mit einer Vielfalt von Verknüpfungsgruppen zur Synthese von Tags oder Tag-Komplementären derivatisiert nützlich. Bevorzugt werden für ausgeprägt parallele Manipulationen von mit einem Tag versehenen Mikropartikeln GMA-Kügelchen mit 5 µm Durchmesser verwendet.

Anhängen von Zielpolynucleotiden an Mikropartikel

[0063] Ein wichtiger Aspekt der Erfindung ist das Sortieren von Populationen identischer Polynucleotide, zum Beispiel aus einer cDNA-Bibliothek, und ihre Bindung an Mikropartikel oder getrennte Regionen eines Festphasenträgers, so dass jedes Mikropartikel oder jede Region nur eine einzige Art von Polynucleotid hat. Diese letzte Bedingung kann im wesentlichen durch Ligierung eines Repertoires von Tags an eine Population von Polynucleotiden erfüllt werden. Die Produkte der Ligierung werden dann cloniert, amplifiziert und eine Auswahl erhoben. Vorausgesetzt, dass die Probe ausreichend klein ist, wie dies mehr im vollen Umfang nachstehend erläutert wird, werden im wesentlichen alle Tag-Polynucleotidkonjugate der sich ergebenden Bibliothek einzigartig sein. Das heißt, dass jedes Polynucleotid einen einzigartigen Tag haben wird und umgekehrt. Die Polynucleotide werden dann durch Hybridisierung der Tags an ihre Komplementäre sortiert.

[0064] Ein Repertoire von Oligonucleotidtags kann an eine Population von Polynucleotiden in einer Menge von Arten, wie durch direkte enzymatische Ligierung, Amplifikation, zum Beispiel über PCR, unter Verwendung von die Tag-Sequenzen enthaltenden Primern, und ähnliches ligiert werden. Der erste Ligierungsschritt erzeugt eine sehr große Population von Tag-Polynucleotidkonjugaten, so dass ein einzelner Tag im Allgemeinen an viele verschiedene Polynucleotide angehängt ist. Durch Erhebung einer ausreichend kleinen Auswahl der Konjugate kann jedoch die Wahrscheinlichkeit „Dubletten“ zu erhalten, das heißt, der gleiche Tag auf zwei verschiedenen Polynucleotiden, vernachlässigbar gemacht werden. (Beachte, dass es auch möglich ist, verschiedene Tags mit dem gleichen Polynucleotid in einer Probe zu erhalten. Dieser Fall führt einfach dazu, dass ein Polynucleotid zweimal verarbeitet, zum Beispiel sequenziert, wird, was somit normalerweise nicht problematisch ist.) Wie nachstehend mehr im vollen Umfang erklärt wird, kann die Wahrscheinlichkeit in einer Probe eine Dublette zu erhalten durch eine Poisson-Verteilung geschätzt werden, da die Zahl von Konjugaten in einer Probe groß sein wird, zum Beispiel in der Größenordnung von Tausenden oder mehr und da, weil das Tag-Repertoire groß ist, zum Beispiel in der Größenordnung von Zehntausenden oder mehr, die Wahrscheinlichkeit für die Auswahl eines bestimmten Tags klein sein wird. Im Allgemeinen gilt, dass je größer die Probe ist die Wahrscheinlichkeit eine Dublette zu erhalten um so größer ist. So besteht beim Entwurf ein Zielkonflikt zwischen der Auswahl einer großen Probe von Tag-Polynucleotidkonjugaten, welche zum Beispiel eine angemessene Abdeckung eines Zielpolynucleotids bei einem „Shotgun“-Sequenzierungsverfahren sicherstellt, und dem Auswählen einer kleinen Probe, welche sicherstellt, dass eine minimale Zahl von Dubletten vorliegen wird. In den meisten Ausführungsformen fügt die Anwesenheit von Dubletten nur eine zusätzliche Quelle für ein Rauschen hinzu oder im Fall des Sequenzierens eine geringfügige Erschwerung des Abtastens und der Signalverarbeitung, da Mikropartikel, welche mehrfache Fluoreszenzsignale ergeben, einfach ignoriert werden können. Wie hier verwendet soll der Begriff „im wesentlichen alle“ im Bezug auf das Anhängen von Tags an Moleküle, insbesondere Polynucleotide, die statistische Natur des Verfahrens zur Erhebung einer Auswahl zum Erhalt einer Population von Tag-Molekülkonjugaten, welche im Wesentlichen frei von Dubletten ist, reflektieren. Die Bedeutung von im Wesentlichen alle in Begriffen von tatsächlichen Prozentsätzen von Tag-Molekülkonjugaten hängt davon ab, wie die Tags verwendet werden. Bevorzugt bedeutet für die Sequenzierung von Nucleinsäuren im Wesentlichen alle, dass an wenigstens achtzig Prozent der Tags einzigartige Polynucleotide angehängt sind. Stärker bevorzugt bedeutet es, dass an wenigstens neunzig Prozent der Tags einzigartige Polynucleotide angehängt sind. Noch stärker bevorzugt bedeutet es, dass an wenigstens fünfundneunzig Prozent der Tags einzigartige Polynucleotide angehängt sind. Und am stärksten bevorzugt bedeutet es, dass an wenigstens neunundneunzig Prozent der Tags einzigartige Polynucleotide angehängt sind.

[0065] Bevorzugt werden, wenn eine Population von Polynucleotiden aus Messenger-RNA (mRNA) besteht, Oligonucleotidtags durch eine reverse Transkription der mRNA mit einem Set von Primern, welches die Komplementären der Tag-Sequenzen enthält, angehängt. Ein beispielhaftes Set solcher Primer könnte die folgende Sequenz haben:

5' -mRNA- [A]_n -3'
 [T]₁₉GG[W, W, W, C]₉ACCAGCTGATC-5' - Biotin

[0066] Worin „[W, W, W, C]“ die Sequenz von einem Oligonucleotidtag aus neun Untereinheiten von jeweils vier Nucleotiden und „[W, W, W, C]“ die vorstehend aufgeführten Sequenzen der Untereinheiten, das heißt, „W“

stellt T oder A dar. Die unterstrichenen Sequenzen identifizieren eine optionale Restriktionsendonucleasestelle, welche verwendet werden kann, um das Polynucleotid aus einer Bindung an einen Festphasenträger über das Biotin, wenn eine verwendet wird, freizusetzen. Für den vorstehenden Primer könnte das an ein Mikropartikel gebundene Komplementär die Form haben:

5'-[G, W, W, W]₉ TGG – Linker – Mikropartikel

[0067] Nach der reversen Transkription wird die RNA, zum Beispiel durch Umsetzung mit RNase H, entfernt und der zweite Strang der cDNA wird zum Beispiel unter Verwendung eines Primers der folgenden Form synthetisiert:

5'-NRRGATCYNNN-3'

[0068] Worin N jedes aus A, T, G oder C ist; R ein Purin enthaltendes Nucleotid ist und Y ein ein Pyrimidin enthaltendes Nucleotid ist. Dieser bestimmte Primer erzeugt eine Bst YI Restriktionsstelle in der sich ergebenden doppelsträngigen DNA, welche zusammen mit der Sal I-Stelle, die Clonierung in einen Vektor mit zum Beispiel Bam HI und Xho I ermöglicht. Nach der Spaltung mit Bst YI und Sal I hätte das beispielhafte Konjugat die Form:

5' -RCGACCA [C, W, W, W]₉GG [T]₁₉- cDNA -NNNR
GGT [G, W, W, W]₉CC [A]₁₉- rDNA -NNNYCTAG-5'

[0069] Bevorzugt werden die mit Bst YI und Sal I gespaltenen Fragmente, wenn ein auf Ligase basierendes Verfahren zur Sequenzierung verwendet wird, in einen mit Bam HI/Xho I gespaltenen Vektor cloniert, welcher die folgenden Restriktionsstellen in einfacher Kopie hat:

5' -GAGGATGCCTTTATGGATCCACTCGAGATCCCAATCCA-3'
FokI BamHI XhoI

[0070] Dies fügt die Fok I-Stelle hinzu, welche die Einleitung des Sequenziervorganges wie nachstehend mehr in vollem Umfang diskutiert erlauben wird.

[0071] Ein allgemeines Verfahren für die Exposition eines einzelsträngigen Tags nach der Amplifikation schließt die Umsetzung eines ein Zielpolynucleotid enthaltenden Konjugates mit der 3'→5'-Exonucleaseaktivität der T4-DNA-Polymerase oder eines ähnlichen Enzyms ein. Wenn sie in der Anwesenheit eines einzelnen Nucleosidtriphosphates verwendet wird, wird eine solche Polymerase Nucleotide von an 3' zurückgesetzten Enden, welche auf dem nicht-Matrizen-Strang eines doppelsträngigen Fragments vorhanden sind, abspalten bis ein Komplementär des einzelnen Nucleosidtriphosphates auf dem Matrizenstrang erreicht ist. Wenn solch ein Nucleotid erreicht ist, endet die 3' → 5'-Umsetzung effektiv, da die Verlängerungsaktivität der Polymerase Nucleotide mit einer höheren Rate hinzufügt, als die Spaltungsaktivität Nucleotide entfernt. Daraus folgt, dass aus drei Nucleotiden erzeugte Tags schnell zur Beladung auf Festphasenträger vorzubereiten sind.

[0072] Die Technik kann auch verwendet werden, um bevorzugt interne Fok I-Stellen eines Zielpolynucleotids zu methylieren, während eine einzelne Fok I-Stelle am Ende des Polynucleotids unmethyliert belassen wird. Zuerst wird die endständige Fok I-Stelle unter Verwendung einer Polymerase mit Desoxycytidintriphosphat einzelsträngig gemacht. Der doppelsträngige Anteil des Fragments wird dann methyliert, wonach das einzelsträngige Ende mit einer DNA-Polymerase in der Anwesenheit aller vier Nucleosidtriphosphate aufgefüllt wird, wobei die Fok I-Stelle wiederhergestellt wird. Dieses Verfahren kann natürlich für andere Nucleasen als Fok I verallgemeinert werden.

[0073] Nachdem die Oligonucleotidtags für die spezifische Hybridisierung, zum Beispiel indem sie wie vorstehend beschrieben einzelsträngig gemacht wurden, vorbereitet wurden, werden die Polynucleotide mit Mikropartikeln, welche die komplementären Sequenzen der Tags enthalten, unter Bedingungen gemischt, welche die Bildung von perfekt zusammenpassenden Duplexsträngen zwischen den Tags und ihren Komplementären begünstigen. In der Literatur gibt es ausführliche Anleitung zur Erzeugung dieser Bedingungen. Beispielhafte Quelle, welche solche Anleitung bereitstellen, schließen Wetmur, Critical Reviews in Biochemistry and Molecular Biology 26 (1991): 227–259; Sambrook et al., Molecular Cloning: A Laboratory Manual, 2. Ausgabe (Cold Spring Harbor Laboratory, New York, 1989); und ähnliches ein. Bevorzugt sind die Hybridisierungsbedingungen ausreichend stringent, so dass nur perfekt zusammenpassende Sequenzen stabile Duplexstränge ausbilden. Unter solchen Bedingungen werden die durch ihre Tags spezifisch hybridisierten Polynucleotide an die an die Mikropartikel angehängten komplementären Sequenzen ligiert. Schließlich werden die Mikropartikel ge-

waschen, um unligierte Polynucleotide zu entfernen.

[0074] Wenn für gewöhnlich als Syntheseträger verwendete CPG-Mikropartikel verwendet werden, ist die Dichte der Tag-Komplementäre auf der Oberfläche der Mikropartikel typischer Weise größer, als jene die für einige Sequenzierungsverfahren nötig ist. Das heißt, dass in Sequenzierungsansätzen, welche eine aufeinanderfolgende Behandlung der angehängten Polynucleotide mit einer Vielfalt von Enzymen erfordern, können dicht gesetzte Polynucleotide dazu neigen den Zugang der relativ sperrigen Enzyme an die Polynucleotide zu behindern. In solchen Fällen werden die Polynucleotide bevorzugt mit den Mikropartikeln so gemischt, dass die Tag-Komplementäre gegenüber den Polynucleotiden in einem wesentlichen Überschuss vorhanden sind, zum Beispiel von 10:1 bis zu 100:1 oder mehr. Dies stellt sicher, dass die Dichte der Polynucleotide auf der Oberfläche der Mikropartikel nicht so hoch sein wird, dass der Zugang der Enzyme behindert wird. Bevorzugt ist der durchschnittliche Abstand zwischen den Polynucleotiden auf der Oberfläche der Mikropartikel in der Größenordnung von 30–100 nm. Anleitung zur Auswahl von Verhältnissen für Standard-CPG-Träger und Balotini-Kügelchen (eine Art von festem Glasträger) ist in Maskos und Southern, *Nucleic Acids Research* 20 (1992): 1679–1684, zu finden. Bevorzugt werden für Sequenzierungsanwendungen Standard-CPG-Kügelchen eines Durchmessers im Bereich von 20–50 µm mit etwa 10^5 Polynucleotiden beladen und GMA-Kügelchen eines Durchmessers im Bereich von 5–10 µm werden mit einigen wenigen zehntausend Polynucleotiden, zum Beispiel 4×10^4 bis zu 6×10^4 , beladen.

[0075] Das vorstehende Verfahren kann verwendet werden, um ein Fingerprinting von mRNA-Populationen durchzuführen, wenn es mit der nachstehend beschriebenen Verfahrensweise zur parallelen Sequenzierung kombiniert wird. Teilweise Sequenzinformationen werden gleichzeitig von einer großen Auswahl, zum Beispiel zehn- bis einhunderttausend, von wie im vorstehenden Verfahren beschrieben an getrennte Mikropartikel angehängten cDNAs erhalten. Die Häufigkeitsverteilung von Teilsequenzen kann mRNA-Populationen aus verschiedenen Zell- oder Gewebearten, genauso wie von erkrankten Geweben wie Krebs, identifizieren. Solche mRNA-Fingerprints sind in der Überwachung und Diagnostik von Krankheitszuständen, zum Beispiel Internationale Anmeldung PCT/US95/21944, welche die Verwendung von Expressed-Sequenztags (ESTs) für den gleichen Zweck beschreibt, nützlich.

Einzelbasen-DNA-Sequenzierung

[0076] Die vorliegende Erfindung kann mit üblichen Verfahren der DNA-Sequenzierung, wie zum Beispiel durch Hultman et al., *Nucleic Acids Research* 17 (1989): 4937–4946 offenbart verwendet werden. Für die parallele oder gleichzeitige Sequenzierung von multiplen Polynucleotiden siehe jedoch folgende Quellen: Cheeseman, U.S.-Patent 5,302,509; Tsien et al., Internationale Anmeldung WO 91/06678; Rosenthal et al., Internationale Anmeldung WO 93/21340; Canard et al., *Gene* 148 (1994): 1–6; und Metzker et al., *Nucleic Acids Research* 22 (1994): 4259–4267.

[0077] Ein „Einzelbasen“-Verfahren der DNA-Sequenzierung, welches für die Verwendung mit der vorliegenden Erfindung geeignet ist und welches keine elektrophoretische Auftrennung der DNA-Fragmente erfordert, ist in der internationalen Anmeldung EP 0703991 beschrieben. Das Verfahren umfasst die folgenden Schritte: (a) Ligierung einer Sonde an ein Ende des Polynucleotids, welches einen hervorstehenden Strang hat, um einen ligierten Komplex zu bilden, wobei die Sonde einen zu dem des Polynucleotids komplementären hervorstehenden Strang hat und die Sonde eine Nucleasenerkennungsstelle hat; (b) Entfernung der nicht-ligierten Sonde aus dem ligierten Komplex; (c) Identifizierung von einem oder mehreren Nucleotiden in dem hervorstehenden Strang des Polynucleotids durch die Identität der ligierten Sonde; (d) Spaltung des ligierten Komplexes mit einer Nuclease; und (e) Wiederholen der Schritte (a) bis (d) bis die Nucleotidsequenz des Polynucleotids bestimmt ist. Wie nachstehend mehr im vollen Umfang beschrieben ist, kann die Identifizierung des einen oder der mehreren Nucleotids/Nucleotide entweder vor oder nach Spaltung des ligierten Komplexes von dem Zielpolynucleotid durchgeführt werden. Bevorzugt schließt das Verfahren, wann immer natürliche Proteinendonucleasen verwendet werden, einen Schritt zur Methylierung des Zielpolynucleotids am Beginn eines Sequenzvorganges ein.

[0078] Eine wichtige Charakteristik des Verfahrens ist die an das Zielpolynucleotid ligierte Sonde. Eine bevorzugte Form der Sonden wird in [Fig. 1a](#) dargestellt. Im Allgemeinen sind die Sonden doppelsträngige DNA mit einem hervorstehenden Strang an einem Ende **10**. Die Sonden enthalten wenigstens eine Nucleaseerkenntnisstelle **12** und eine Platzhalterregion **14** zwischen der Erkennungsstelle und dem hervorstehenden Ende **10**. Bevorzugt schließen Sonden auch eine Markierung **16** ein, welche in dieser besonderen Ausführungsform am dem hervorstehenden Strang entgegengesetzten Ende dargestellt ist. Die Sonden können durch eine Vielfalt von Mitteln und an einer Vielfalt von Positionen, mit der einzigen Einschränkung, dass das ausgewählte

Markierungsmittel den Ligierungsschritt oder die Erkennung der Sonde durch die Nuclease nicht stören darf, markiert werden.

[0079] Es ist nicht entscheidend, ob der hervorstehende Strang **10** der Sonde am 5'- oder 3'-Ende ist. Es ist jedoch wichtig, dass die hervorstehenden Stränge des Zielpolynucleotids und der Sonde fähig sind, perfekt zusammenpassende Duplexstränge zu bilden, um so eine spezifische Ligierung zu erlauben. Wenn die hervorstehenden Stränge des Zielpolynucleotids und der Sonde verschiedene Längen haben, kann die sich ergebende Lücke mit einer Polymerase vor der Ligierung, zum Beispiel wie in einer in Backman et al., Europäische Patentanmeldung EP 0439182 offenbarten „Gap-LCR“, aufgefüllt werden. Bevorzugt ist die Zahl von Nucleotiden in den jeweiligen hervorstehenden Strängen die gleiche, so dass die beiden Stränge der Sonde und des Zielpolynucleotids ohne einen Auffüllungsschritt ligiert werden können. Bevorzugt ist der hervorstehende Strang der Sonde 2 bis zu 6 Nucleotide lang. Wie nachstehend aufgezeigt, ist die Komplexität des während jeden Ligierungs- und Spaltungszyklus auf das Zielpolynucleotid aufgetragenen Gemisches der Sonden um so größer je größer die Länge des hervorstehenden Stranges ist.

[0080] Die komplementären Stränge der Sonden werden günstiger Weise auf einem automatisierten DNA-Syntheseautomaten, zum Beispiel einem Applied Biosystems, Inc. (Foster City, California) Modell 392 oder 394 DNA/RNA Synthesizer, unter Verwendung von Standardchemien synthetisiert. Nach der Synthese werden die komplementären Stränge kombiniert, um eine doppelsträngige Sonde zu erzeugen. Im Allgemeinen wird der hervorstehende Strang einer Sonde als Gemisch synthetisiert, so dass jede mögliche Sequenz im hervorstehenden Anteil vertreten ist. Zum Beispiel werden vier Gemische wie folgt hergestellt, wenn in einer Ausführungsform der hervorstehende Anteil aus vier Nucleotiden besteht:

$X_1X_2 \dots X_iNNNA,$
 $X_1X_2 \dots X_iNNNC,$
 $X_1X_2 \dots X_iNNNG,$ und
 $X_1X_2 \dots X_iNNNT$

[0081] Worin die „NNN“ jedes mögliche 3-mer darstellen und die „X“ den Anteil des Stranges darstellen, welcher den Duplexstrang bildet. Somit enthält jede der vorstehend aufgeführten Sonden 4^3 oder 64 unterschiedliche Sequenzen; oder in anderen Worten, jede der vier Sonden hat eine Degeneriertheit von 64. Zum Beispiel enthält $X_1X_2 \dots X_iNNNA$ die folgenden Sequenzen:

$$X_1X_2 \dots X_iAAAA$$

$$X_1X_2 \dots X_iAACA$$

$$X_1X_2 \dots X_iAAGA$$

$$X_1X_2 \dots X_iAATA$$

$$X_1X_2 \dots X_iACAA$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$X_1X_2 \dots X_iTGTA$$

$$X_1X_2 \dots X_iTTAA$$

$$X_1X_2 \dots X_iTTCA$$

$$X_1X_2 \dots X_iTTGA$$

$$X_1X_2 \dots X_iTTTA$$

[0082] Solche Gemische werden einfach unter Verwendung wohl bekannter Verfahren, zum Beispiel wie in Telenius et al. (vorstehend zitiert) offenbart, synthetisiert. Im Allgemeinen erfordern diese Verfahren einfach während der Kopplungsschritte die Anwendung von Gemischen der aktivierten Monomere an dem wachsenden Oligonucleotid, wo man die Degeneriertheit einzuführen wünscht. In einigen Ausführungsformen kann es wünschenswert sein die Degeneriertheit der Sonden zu reduzieren. Dies kann durch die Verwendung von Analoga, wie Desoxyinosin, 2-Aminopurin oder ähnlichem, welche die Degeneriertheit herabsetzen, wie zum Beispiel in Kong Thoo Lin et al., Nucleic Acids Research 20: 5149–5152, oder durch das U.S. Patent 5,002,867 gelehrt, bewerkstelligt werden.

[0083] Bevorzugt ist die den Duplexstrang bildende Region einer Sonde für Oligonucleotide mit Phosphodiesterbindungen zwischen etwa 12 bis etwa 30 Basenpaare lang; stärker bevorzugt ist ihre Länge zwischen etwa 15 bis etwa 25 Basenpaaren.

[0084] Wenn übliche Ligasen in der Erfindung verwendet werden, wie dies nachstehend mehr in vollem Umfang beschrieben ist, kann das 5'-Ende der Sonde in einigen Ausführungsformen phosphoryliert werden. Ein 5'-Monophosphat kann an ein zweites Oligonucleotide entweder chemisch oder enzymatisch mit einer Kinase, zum Beispiel Sambrook et al. (vorstehend zitiert) angehängt werden. Die chemische Phosphorylierung wird durch Horn und Urdea, Tetrahedron Lett. 27 (1986): 4705 beschrieben, und Reagenzien zur Durchführung der offenbarten Protokolle sind kommerziell erhältlich, zum Beispiel 5' Phosphate-ON™ von Clontech Laboratories (Palo Alto, California). Somit können in einigen Ausführungsformen Sonden die Form haben:

$$\begin{array}{l} 5' - X_1X_2 \dots X_iTTGA \\ \quad Y_1Y_2 \dots Y_iP \end{array}$$

[0085] Worin die Y die komplementären Nucleotide der X sind und „p“ eine Monophosphatgruppe ist.

[0086] Die vorstehenden Sonden können in einer Vielfalt von Arten, einschließlich des direkten oder indirekten Anhängens von radioaktiven Einheiten, fluoreszenten Einheiten, colorimetrischen Einheiten, chemilumineszenten Markierungen und ähnlichem, markiert werden. Viele umfassende Übersichten über Verfahrensweisen zur Markierung von DNA und der Erzeugung von DNA-Sonden stellen Anleitungen bereit, welche bei der Erzeugung von erfindungsgemäßen Sonden anwendbar sind. Solche Übersichten schließen Kricka, Heraus-

geber, Nonisotopic DNA Probe Techniques (Academic Press, San Diego, 1992); Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Inc., Eugene, 1992); Keller und Manak, DNA Probes, 2. Ausgabe (Stockton Press, New York, 1993); und Eckstein, Herausgeber, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991); Kessler, Herausgeber, Nonradioactive Labeling and Detection of Biomolecules (Springer-Verlag, Berlin, 1992); Wetmur (vorstehend zitiert); und ähnliches ein.

[0087] Bevorzugt werden die Sonden mit einem oder mehreren Fluoreszenzfarbstoffen, zum Beispiel wie durch Menchen et al., U.S. Patent 5,188,934; Begot et al. Internationale Anmeldung EP 0496749 offenbart, markiert.

[0088] In Übereinstimmung mit dem Verfahren wird eine Sonde an das Ende eines Zielpolynucleotids ligiert, um in jedem Zyklus von Ligierung und Spaltung einen ligierten Komplex zu erzeugen. Der ligierte Komplex ist die doppelsträngige Struktur, welche gebildet wird, nachdem sich die hervorstehenden Stränge des Zielpolynucleotids und der Sonde aneinanderlagern und wenigstens ein Paar der identisch orientierten Stränge der Sonde und des Ziels ligiert werden, das heißt, dass sie zur kovalenten Bindung aneinander gebracht werden. Die Ligierung kann entweder enzymatisch oder chemisch erreicht werden. Chemische Ligierungsverfahren sind im Fachgebiet wohl bekannt, zum Beispiel Ferris et al., Nucleosides & Nucleotides 8 (1989): 407–414; Shabarova et al., Nucleic Acids Research 19 (1991): 4247–4251; und ähnliches. Bevorzugt wird die Ligierung jedoch enzymatisch unter Verwendung einer Ligase nach einem Standardprotokoll durchgeführt. Viele Ligasen sind bekannt und sind für die Verwendung in der Erfindung geeignet, zum Beispiel Lehman, Science 186 (1974): 790–797; Engler et al., DNA Ligases, Seiten 3–30 in Boyer, Herausgeber, The Enzymes, Band 15B (Academic Press, New York, 1982); und ähnliches. Bevorzugte Ligasen schließen T4-DNA-Ligase, T7-DNA-Ligase, E. coli-DNA-Ligase, Taq-Ligase, Pfu-Ligase und Tth-Ligase ein. Protokolle für deren Verwendung sind wohl bekannt, zum Beispiel Sambrook et al. (vorstehend zitiert), Barany, PCR Methods and Applications 1 (1991): 5–16; Marsh et al., Strategies 5 (1992): 73–76; und ähnliches. Im Allgemeinen erfordern Ligasen, dass eine 5'-Phosphatgruppe für die Ligierung an das 3'-Hydroxyl eines angrenzenden Stranges vorhanden ist. Dies wird günstigerweise dadurch für wenigstens einen Strang des Zielpolynucleotids bereitgestellt, dass eine Nuclease, wie zum Beispiel Fok I ausgewählt wird, welche ein 5'-Phosphat hinterlässt.

[0089] In einer Ausführungsform des Sequenzierungsverfahrens, welche nichtphosphorylierte Sonden verwendet, schließt der Schritt der Ligierung (i) das Ligieren der Sonde an das Zielpolynucleotid mit einer Ligase, so dass ein ligierter Komplex erzeugt wird, welcher einen Bruch in einem Strang besitzt, (ii) das Phosphorylieren des 5'-Hydroxyls am Bruch mit einer Kinase unter Verwendung üblicher Protokolle, zum Beispiel Sambrook et al. (vorstehend zitiert), und (iii) das erneute Ligieren, um die Stränge am Bruch kovalent zu verbinden, das heißt, den Bruch zu entfernen, ein.

Gerät zur Beobachtung enzymatischer Vorgänge und/oder von Bindungsereignissen auf Mikropartikeloberflächen

[0090] Ein erfindungsgemäßes Ziel ist es, identische Moleküle, insbesondere Polynucleotide, auf die Oberflächen von Mikropartikeln durch die spezifische Hybridisierung von Tags und ihren Komplementären zu sortieren. Hat ein solches Sortieren stattgefunden, kann die Anwesenheit der Moleküle oder die an ihnen vorgenommenen Maßnahmen auf verschiedene Arten nachgewiesen werden, abhängig von der Art der mit einem Tag markierten Moleküle, ob Mikropartikel getrennt oder in „Reihen“, ob wiederholte Messungen gewünscht sind, und ähnliches, nachgewiesen werden. Typischer Weise werden die sortierten Moleküle gegenüber Liganden zur Bindung ausgesetzt, zum Beispiel in der Arzneimittelentwicklung oder werden chemischen oder enzymatischen Vorgängen unterzogen, zum Beispiel bei der Polynucleotidsequenzierung. In beiden dieser Anwendungen ist es oft wünschenswert gleichzeitig Signale, welche mit solchen Ereignissen oder Vorgängen korrespondieren auf großen Mengen von Mikropartikeln zu beobachten. Sortierte Moleküle tragende Mikropartikel (hier als „beladene“ Mikropartikel bezeichnet) bieten sich für solche parallele Maßnahmen in großem Maßstab, wie zum Beispiel durch Lam et al. (vorstehend zitiert) gezeigt, an.

[0091] Bevorzugt werden, wann immer lichterzeugende Signale, zum Beispiel chemilumineszent, fluorescent oder ähnliches, verwendet werden, um Ereignisse oder Vorgänge nachzuweisen, beladene Mikropartikel zur Untersuchung mit einem Abtastungssystem, so wie in den Internationalen Patentanmeldungen WO 9210587, EP 0476014 und WO 9522058 beschrieben, auf einem Planaren Substrat, zum Beispiel einem Glasobjektträger, verteilt. Das Abtastungssystem sollte fähig sein, das Substrat reproduzierbar abzutasten und die Positionen von jedem der Mikropartikel in einer vorbestimmten Region in der Art eines Koordinatensystems zu definieren. In Anwendungen der Polynucleotidsequenzierung ist es wichtig, dass die Identifizierung von Mikropartikeln bezüglich der Position in aufeinanderfolgenden Abtastschritten wiederholbar ist.

[0092] Solche Abtastungssysteme können aus kommerziell erhältlichen Komponenten zusammengesetzt werden, zum Beispiel einem durch einen digitalen Computer gesteuerten bezüglich x- und y-Achsen linear bewegten Tisch, der zusammen mit einem Nachweissystem verwendet wird, das eine oder mehrere Photomultiplerröhren oder alternativ eine CCD-Anordnung und geeigneten Optiken umfasst, zum Beispiel zur Anregung, Sammlung oder Sortierung von Fluoreszenzsignalen. In einigen Ausführungsformen kann ein konfokales optisches System wünschenswert sein. Ein beispielhaftes Abtastsystem, welches für die Verwendung in der Vierfarbensequenzierung geeignet ist, wird diagrammartig in [Fig. 5](#) dargestellt. Das Substrat **300**, zum Beispiel ein Mikroskopobjektträger mit fixierten Mikropartikeln, wird auf einen bezüglich x- und y-Achsen linear bewegten Tisch **302** gesetzt, welcher an einen in geeigneter Weise programmierten digitalen Computer **304** angeschlossen und durch ihn gesteuert ist, welcher jeder aus einer Vielfalt von kommerziell erhältlichen Personal-Computern, zum Beispiel einem auf dem 486 basierten Gerät oder einem PowerPC Modell 7100 oder 8100, welcher von Apple Computer (Cupertino, CA) erhältlich ist, sein kann. Computer-Software für die lineare Bewegung des Tisches und Sammelfunktionen für Daten können durch kommerziell erhältliche Laborsoftware, wie Lab Windows, erhältlich von National Instruments, bereitgestellt werden.

[0093] Das Substrat **300** und der Tisch **302** sind funktionell mit dem Mikroskop **306** assoziiert, welches eine oder mehrere Objektivlinsen **308** besitzt, welche zur Sammlung und Abgabe von Licht von/an auf dem Substrat **300** fixierte/n Mikropartikel/n fähig ist/sind. Der Anregungsstrahl **310** von der Lichtquelle **312**, welche bevorzugt ein Laser ist, wird auf den Strahlteiler **314** gerichtet, zum Beispiel einen dichroitischen Spiegel, welcher den Strahl durch das Mikroskop **306** und die Objektivlinse **308** umleitet, welche ihrerseits den Strahl auf das Substrat **300** fokussiert. Die Linse **308** sammelt die Fluoreszenz **316**, welche von den Mikropartikeln emittiert wird und richtet sie durch den Strahlteiler **314** zur Signalverteilungsoptik **318**, welche ihrerseits die Fluoreszenz auf eine oder mehrere geeignete optoelektronische Einheiten zur Umwandlung einiger Fluoreszenzcharakteristiken in ein elektrisches Signal, zum Beispiel der Intensität, der Lebensdauer oder ähnlichem, richtet. Die Signalverteilungsoptik **318** kann eine Vielfalt von Komponenten, welche im Fachgebiet Standard sind, wie Bandpassfilter, Glasfaseroptiken, rotierende Spiegel, Spiegel und Linsen mit fixierten Positionen, Brechungsgitter und ähnliches umfassen. Wie in [Fig. 5](#) dargestellt richtet die Signalverteilungsoptik **318** die Fluoreszenz **316** auf vier getrennte Photomultiplerröhren, **330**, **332**, **334** und **336**, deren Ausgabe dann zu Vorverstärkern und Photonen-zählern **350**, **352**, **354** und **356** geleitet wird. Die Ausgabe der Photonen-zähler wird durch den Computer **304** gesammelt, wo sie gespeichert, untersucht und auf einem Bildschirm **360** angesehen werden kann. Alternativ kann die Signalverteilungsoptik **318** ein Brechungsgitter sein, welches das Fluoreszenzsignal **318** auf eine CCD-Anordnung richten kann.

[0094] Die Stabilität und Reproduzierbarkeit der Lokalisierung bezüglich der Position in der Abtastung wird zu einem großen Umfang die Auflösung für die Trennung eng zusammenliegender Mikropartikel bestimmen. Bevorzugt sollte das Abtastungssystem fähig sein, nahe zusammenliegende Mikropartikel, zum Beispiel durch einen Durchmesser eines Mikropartikels oder weniger getrennt, aufzulösen. Somit sollte für die meisten Anwendungen, zum Beispiel solche, die CPG-Mikropartikel verwenden, das Abtastungssystem wenigstens die Fähigkeit haben, Objekte in der Größenordnung von 10–100 µm aufzulösen. Noch höhere Auflösung kann in einigen Ausführungsformen wünschenswert sein, mit gesteigerter Auflösung wird aber die benötigte Zeit zur vollständigen Abtastung eines Substrats zunehmen; somit wird in einigen Ausführungsformen ein Kompromiss zwischen Geschwindigkeit und Auflösung zu machen sein. Eine Steigerung in der Abtastzeit kann durch ein System erreicht werden, welches nur Positionen abtastet, wo, zum Beispiel aus einer anfänglichen vollständigen Abtastung, bekanntermaßen Mikropartikel lokalisiert sind. Bevorzugt werden die Größe der Mikropartikel und die Auflösung des Abtastungssystems so ausgewählt, dass es die Auflösung von fluoreszenzmarkierten Mikropartikeln erlaubt, welche zufällig auf einer Ebene mit einer Dichte zwischen etwa zehntausend bis zu etwa einhunderttausend Mikropartikeln pro cm² angeordnet sind.

[0095] In Sequenzierungsanwendungen können beladene Mikropartikel auf der Oberfläche eines Substrates in einer Vielfalt von Arten fixiert werden. Die Fixierung sollte stark genug sein, um es den Mikropartikeln zu erlauben aufeinanderfolgende Zyklen der Aussetzung gegenüber Reagenzien und des Waschens ohne wesentliche Verluste zu durchlaufen. Wenn das Substrat Glas ist, kann dessen Oberfläche mit einem Alkylamino-Linker unter Verwendung von kommerziell erhältlichen Reagenzien, zum Beispiel Pierce Chemical, derivatisiert werden, welcher seinerseits mit Avidin, erneut unter Verwendung von üblichen Chemien, quervernetzt werden kann, um eine mit Avidin besetzte Oberfläche zu erzeugen. Biotin-Einheiten können auf die beladenen Mikropartikel über eine Menge von Wegen aufgebracht werden. Zum Beispiel wird ein Anteil, zum Beispiel 10–15 Prozent, der zum Anhängen der Tags an Polynucleotide verwendeten Clonierungsvektoren so entworfen, dass sie eine einzigartige Restriktionsstelle (stellt bei Spaltung haftende Enden bereit) unmittelbar dem Polynucleotidinsert angrenzend, am dem Tag gegenüberliegenden Ende des Polynucleotids enthalten. Die Stelle wird mit dem Polynucleotid und dem Tag für die Beladung auf die Mikropartikel ausgeschnitten. Nach

der Beladung werden etwa 10–15 Prozent der geladenen Polynucleotide die einzigartige Restriktionsstelle entfernt von der Mikropartikeloberfläche besitzen. Nach Spaltung mit der assoziierten Restriktionsendonuclease wird ein geeigneter, doppelsträngiger Adapter, welcher einen Biotineinheit enthält, an das haftende Ende ligiert. Die sich ergebenden Mikropartikel werden dann auf der mit Avidin besetzten Glasoberfläche ausgebreitet, wo sie über die Avidin-Biotin-Bindungen fixiert werden.

[0096] Alternativ, und bevorzugt wird, wenn eine Sequenzierung über Ligierung verwendet wird, im ersten Ligierungsschritt ein Gemisch von Sonden auf den beladenen Mikropartikeln angewendet: ein Anteil der Sonden enthält eine Typ II-Restriktionserkennungsstelle, wie dies vom Sequenzierverfahren gefordert wird, und ein Anteil der Sonden hat eine solche Erkennungsstelle nicht, aber stattdessen enthalten sie eine Biotineinheit an ihrem Ende, welches nicht ligiert. Bevorzugt umfasst das Gemisch etwa 10–15 Prozent der biotinylierten Sonde.

[0097] In noch einer anderen Alternative, wenn DNA-beladene Mikropartikel auf ein Glassubstrat aufgebracht werden, kann die DNA über mehrere Stunden Inkubation, zum Beispiel 24 Stunden, unspezifisch an die Glasoberfläche adsorbieren, um einen ausreichend starke Bindung zu erzeugen, um die wiederholte Aussetzung gegenüber Reagenzien und Waschungen ohne wesentlichen Verlust von Mikropartikeln zu ermöglichen. Bevorzugt ist ein solches Glassubstrat eine Flusszelle, welche einen in einen Glasobjektträger geätzten Kanal umfassen kann. Bevorzugt ist solch ein Kanal geschlossen, so dass Flüssigkeiten durch ihn gepumpt werden können und hat eine Tiefe, welche ausreichend nah am Durchmesser der Mikropartikel liegt, so dass eine einlagige Schicht von Mikropartikeln in einer definierten Beobachtungsregion eingefangen wird.

Parallele Sequenzierung

[0098] Das erfindungsgemäße Tag-Markierungssystem kann mit Einzelbasen-Sequenzierungsverfahren zur Sequenzierung von Polynucleotiden bis zu einer Länge von einigen Kilobasen verwendet werden. Das Tag-Markierungssystem erlaubt es, viele tausend Fragmente eines Zielpolynucleotids auf einen oder mehrere Festphasenträger zu sortieren und gleichzeitig zu sequenzieren. In Übereinstimmung mit einer bevorzugten Ausführung des Verfahrens, wird ein Teil von jedem sortierten Fragment in einer schrittweisen Art auf jedem der vielen tausend beladenen Mikropartikel, welche auf einem üblichen, mit einem Abtastungssystem oder einem Bildauswertungssystem, wie dies nachstehend beschrieben wird, assoziierten Substrat wie einem Mikroskopobjektträger fixiert sind, sequenziert. Die Größe des Anteils der sequenzierten Fragmente hängt von einigen Faktoren, wie der Zahl von erzeugten und sortierten Fragmenten, der Länge des Zielpolynucleotids, der Geschwindigkeit und der Genauigkeit des verwendeten Einzelbasen-Verfahrens, der Zahl der Mikropartikel und/oder der abgegrenzten Regionen, welche gleichzeitig überwacht werden können; und ähnlichem ab. Bevorzugt werden von 12 bis 50 Basen auf jedem Mikropartikel oder jeder Region identifiziert, und stärker bevorzugt werden 18–30 Basen auf jedem Mikropartikel oder jeder Region identifiziert. Mit dieser Information wird die Sequenz des Zielpolynucleotids durch Zuordnen der 12–50-Basenfragmente über deren überlappende Regionen, wie zum Beispiel im U.S. Patent 5,002,867 beschrieben, bestimmt. Die nachfolgenden Quellen stellen zusätzliche Anleitung in der Bestimmung des Anteils der Fragmente bereit, welcher sequenziert werden muss, um das Zielpolynucleotid einer gegebenen Länge erfolgreich zu Rekonstruieren: Lander und Waterman, *Genomics*, 2 (1988): 231–239; Drmanac et al., *Genomics* 4 (1989): 114–128; Bains, *DNA Sequencing and Mapping* 4 (1993): 143–150; Bains, *Genomics* 11 (1991): 294–301; Drmanac et al., *J. Biomolecular Structure and Dynamics* 8 (1991): 1085–1102; und Pevzner J. *Biomolecular Structure and Dynamics* 7 (1989): 63–73. Bevorzugt ist die Länge des Zielpolynucleotids zwischen 1 Kilobase und 50 Kilobasen. Stärker Bevorzugt ist die Länge zwischen 10 Kilobasen und 40 Kilobasen. Lander und Waterman (vorstehend zitiert) stellen Anleitung bezüglich der Beziehung zwischen der Zahl der sequenzierten Fragmente (das heißt der Probengröße), der Menge der von jedem Fragment erhaltenen Sequenzinformation und der Wahrscheinlichkeit dass das Zielpolynucleotid von den Teilsequenzen ohne Lücken oder „Inseln“ rekonstruiert werden kann. Für die vorliegende Erfindung werden die maximalen Polynucleotidgrößen, welche für gegebene Probengrößen und Größen der Fragmentsequenzen, erhalten werden können nachstehend gezeigt:

Größe der Probe	Annähernde maximale Länge des Zielpolynucleotids	
-----------------	--	--

	30 Basen / Fragment	50 Basen / Fragment
1000	3 Kilobasen	4 Kilobasen
10000	22 Kilobasen	32 Kilobasen
20000	40 Kilobasen	65 Kilobasen
30000	60 Kilobasen	85 Kilobasen
100000	180 Kilobasen	300 Kilobasen

[0099] Fragmente können aus über eine Vielfalt von Wegen aus einem Zielpolynucleotid erzeugt werden, einschließlich sogenannter „direkter“ Ansätze, wo man versucht Sets von Fragmenten zu erzeugen, welche das Zielpolynucleotid mit einer minimalen Überlappung abdecken und sogenannten „Shotgun“-Ansätzen, wo zufällig überlappende Fragmente erzeugt werden. Bevorzugt werden „Shotgun“-Ansätze zur Erzeugung von Fragmenten wegen ihrer Einfachheit und ihrer innewohnenden Redundanz, verwendet. Zum Beispiel werden zufällig überlappende Fragmente, welche ein Zielpolynucleotid abdecken, in dem nachfolgenden üblichen „Shotgun“-Sequenzierungsprotokoll, wie zum Beispiel in Sambrook et al. (vorstehend zitiert) offenbart, erzeugt. Wie hier verwendet bedeutet „abdecken“ in diesem Zusammenhang, dass jeder Anteil der Zielpolynucleotidsequenz in jedem Größenbereich der erzeugten Fragmente, zum Beispiel allen Fragmenten einer Länge zwischen 100 und 200 Basenpaaren, vertreten ist. Kurz gesagt wird beginnend mit einem Zielpolynucleotid als Insert in einem geeigneten Clonierungsvektor, zum Beispiel einem λ -Phagen, der Vektor vermehrt, gereinigt und mit geeigneten Restriktionsenzymen gespalten, um etwa 10–15 μ g gereinigten Inserts zu ergeben. Typischer Weise ergibt das Protokoll etwa 500–1000 Subclone pro Mikrogramm Ausgangs-DNA. Das Insert wird durch präparative Gelelektrophorese von den Vektorfragmenten getrennt, durch übliche Verfahren aus dem Gel entfernt und in einem Standardpuffer, wie TE (Tris-EDTA), resuspendiert. Die zum Ausschneiden des Inserts aus dem Vektor ausgewählten Restriktionsenzyme belassen bevorzugt haftende Enden auf dem Insert, so dass das Insert in der Vorbereitung für die Herstellung zufällig überlappender Fragmente mit sich selbst ligiert werden kann. Wie in Sambrook et al. (vorstehend zitiert) erklärt ist, ergibt die zirkularisierte DNA in den nachstehend verwendeten Fragmentierungsverfahren eine bessere zufällige Verteilung von Fragmenten als lineare DNA. Nach der Ligierung des Inserts mit sich selbst, zum Beispiel mit T4-Ligase unter Verwendung üblicher Protokolle, wird das gereinigte ligierte Insert über ein Standardprotokoll fragmentiert, zum Beispiel Ultraschallbehandlung oder Spaltung mit DNase I in Anwesenheit von Mn^{++} . Nach der Fragmentierung werden die Enden der Fragmente repariert, zum Beispiel wie in Sambrook et al. (vorstehend zitiert) beschrieben, und die reparierten Fragmente werden dann der Größe nach unter Verwendung einer Gelelektrophorese aufgetrennt. Fragmente in dem Größenbereich von 300 – 500 Basenpaaren werden ausgewählt und durch übliche Mittel aus dem Gel eluiert und in einen einen Tag tragenden Vektor, wie vorstehend beschrieben, ligiert, um eine Bibliothek von Tag-Fragment-Konjugaten zu erzeugen.

[0100] Wie vorstehend beschrieben, wird eine Auswahl, welche einige tausend Tag-Fragment-Konjugate enthält, aus der Bibliothek entnommen und vermehrt, wonach die Tag-Fragment-Inserts aus dem Vektor ausgeschnitten werden und auf eine spezifische Hybridisierung mit den Tag-Komplementären auf Mikropartikeln, wie vorstehend beschrieben, vorbereitet werden. Abhängig von der Größe des Zielpolynucleotids können vielfache Proben aus der Tag-Fragment-Bibliothek erhoben und getrennt vermehrt, auf die Mikropartikel geladen und sequenziert werden. Die Zahl der ausgewählten Dubletten wird vom Anteil des in einer Probe repräsentierten Tag-Repertoires abhängen. (Die Wahrscheinlichkeit Triplette, drei unterschiedliche Polynucleotide mit dem selben Tag, oder mehr zu erhalten kann sicher ignoriert werden). Wie vorstehend erwähnt kann die Wahrscheinlichkeit von Dubletten in einer Probe aus der Poisson-Verteilung $p(\text{Dublette}) = m^2 e^{-m} / 2$, worin m der Anteil des Tag-Repertoires in der Probe ist, geschätzt werden. In der nachstehenden Tabelle V sind die Wahrscheinlichkeiten aufgeführt Dubletten in einer Probe zu erhalten für eine gegebene Tag-Größe, Probengröße und Diversität des Repertoires.

Tabelle V

Zahl der Wörter im Tag aus einem 8- Wort-Set	Größe des Tag- Repertoires	Größe der Probe	Anteil des Repertoires in der entnommenen Probe		Wahrscheinlichkeit einer Dublette
			der		
7	$2,1 \times 10^6$	3000	$1,43 \times 10^{-3}$		10^{-6}
8	$1,68 \times 10^7$	3×10^4	$1,78 \times 10^{-3}$		$1,6 \times 10^{-6}$
		3000	$1,78 \times 10^{-4}$		$1,6 \times 10^{-8}$
9	$1,34 \times 10^8$	3×10^5	$2,24 \times 10^{-3}$		$2,5 \times 10^{-6}$
		3×10^4	$2,24 \times 10^{-4}$		$2,5 \times 10^{-8}$
10	$1,07 \times 10^9$	3×10^6	$2,8 \times 10^{-3}$		$3,9 \times 10^{-6}$
		3×10^5	$2,8 \times 10^{-4}$		$3,9 \times 10^{-8}$

[0101] In jedem Fall werden die beladenen Mikropartikel dann auf einem Mikroskopobjektträger aus Glas verteilt und fixiert, bevorzugt über eine Avidin-Biotin-Kopplung. Bevorzugt werden wenigstens 15–20 Nucleotide von jedem der zufälligen Fragmente gleichzeitig über ein Einzelbasenverfahren sequenziert. Die Sequenz des Zielpolynucleotids wird dann durch Zuordnen der Teilsequenzen der zufälligen Fragmente mittels ihrer überlappenden Anteile unter Verwendung von Algorithmen, welche jenen ähnlich sind, die für die Zusammensetzung von Contigs verwendet werden, oder wie sie für die Sequenzierung durch Hybridisierung, offenbart in vorstehenden Quellen, entwickelt wurden, rekonstruiert.

Kits zur Ausführung des erfindungsgemäßen Verfahrens

[0102] Die Erfindung schließt Kits zur Durchführung der verschiedenen erfindungsgemäßen Ausführungsformen ein. Bevorzugt schließen erfindungsgemäße Kits ein Repertoire von an einen Festphasenträger angehängten Tag-Komplementären ein. Zusätzlich können erfindungsgemäße Kits das korrespondierende Repertoire von Tags einschließen, zum Beispiel als Primer zur Amplifikation der zu sortierenden Polynucleotide. Bevorzugt werden die Repertoires von Tag-Komplementären an Mikropartikel angehängt. Kits können auch geeignete Puffer für die enzymatische Umsetzung, Nachweis-Chemikalien, zum Beispiel fluoreszente oder chemilumineszente Tags und ähnliches, Anweisungen für die Anwendung, Enzyme für die Umsetzung, wie Ligasen, Polymerasen, Transferasen und so weiter, enthalten. In einer wichtigen Ausführungsform für die Sequenzierung können Kits auch Substrate, wie mit Avidin besetzte Mikroskopobjektträger, für die Fixierung beladener Mikropartikel zur Umsetzung einschließen.

Identifizierung von neuen Polynucleotiden in cDNA-Banken

[0103] Neue Polynucleotide in einer cDNA-Bank können durch die Erzeugung einer Bank aus cDNA-Molekülen, die an Mikropartikel angehängt sind, wie vorstehend beschrieben, identifiziert werden. Ein großer Anteil der Genbank oder sogar die ganze Genbank kann dann teilweise parallel sequenziert werden. Nach der Isolierung von mRNA und vielleicht einer Normalisierung der Populationen, wie durch Soares et al., Proc. Natl. Acad. Sci. 91 (1994): 9228–9232 oder ähnliche Quellen gelehrt, kann der folgende Primer für die Synthese des ersten Stranges mit einer reversen Transkriptase unter Verwendung von üblichen Protokollen an die Poly-A-Schwänze hybridisiert werden:

5' –mRNA– [A]_n –3'
 [T]₁₉ – (Primerstelle) –GG[W, W, W, C]₉ ACCAGCTGATC –5'

worin [W, W, W, C]₉ einen wie vorstehend beschriebenen Tag darstellt, „ACCAGCTGATC“ eine optionale Sequenz, welche in der doppelsträngigen Form eine Restriktionsstelle bildet und „Primer-Stelle“ eine allen Mitgliedern der Bibliothek gemeinsame Sequenz ist, welche später als Primer-Bindungsstelle zur Amplifizierung

von interessierenden Polynucleotiden durch PCR verwendet wird.

[0104] Nach der reversen Transkription und der Synthese des zweiten Stranges durch übliche Verfahren, werden die doppelsträngigen Fragmente wie vorstehend beschrieben in einen Clonierungsvektor eingefügt und amplifiziert. Von der amplifizierten Bibliothek wird dann eine Auswahl erhoben und die Auswahl amplifiziert. Die Clonierungsvektoren aus der amplifizierten Auswahl werden isoliert und die mit einem Tag versehenen cDNA-Fragmente ausgeschnitten und gereinigt. Nachdem der Tag mit einer Polymerase wie vorstehend beschrieben einzelsträngig gemacht wurde, werden die Fragmente methyliert und auf Mikropartikeln in Übereinstimmung mit der Erfindung sortiert. Bevorzugt wird der Clonierungsvektor wie vorstehend beschrieben so erzeugt, dass die mit einem Tag versehenen cDNAs mit einer Endonuclease, wie Fok I, ausgeschnitten werden können, welche nach Sortierung und Ligierung an Mikropartikel eine unmittelbare Sequenzierung durch das bevorzugte Einzelbasenverfahren erlauben wird.

[0105] Eine schrittweise Sequenzierung wird dann in Übereinstimmung mit der Erfindung gleichzeitig für die gesamte Bibliothek oder einen oder mehrere Anteile der Bibliothek durchgeführt bis eine ausreichende Zahl von Nucleotiden auf jeder cDNA zur einzigartigen Repräsentation im Genom des Organismus, aus welchem die Bibliothek abgeleitet wurde, identifiziert ist. Zum Beispiel wird von einer zufällig ausgewählten Sequenz von 14–15 Nucleotiden Länge erwartet, wenn die Bibliothek aus einer Säuger-mRNA abgeleitet ist, dass sie eine einzigartige Repräsentation unter den zwei- bis dreitausend Megabasen des typischen Säugergenoms besitzt. Natürlich wäre eine Identifizierung von weit weniger Nucleotiden für eine einzigartige Repräsentation in einer von Bakterien oder den niederen Organismen abgeleiteten Bibliothek ausreichend. Bevorzugt werden wenigstens 20–30 Nucleotide identifiziert, um eine einzigartige Repräsentation sicherzustellen und um die Erzeugung eines geeigneten Primers wie nachstehend beschrieben zu erlauben. Die geordneten Sequenzen können dann mit bekannten Sequenzen verglichen werden, um einzigartige cDNAs zu identifizieren.

[0106] Einzigartige cDNAs werden dann durch übliche Verfahren, zum Beispiel Erzeugung einer Sonde aus dem PCR-Amplicon, welches mit auf die Primerstelle gerichteten Primern und dem Anteil der cDNA, dessen Sequenz bestimmt worden ist, isoliert. Die Sonde kann verwendet werden, um die cDNA in einer Bibliothek unter Verwendung eines üblichen Durchmusterungsprotokolls zu identifizieren.

Beispiel I

Sortieren von aus pUC19 abgeleiteten multiplen Zielpolynucleotiden

[0107] Ein Gemisch von drei Zielpolynucleotid-Tag-Konjugaten wird wie folgt gewonnen: Zuerst werden die folgenden sechs Oligonucleotide synthetisiert und paarweise kombiniert, um Tag 1, Tag 2 und Tag 3 zu erzeugen:

5' -pTCGACC (w₁) (w₂) (w₃) (w₄) (w₅) (w₆) (w₇) (w₈) (w₁) A
GG (**) (**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

Tag 1

5' -pTCGACC (w₆) (w₇) (w₈) (w₁) (w₂) (w₆) (w₄) (w₂) (w₁) A
GG (**) (**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

Tag 2

5' -pTCGACC (w₃) (w₂) (w₁) (w₁) (w₅) (w₈) (w₈) (w₄) (w₄) A
GG (**) (**) (**) (**) (**) (**) (**) (**) (**) (**) TTCGAp-5'

Tag 3

worin „p“ ein Monophosphat anzeigt, die w_is eine in Tabelle I definierte Untereinheit darstellen und die Begriffe „(**)“ ihre jeweiligen Komplementäre darstellen. Ein pUC19 wird mit Sal I und Hind III gespalten, das große

Fragment wird gereinigt und getrennt mit den Tags 1, 2 und 3 ligiert, um jeweils pUC19-1, pUC19-2 und pUC19-3 zu erzeugen. Die drei Rekombinanten werden getrennt amplifiziert und isoliert, wonach pUC19-1 mit Hind III und Aat I gespalten wird, pUC19-2 mit Hind III und Ssp I gespalten wird und pUC19-3 mit Hind III und Xmn I gespalten wird. Die kleinen Fragmente werden unter Verwendung üblicher Protokolle isoliert, um drei doppelsträngige Fragmente von jeweils etwa 250, 375 und 575 Basenpaaren Länge zu ergeben und wobei jedes einen an 3' zurückgesetzten, dem Tag benachbarten Strang und einen mit glatten Enden versehenen oder an 3' hervorstehenden Strang am gegenüberliegenden Ende besitzt. Annähernd 12 nmol von jedem Fragment werden mit 5 Einheiten T4-DNA-Polymerase in dem vom Hersteller empfohlenen Reaktionspuffer, welcher 33 µM Desoxycytosintriphosphat enthält, gemischt. Man lässt das Reaktionsgemisch bei 37 °C über 30 Minuten inkubieren, wonach die Reaktion durch Setzen auf Eis gestoppt wird. Die Fragmente werden dann durch übliche Mittel gereinigt.

[0108] CPG-Mikropartikel (Partikelgröße 37–74 nm, Porengröße 500 Å, Pierce Chemical) werden mit dem durch Maskos und Southern, *Nucleic Acids Research* 20 (1992): 1679–1684; offenbarten Linker derivatisiert. Nach der Aufteilung in drei Proben werden die Komplementäre der Tags 1, 2 und 3 auf den Mikropartikeln unter Verwendung eines üblichen automatisierten DNA-Syntheseautomaten, zum Beispiel einem Modell 392 DNA Synthesizer (Applied Biosystems, Foster City, CA) synthetisiert. Annähernd 1 mg von jedem der unterschiedlich derivatisierten Mikropartikel wird in getrennte Gefäße platziert.

[0109] Die mit T4-DNA-Polymerase behandelten, aus pUC19-1, -2 und -3 ausgeschnittenen Fragmente werden in 50 µl des durch den Hersteller für die Taq-DNA-Ligase (New England Biolabs) empfohlenen Puffers resuspendiert. Das Gemisch wird dann gleichmäßig zwischen den drei Gefäßen aufgeteilt, welche jeweils die 1 mg der derivatisierten CPG-Mikropartikel enthalten. 5 Einheiten von Taq-DNA-Ligase werden jedem Gefäß hinzugefügt, wonach diese bei 55 °C über 15 Minuten inkubiert werden. Die Reaktion wird durch Setzen auf Eis gestoppt und die Mikropartikel werden einige Male durch wiederholte Zentrifugation und Resuspension in TE gewaschen. Schließlich werden die Mikropartikel in Nde I-Reaktionspuffer (New England Biolabs) resuspendiert, worin die angehängten Polynucleotide gespalten werden. Nach Trennung von den Mikropartikeln werden die durch Nde I-Spaltung freigesetzten Polynucleotidfragmente durch Inkubation mit Sequenase-DNA-Polymerase und fluoreszenzmarkiertem Thymidintriphosphat (Applied Biosystems, Foster City, CA) durch Fluoreszenz markiert. Die Fragmente werden dann getrennt auf einem nicht-denaturierenden Polyacrylamidgel unter Verwendung eines Applied Biosystems Modell 373 DNA Sequenzers untersucht.

Beispiel II

Paralleles Sequenzieren von SV40-Fragmenten

[0110] Ein Repertoire von aus neun aus der Tabelle I ausgewählten Untereinheiten von 4 Nucleotiden bestehenden 36-mer-Tags wird durch getrennte Synthetisierung von Tags und Tag-Komplementären durch einen Aufteil- und Mischansatz wie vorstehend beschrieben hergestellt. Das Repertoire wird so synthetisiert, dass es die Ligierung in einen mit Sma I/Hind III gespaltenen M13mp19 erlaubt. Somit beginnt wie in Beispiel I ein Set von Oligonucleotiden mit der Zugabe von A gefolgt von neun Runden einer Aufteil- und Misch-Synthese, worin das Oligonucleotid über Untereinheiten durch mit 3'-Phosphoramidit derivatisierte 4-mere, welche den Untereinheiten in Tabelle I entsprechen, verlängert wird. Die Synthese wird dann mit dem Hinzufügen von Nucleotid-um-Nucleotid der einen Hälfte der Sma I-Erkennungsstelle (GGG), zwei C und einem 5'-Monophosphat, zum Beispiel über das Phosphate-ON Reagent, welches von Clontech Laboratories (Palo Alto, CA) erhältlich ist, abgeschlossen. Das andere Set von Oligonucleotiden beginnt mit der Zugabe von drei C (Teil der Sma I-Erkennungsstelle) und zwei Gs gefolgt durch neun Runden einer Aufteil- und Misch-Synthese, worin das Oligonucleotid durch mit 3'-Phosphoramidit derivatisierte 4-mere, welche den Komplementären der Untereinheiten von Tabelle I entsprechen, verlängert wird. Die Synthese wird durch das Hinzufügen von Nucleotid-um-Nucleotid der Hind III-Erkennungsstelle und eines 5'-Monophosphats abgeschlossen. Nach Trennung von den Syntheseträgern werden die Oligonucleotide unter Bedingungen gemischt, welche die Bildung der folgenden Duplexstränge erlauben:

$$5' - \text{pGGGCC} (w_1) (w_1) (w_1) (w_1) (w_1) (w_1) (w_1) (w_1) (w_1) \text{A} \\ \text{CCCGG} (**) (**) (**) (**) (**) (**) (**) (**) (**) \text{TTCGAp} - 5'$$

[0111] Das Gemisch von Duplexsträngen wird dann in einen mit Sma I/Hind III gespaltenen M13mp19 ligiert. Ein Repertoire von Tag-Komplementären wird auf CPG-Mikropartikeln wie vorstehend beschrieben synthetisiert.

[0112] Als nächstes wird der folgende Adapter hergestellt, welcher eine Fok I-Stelle und Teile von Eco RI und Sma I-Stellen enthält.

5' -pAATTCGGATGATGCATGCATCGACCC
GCCTACTACGTACGTAGCTGGGp-5'

Eco RI Fok I

Sma I

[0113] Der Adapter wird in den vorstehend beschriebenen mit Eco RI/Sma I gespaltenen M13 ligiert.

[0114] Getrennt wird SV40-DNA durch Ultraschallbehandlung dem in Sambrook et al. (vorstehend zitiert) dargestellten Protokoll folgend fragmentiert. Die sich ergebenden Fragmente werden unter Verwendung von Standardprotokollen repariert und der Größe nach getrennt. Fragmente im Bereich von 300–500 Basenpaaren werden ausgewählt und in den vorstehend beschriebenen mit Sma I gespaltenen M13 ligiert, um eine Bibliothek von Fragment-Tag-Konjugaten zu erzeugen, welche dann amplifiziert wird. Eine einige tausend verschiedene Fragment-Tag-Konjugate enthaltende Auswahl wird aus der Bibliothek entnommen, weiter amplifiziert und die Fragment-Tag-Inserts durch Spaltung mit Eco RI und Hind III ausgeschnitten. Die ausgeschnittenen Fragment-Tag-Konjugate werden mit T4-DNA-Polymerase in der Anwesenheit von Desoxycytidintriphosphat wie in Beispiel I beschrieben behandelt, um die Oligonucleotidtags für die spezifische Hybridisierung mit den CPG-Mikropartikeln freizulegen.

[0115] Nach Hybridisierung und Ligierung wie in Beispiel I beschrieben werden die beladenen Mikropartikel mit Fok I behandelt, um einen hervorstehenden Strang von 4 Nucleotiden mit vorbestimmter Sequenz zu erzeugen. Ein 10:1-Gemisch (Sonde 1:Sonde 2) der folgenden Sonden wird an die Polynucleotide auf Mikropartikeln ligiert.

Sonde 1

FAM- ATCGGATGAC
 TAGCCTACTGAGCT

Sonde 2

Biotin- ATCGGATGAC
 TAGCCTACTGAGCT

[0116] FAM stellt einen Fluoreszenzfarbstoff dar, welcher über einen Aminophosphatlinker, erhältlich von Applied Biosystems (Aminolinker), an das 5'-Hydroxyl des oberen Stranges der Sonde 1 angehängt ist. Das Biotin kann auch durch eine Aminolinker-Gruppe angehängt werden und kann optional über Polyethylenoxidlinker weiter verlängert werden, zum Beispiel Jaschke et al. (vorstehend zitiert).

[0117] Die beladenen Mikropartikel werden dann auf der Oberfläche eines mit Avidin besetzten Glasobjektträgers angeordnet, an welchen und von welchem Reagenzien und Waschlösungen aufgebracht und entfernt werden können. Der mit Avidin besetzte Objektträger mit den angehängten Mikropartikeln wird mit einem abtastenden Fluoreszenzmikroskop untersucht (zum Beispiel ein mit einem Newport Modell PM500-C Bewegungskontrollgerät, einem Spectra-Physics Modell 2020 Argon-Laser, welcher einen 488 nm Anregungsstrahl erzeugt und einem 520 nm Langpassemissionsfilter ausgestatteten Zeiss Axioskop oder einem ähnlichen Gerät). Der Anregungsstrahl und die Fluoreszenzemissionen werden jeweils durch die gleiche Objektivlinse abgegeben und gesammelt. Der Anregungsstrahl und die gesammelte Fluoreszenz werden durch einen dichroitischen Spiegel getrennt, welcher die gesammelte Fluoreszenz durch eine Serie von Bandpassfiltern und auf den beobachteten Fluorophoren entsprechende Photonenzählgeräte richtet, zum Beispiel umfassend Hamamatsu Modell 9403-02 Photomultiplier, einen Stanford Research Systems Modell SR445 Verstärker und einen Modell SR430 Multikanalfrequenzteiler und einen digitalen Computer, zum Beispiel einem auf dem 486 basierendem Computer. Der Computer erzeugt eine zweidimensionale Karte des Objektträgers, welche die Positionen der Mikropartikel aufzeichnet.

[0118] Nach Spaltung mit Fok I zur Entfernung der anfänglichen Sonde, durchlaufen die Polynucleotide auf den angehängten Mikropartikeln 20 Zyklen von Sondenligierung, Waschen, Nachweis, Spaltung und Waschen in Übereinstimmung mit der nachstehend beschriebenen bevorzugten Sequenzierungsverfahrensweise. Innerhalb jedes Nachweisschrittes zeichnet das Abtastungssystem die der bei jedem Mikropartikel identifizierten Base entsprechende Fluoreszenzemission auf. Nachstehende Reaktionen und Waschschritte werden im All-

gemeinen, solange dies nicht anders angegeben ist, mit den vom Hersteller (New England Biolabs) für die verwendeten Enzyme empfohlenen Puffern durchgeführt. Standardpuffer werden auch in Sambrook et al. (vorstehend zitiert) beschrieben.

[0119] Die folgenden vier Sets von gemischten Sonden werden für die Hinzufügung an die Zielpolynucleotide bereitgestellt:

TAMRA- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGANNN

FAM- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGCNNN

ROX- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGGNNN

JOE- ATCGGATGACATCAAC
TAGCCTACTGTAGTTGTNNN

[0120] Worin TAMRA, FAM, ROX und JOE spektral auflösbare Fluoreszenzfarbstoffe sind, welche über den Aminolinker II (alle erhältlich von Applied Biosystems, Inc., Foster City, California) angehängt sind; die fett gedruckten Nucleotide sind Erkennungsstellen für die Fok I-Endonuclease und „N“ stellt jegliches der vier Nucleotide A, C, G und T dar. TAMRA (Tetramethylrhodamin), FAM (Fluorescein), ROX (Rhodamin X) und JOE (2',7'-Dimethoxy-4',5'-Dichlorofluorescein) und das Verbinden mit den Oligonucleotiden wird auch in Fung et al., U.S. Patent 4,855,225 beschrieben.

[0121] Die vorstehenden Sonden werden in annähernd 5 molarem Überschuss gegenüber den Enden des Zielpolynucleotids wie folgt inkubiert: Die Sonden werden über 60 Minuten bei 16 °C mit 200 Einheiten T4-DNA-Ligase und den verankerten Zielpolynucleotiden in T4-DNA-Ligasepuffer inkubiert; nach dem Waschen wird das Zielpolynucleotid dann mit 100 Einheiten T4-Polynucleotidkinase in dem vom Hersteller empfohlenen Puffer über 30 Minuten bei 37 °C inkubiert, gewaschen und wieder über 30 Minuten bei 16 °C mit 200 Einheiten T4-DNA-Ligase und dem verankerten Zielpolynucleotid in T4-DNA-Ligasepuffer inkubiert. Waschen wird durch aufeinanderfolgendes Laufen lassen der Waschpuffervolumina, zum Beispiel TE, offenbart in Sambrook et al. (vorstehend zitiert) über den Objektträger bewerkstelligt. Nach dem Zyklus der Ligierung-Phosphorylierung-Ligierung und einem letzten Waschen, werden die angehängten Mikropartikel auf die Anwesenheit der Fluoreszenzmarkierung hin abgetastet, wobei deren Positionen und Charakteristiken durch das Abtastungssystem aufgezeichnet werden. Das markierte Zielpolynucleotid, das heißt, der ligierte Komplex, wird dann mit 10 Einheiten von Fok I in dem vom Hersteller empfohlenen Puffer über 30 Minuten bei 37 °C inkubiert, gefolgt durch Waschen in TE. Als ein Ergebnis wird das Zielpolynucleotid auf jedem Strang um ein Nucleotid gekürzt und ist für den nächsten Zyklus von Ligierung und Spaltung bereit. Der Vorgang wird fortgesetzt bis zwanzig Nucleotide identifiziert sind.

```

Program minxh
c
c
c
integer*2 sub1(6),mset1(1000,6),mset2(1000,6)
dimension nbase(6)

c
c
write(*,*) "GEBEN SIE DIE LÄNGE DER UNTEREINHEITEN EIN"
read(*,100)nsub
format(i1)
open(1,file='sub4.dat',form='formatted',status='new')

c
c
nset=0
do 7000 m1=1,3
    do 7000 m2=1,3
        do 7000 m3=1,3
            do 7000 m4=1,3
                sub1(1)=m1
                sub1(2)=m2
                sub1(3)=m3
                sub1(4)=m4

c
c
ndiff=3

c
c
Erzeuge ein Set aus Untereinheiten das sich
von sub1 durch mindestens ndiff Nucleotide unterscheidet.
Sichere in mset1.

c
c
jj=1
do 900 j=1,nsub
    mset1(1,j)=sub1(j)

900
c
c
do 1000 k1=1,3
    do 1000 k2=1,3
        do 1000 k3=1,3
            do 1000 k4=1,3

c
c
nbase(1)=k1
nbase(2)=k2
nbase(3)=k3
nbase(4)=k4

c
n=0

```

```

do 1200 j=1,nsub
  if(sub1(j).eq.1 .and. nbase(j).ne.1 .or.
1    sub1(j).eq.2 .and. nbase(j).ne.2 .or.
3    sub1(j).eq.3 .and. nbase(j).ne.3) then
    n=n+1
    endif
1200  continue
c
c
c    if(n.ge.ndiff) then
c
c
c        Wenn die Anzahl der Fehlpaarungen
c        größer oder gleich ndiff ist
c        dann nimm die Untereinheit
c        in Matrix mset auf
c
c
c        jj=jj+1
c        do 1100 i=1,nsub
1100      mset1(jj,i)=nbase(i)
c        endif
c
c
1000  continue
c
c
c        do 1325 j2=1,nsub
c        mset2(1,j2)=mset1(1,j2)
1325      mset2(2,j2)=mset1(2,j2)
c
c
c
c        Vergleiche Untereinheit 2 von mset1
c        mit jeder darauffolgenden Untereinheit
c        in mset1; d.h. 3, 4, 5, ... usw.
c        Sichere diejenigen mit Fehlpaarungen
c        .ge. ndiff in Matrix mset2
c        beginnend an Position 2.
c
c        Anschließend transferiere den Inhalt
c        von mset2 in mset1 und
c        beginne die Vergleiche nochmal
c        diesmal beginnend mit Untereinheit 3.
c        Setze fort bis alle Untereinheiten
c        Vergleichen unterzogen wurden.
c
c
c
c
c        npass=0
c
c
1700  continue
      kk=npass+2
      npass=npass+1
c
c
      do 1500 m=npass+2,jj

```



```

n=0
do 1600 j=1,nsub
  if (mset1(npass+1,j).eq.1.and.mset1(m,j).ne.1.or.
2      mset1(npass+1,j).eq.2.and.mset1(m,j).ne.2.or.
2      mset1(npass+1,j).eq.3.and.mset1(m,j).ne.3) then
    n=n+1
    endif
1600    continue
    if (n.ge.ndiff) then
      kk=kk+1
      do 1625 i=1,nsub
1625        mset2(kk,i)=mset1(m,i)
      endif
1500    continue
c
c      kk ist die Anzahl der Untereinheiten
c      die in mset2 gespeichert sind.
c
c      Transferiere den Inhalt von mset2
c      in mset1 für den nächsten Durchlauf.
c
      do 2000 k=1,kk
        do 2000 m=1,nsub
2000          mset1(k,m)=mset2(k,m)
        if (kk.lt.jj) then
          jj=kk
          goto 1700
        endif
c
c      nset=nset+1
      write(1,7009)
7009    format(/)
      do 7008 k=1,kk
7008        write(1,7010) (mset1(k,m),m=1,nsub)
7010    format(4i1)
      write(*,*)
      write(*,120) kk,nset
120    format(1x,'Subunits in set=',i5,2x,'Set No=',i5)
7000    continue
      close(1)
c
c
end

```

SEQUENZPROTOKOLL

(1) Allgemeine Information

- (i) ANMELDER: Sydney Brenner
- (ii) TITEL DER ERFINDUNG: Molekulares Markierungssystem
- (iii) ZAHL DER SEQUENZEN: 5
- (iv) KORRESPONDENZADRESSE:
 - (A) ADRESSAT: Stephen C. Macevicz, Lynx Therapeutics, Inc.
 - (B) STRASSE: 3832 Bay Center Place
 - (C) STADT: Hayward
 - (D) STAAT: California
 - (E) LAND: USA.
 - (F) ZIP: 94545
- (v) IM COMPUTER LESBARE FORM:
 - (A) MEDIUMTYP: 3,5 Zoll Diskette
 - (B) COMPUTER: IMB-kompatibel
 - (C) BETRIEBSSYSTEM: Windows 3.1 / DOS 5.0
 - (D) SOFTWARE: Microsoft Word für Windows, Version 2.0
- (vi) DATEN DER AKTUELLEN ANMELDUNG:
 - (A) ANMELDUNGSNUMMER:
 - (B) ANMELDETAG:
 - (C) KLASSIFIZIERUNG:
- (vii) DATEN FRÜHERER ANMELDUNG:
 - (A) ANMELDUNGSNUMMER: 08/322,348
 - (B) ANMELDETAG: 13. Okt. 94
- (viii) DATEN FRÜHERER ANMELDUNG:
 - (A) ANMELDUNGSNUMMER: 08/358,810
 - (B) ANMELDETAG: 19. Dez. 94
- (ix) ANWALTS / VERTRETER INFORMATION:
 - (A) NAME: Stephen C. Macevicz
 - (B) REGISTRIERUNGSNUMMER: 30,285
 - (C) REFERENZ/VERFAHRENSNUMMER: cbd3wo
- (x) TELEKOMMUNIKATIONSINFORMATION:
 - (A) TELEFON: (510) 670-9365
 - (B) TELEFAX: (510) 670-9302

(2) INFORMATION FÜR SEQ ID NO: 1:

- (i) SEQUENZCHARAKTERISTIK:
 - (A) LÄNGE: 38 Nucleotide
 - (B) ART: Nucleinsäure
 - (C) STRANGFORM: einzeln
 - (D) TOPOLOGIE: linear

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

GAGGATGCCT TTATGGATCC ACTCGAGATC CCAATCCA

38

(2) INFORMATION FÜR SEQ ID NO: 2:

(i) SEQUENZCHARAKTERISTIK:

- (A) LÄNGE: 26 Nucleotide
- (B) ART: Nucleinsäure
- (C) STRANGFORM: einzeln
- (D) TOPOLOGIE: linear

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

AATTCGGATG ATGCATGCAT CGACCC

26

(2) INFORMATION FÜR SEQ ID NO: 3:

(i) SEQUENZCHARAKTERISTIK:

- (A) LÄNGE: 14 Nucleotide
- (B) ART: Nucleinsäure
- (C) STRANGFORM: einzeln
- (D) TOPOLOGIE: linear

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

TAGCCTACTG AGCT

14

(2) INFORMATION FÜR SEQ ID NO: 4:

(i) SEQUENZCHARAKTERISTIK:

- (A) LÄNGE: 16 Nucleotide
- (B) ART: Nucleinsäure
- (C) STRANGFORM: einzeln
- (D) TOPOLOGIE: linear

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

ATCGGATGAC ATCAAC

16

(2) INFORMATION FÜR SEQ ID NO: 5:

(i) SEQUENZCHARAKTERISTIK:

- (A) LÄNGE: 11 Nucleotide
- (B) ART: Nucleinsäure
- (C) STRANGFORM: einzeln
- (D) TOPOLOGIE: linear

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

ACCAGCTGAT C

11

Patentansprüche

1. Verfahren zum Sortieren von Polynucleotiden aus einer Population von Polynucleotiden auf einem oder mehreren Festphasenträgern, wobei das Verfahren die Schritte umfasst:

- (a) Binden eines Oligonucleotidtags aus einem Repertoire von Tags, ausgewählt aus einem minimal kreuzhybridisierenden Set, die von 12 bis 60 Nucleotiden reichen und sich voneinander durch mindestens 3 Nucleotide unterscheiden, an jedes Polynucleotid einer Population von Polynucleotiden, so dass alle unterschiedlichen Polynucleotide der Population unterschiedliche Oligonucleotidtags gebunden haben, um eine Population von mit Tags versehenen Polynucleotiden zu bilden; und
- (b) Sortieren der mit Tags versehenen Polynucleotide durch spezifisches Hybridisieren der Oligonucleotidtags der mit Tags versehenen Polynucleotide mit ihren entsprechenden Komplementären, wobei die jeweiligen Komplementäre als einheitliche Population in räumlich getrennten Regionen eines oder mehrerer Festphasenträger gebunden sind.

2. Verfahren nach Anspruch 1, wobei jeder Oligonucleotidtag des Repertoires eine Vielzahl von Untereinheiten umfasst und jede Untereinheit dieser Vielzahl aus einem Oligonucleotid mit einer Länge von 3 bis 6 Nucleotiden oder 3 bis 6 Basenpaaren besteht, wobei die Untereinheiten aus einem minimal kreuzhybridisierenden Set ausgewählt sind.

3. Verfahren nach Anspruch 1 oder 2, wobei die ein oder mehreren Festphasenträger Mikropartikel sind.

4. Verfahren nach einem der Ansprüche 1 bis 3, wobei der Oligonucleotidtag und dessen Komplementär einzelsträngige Oligonucleotide sind.

5. Verfahren nach Anspruch 3 oder 4, wobei der Mikropartikel aus der Gruppe bestehend aus Glasmikropartikeln, magnetischen Kügelchen, Glycidalmethacrylat-Mikropartikeln und Polystyrol-Mikropartikeln ausgewählt wird.

6. Verfahren nach einem der Ansprüche 3 bis 5, wobei der Mikropartikel einen Durchmesser zwischen 1 und 100 µm hat.

7. Verfahren nach Anspruch 1 oder 2, wobei der Festphasenträger ein planares Substrat mit einer Vielzahl räumlich getrennter Oberflächenregionen ist, an die einheitliche Populationen der Komplementäre gebunden sind.

8. Verfahren nach Anspruch 7, wobei verschiedene der vielen räumlich getrennten Oberflächenregionen einheitliche Populationen von verschiedenen Komplementären aufweisen.

9. Verfahren nach einem der Ansprüche 1 bis 8, wobei die Oligonucleotidtags perfekt zusammenpassende Duplexstränge gleicher Stabilität bilden.

10. Verfahren nach einem der Ansprüche 1 bis 9, wobei:
 der Bindungsschritt (a) die Erhebung einer Auswahl aus der Population mit Tags versehener Polynucleotide mit einschließt, um eine Subpopulation mit Tags versehener Polynucleotide zu bilden, wobei die Subpopulation mit Tags versehener Polynucleotide vor dem Sortierschritt (b) amplifiziert wird; und
 Sortierschritt (b) die Bildung perfekt zusammenpassender Duplexstränge zwischen den Oligonucleotidtags und ihren entsprechenden Komplementären auf einem oder mehreren Festphasenträgern mit einschließt.

11. Verfahren nach einem der Ansprüche 1 bis 10 zur Klassifizierung einer Population von Polynucleotiden, wobei das Verfahren außerdem die Bestimmung der Nucleotidsequenz eines Teils jedes Polynucleotids nach Sortierschritt (b) einschließt; und die Klassifizierung der Population durch die Häufigkeitsverteilung der Teilsequenzen der Polynucleotide.

12. Verfahren nach einem der Ansprüche 1 bis 10 zur Bestimmung der Nucleotidsequenz eines Zielpolynucleotids, wobei die Population der Polynucleotide eine Vielzahl von Fragmenten, die aus dem Zielpolynucleotid erzeugt wurden, mit einschließt, wobei die Vielzahl der Fragmente das Zielpolynucleotid mit abdeckt; und das Verfahren ferner das Bestimmen der Nucleotidsequenz eines Teils eines jeden Fragments aus der Vielzahl nach Sortierschritt (b) mit einschließt; und Bestimmen der Nucleotidsequenz des Zielpolynucleotids durch Zuordnen der Sequenzen der Fragmente.

13. Verfahren nach Anspruch 12, wobei die Vielzahl der Fragmente zufällig überlappende Fragmente der Zielpolynucleotide einschließt.

14. Verfahren nach Anspruch 12 oder 13, wobei der Festphasenträger eine Vielzahl von Mikropartikeln ist, an die jeweils eine einheitliche Population der Komplementäre gebunden sind, und wobei der Teil jedes Fragments 12 bis 25 Nucleotide einschließt.

15. Verfahren nach einem der Ansprüche 1 bis 14, wobei nach dem Sortierschritt (b) die Vielzahl der Mikropartikel an einem planaren Substrat fixiert werden.

16. Verfahren nach Anspruch 15, wobei die Vielzahl der Mikropartikel zufällig auf der Oberfläche des planaren Substrats angeordnet sind mit einer Dichte von etwa 1000 Mikropartikeln bis etwa 100 000 Mikropartikeln pro cm².

17. Verfahren zum Binden einer mit Tags versehenen doppelsträngigen DNA an einen Festphasenträger, wobei das Verfahren die Schritte umfasst:
Bereitstellen eines 5'-Stranges an einem Ende einer mit Tags versehenen doppelsträngigen DNA, wobei der 5'-Strang aus Nucleotiden besteht, die aus einer ersten Gruppe ausgewählt sind, bestehend aus drei oder weniger Arten von Nucleotiden, und wobei der 5'-Strang eine vorgegebene Länge hat, die definiert ist durch ein zweites Nucleotid einer Art, die nicht in der ersten Gruppe vorkommt, wobei das zweite Nucleotid am 3'-Ende des 5'-Stranges positioniert ist;
Aussetzen des Endes der mit Tags versehenen doppelsträngigen DNA einer DNA-Polymerase mit 3'→5'-Exonuclease-Aktivität in Gegenwart von Nucleosidtriphosphaten des zweiten Nucleotids, so dass der 5'-Strang bis zur Position des zweiten Nucleotids abgebaut wird, um einen einzelsträngigen Abschnitt herzustellen, und Binden der mit Tags versehenen doppelsträngigen DNA an einen Festphasenträger durch spezifisches Hybridisieren des einzelsträngigen, mit Tags versehenen Abschnitts mit seinem Komplementär, das an den Festphasenträger gebunden ist.

18. Verfahren nach Anspruch 17, wobei die DNA-Polymerase T4-DNA-Polymerase ist.

19. Verfahren zum Binden von Oligonucleotidtags an Polynucleotide in einer Population, wobei das Verfahren die Schritte umfasst:
Bereitstellen eines Repertoires von Oligonucleotidtags, ausgewählt aus einem minimal kreuzhybridisierenden Set, die von 12 bis 60 Nucleotiden reichen, und die sich untereinander durch mindestens drei Nucleotide unterscheiden, mit einer Größe, die größer ist als die einer Population von Polynucleotiden;
Ligieren des Repertoires an Oligonucleotidtags an die Polynucleotid-Population, um eine Population von mit Tags versehenen Polynucleotidkonjugaten zu bilden; und
Erhebung einer Auswahl von der Population der mit Tags versehenen Polynucleotidkonjugate, so dass alle mit Tags versehenen Polynucleotidkonjugate in der Probe, die unterschiedliche Polynucleotide haben, unterschiedliche Oligonucleotidtags haben.

20. Verfahren nach Anspruch 19, wobei die Größe des Repertoires an Oligonucleotidtags und die Größe der Probe so gewählt sind, dass mindestens 90% der mit Tags versehenen Polynucleotidkonjugate einzigartig sind.

21. Stoffzusammensetzung umfassend:

einen Festphasenträger mit einem oder mehreren räumlich abgegrenzten Bereichen; und eine einzigartige Population an Oligonucleotidtag-Komplementären, die in mindestens einem der ein oder mehreren räumlich abgegrenzten Regionen kovalent an den Festphasenträger gebunden sind, wobei die Oligonucleotidtag-Komplementäre ausgewählt sind aus einem minimal kreuzhybridisierenden Set von Oligonucleotiden mit einer Länge im Bereich von 12 bis 60 Nucleotiden, welche sich jeweils durch mindestens zwei Nucleotide von einander unterscheiden.

22. Zusammensetzung nach Anspruch 21, wobei jedes der Oligonucleotidtag-Komplementäre eine Vielzahl von Untereinheiten umfasst, so dass jede Untereinheit dieser Vielzahl aus einem Oligonucleotid mit einer Länge von drei bis sechs Nucleotiden besteht, wobei die Untereinheiten aus einem minimal kreuzhybridisierenden Set ausgewählt werden.

23. Stoffzusammensetzung nach Anspruch 22, wobei die Vielzahl der Untereinheiten in einem Bereich von 4 bis 10 liegt.

24. Stoffzusammensetzung nach einem der Ansprüche 21 bis 23, wobei der Festphasenträger eine Vielzahl von Mikropartikeln ist, von denen jedes eine einzelne, räumlich getrennte Region aufweist.

25. Zusammensetzung nach einem der Ansprüche 21 bis 24, wobei die Oligonucleotidtag-Komplementäre zusätzlich ausgewählt werden, um perfekt passende Duplexstränge gleicher Stabilität zu bilden.

26. Zusammensetzung nach einem der Ansprüche 21 bis 23, wobei die räumlich getrennten Regionen spezifische Orte auf einem Array von Syntheseorten auf einem einzelnen Träger sind.

27. Stoffzusammensetzung umfassend ein Gemisch aus Mikropartikeln, wobei jedes Mikropartikel mit Tags versehene Polynucleotide einer Population aufweist, die an Tag-Komplementäre auf dem Mikropartikel gebunden sind, so dass alle unterschiedlichen mit Tags versehenen Polynucleotide der Population an verschiedene Mikropartikel gebunden sind, wobei die mit Tags versehenen Polynucleotide ausgewählt sind aus der Gruppe bestehend aus cDNAs und genomischen DNA-Fragmenten; und die Tag-Komplementäre ausgewählt sind aus einem minimal kreuzhybridisierenden Set von Oligonucleotiden mit einer Länge im Bereich von 12 bis 60 Nucleotiden, wobei sich alle durch mindestens zwei Nucleotide voneinander unterscheiden.

28. Zusammensetzung nach Anspruch 27, wobei etwa 10^5 der mit Tags versehenen Polynucleotide an jedes der Mikropartikel gebunden sind.

29. Zusammensetzung nach Anspruch 27 oder 28, wobei die Population eine Größe im Bereich von zehn- bis hunderttausend hat.

30. cDNA-Bibliothek die eine Vielzahl mit Tags versehener cDNAs umfaßt, die an Mikropartikel gebunden sind, so dass verschiedene, mit Tags versehene cDNAs der Bibliothek an verschiedene Mikropartikel gebunden sind, wobei

(a) alle unterschiedlichen Polynucleotide unterschiedliche Oligonucleotidtags gebunden haben, und
(b) das Repertoire an Oligonucleotidtags ausgewählt ist aus der Gruppe von Oligonucleotiden der Form: $S_1, S_2, S_3 \dots S_n$, wobei S_1 bis S_n jeweils eine Untereinheit ist, die aus einem Oligonucleotid mit einer Länge von 3 bis 6 Nucleotiden besteht und aus einem minimal kreuzhybridisierenden Set ausgewählt ist, so dass sich jede Untereinheit von jeder anderen Untereinheit durch mindestens zwei Nucleotide unterscheidet, und n im Bereich von 4 bis 10 liegt.

31. cDNA-Bibliothek nach Anspruch 30, wobei die Vielzahl zwischen 10 000 und 100 000 liegt.

32. Bibliothek genomischer DNA-Fragmente, die eine Vielzahl mit Tags versehener genomischer DNA-Fragmente umfaßt, die an Mikropartikel gebunden sind, so dass alle unterschiedlichen, mit Tags versehenen genomischen DNA-Fragmente der Bibliothek an unterschiedliche Mikropartikel gebunden sind, wobei
(a) alle unterschiedlichen Polynucleotide unterschiedliche Oligonucleotidtags gebunden haben, und
(b) das Repertoire an Oligonucleotidtags ausgewählt ist aus der Gruppe von Oligonucleotiden der Form: $S_1, S_2, S_3 \dots S_n$, wobei S_1 bis S_n jeweils eine Untereinheit ist, die aus einem Oligonucleotid mit einer Länge von 3 bis 6 Nucleotiden besteht und aus einem minimal kreuzhybridisierenden Set ausgewählt ist, so dass sich jede Untereinheit von jeder anderen Untereinheit durch mindestens zwei Nucleotide unterscheidet, und n im Bereich von 4 bis 10 liegt.

33. Bibliothek nach Anspruch 32, wobei die Vielzahl der genomischen DNA-Fragmente mindestens 1000 ist.

34. Verwendung eines Repertoires von Oligonucleotidtags in einem Verfahren zum Markieren oder Sortieren eines Moleküls oder einer Subpopulation von Molekülen einer Population von Molekülen durch spezifisches Hybridisieren der Oligonucleotidtags mit ihren entsprechenden Komplementären, wobei
(a) alle unterschiedlichen Polynucleotide unterschiedliche Oligonucleotidtags gebunden haben, und
(b) das Repertoire an Oligonucleotidtags ausgewählt ist aus der Gruppe von Oligonucleotiden der Form: $S_1, S_2, S_3 \dots S_n$, wobei S_1 bis S_n jeweils eine Untereinheit ist, die aus einem Oligonucleotid mit einer Länge von 3 bis 6 Nucleotiden besteht und aus einem minimal kreuzhybridisierenden Set ausgewählt ist, so dass sich jede Untereinheit von jeder anderen Untereinheit durch mindestens zwei Nucleotide unterscheidet, und n im Bereich von 4 bis 10 liegt.

35. Verwendung nach Anspruch 34, wobei die Oligonucleotidtags Inserts eines Clonierungsvektors sind.

36. Verwendung nach Anspruch 34 oder 35 zum Binden eines Moleküls oder einer Subpopulation von Molekülen an einen Festphasenträger.

37. Verwendung nach Anspruch 36, wobei der Festphasenträger ein Mikropartikel ist, an den eine einheitliche Population der Komplementäre gebunden ist, oder ein planares Substrat mit einer Vielzahl räumlich getrennter Oberflächenregionen, an die eine einheitliche Population der Komplementäre gebunden ist.

38. Verwendung nach Anspruch 34 oder 35 zum Bestimmen der Nucleotidsequenz eines Zielpolynucleotids oder einer Vielzahl von Polynucleotiden.

39. Verwendung nach Anspruch 34 oder 35 zum Klassifizieren einer Population von Polynucleotiden.

40. Verwendung nach Anspruch 34 oder 35 zum Nachweis neuer cDNA-Moleküle in einer cDNA-Bibliothek.

41. Verwendung nach einem der Ansprüche 34 bis 40, wobei das Molekül oder die Subpopulation von Molekülen ein Polynucleotid oder eine Subpopulation von Polynucleotiden, ein Polypeptid oder eine Subpopulation von Polypeptiden oder ein lineares Polymer der Form:



ist, wobei:

L eine Phosphor (V)-Verknüpfungseinheit ist;

M eine geradkettige, cyclische oder verzweigte organisch molekulare Struktur ist, die 1 bis 20 Kohlenstoffatome und 0 bis 10 Heteroatome ausgewählt aus der Gruppe bestehend aus Sauerstoff, Stickstoff und Schwefel enthält, und

n im Bereich von 3 bis 100 liegt.

42. Verwendung nach Anspruch 41, wobei M ein Alkyl, Alkoxy, Alkenyl oder Aryl mit 1 bis 16 Kohlenstoffatomen oder ein Heterocyclus mit 3 bis 8 Kohlenstoffatomen und 1 bis 3 Heteroatomen, ausgewählt aus der Gruppe bestehend aus Sauerstoff, Stickstoff und Schwefel, ist.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

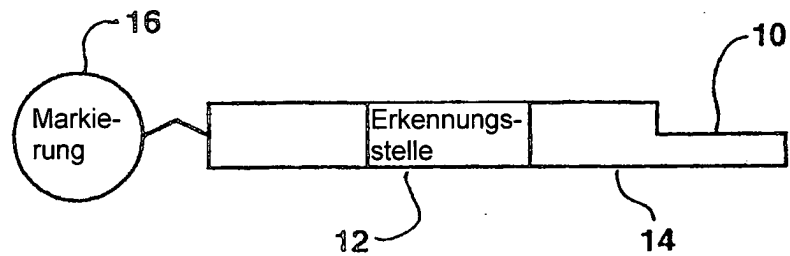


Fig. 1a

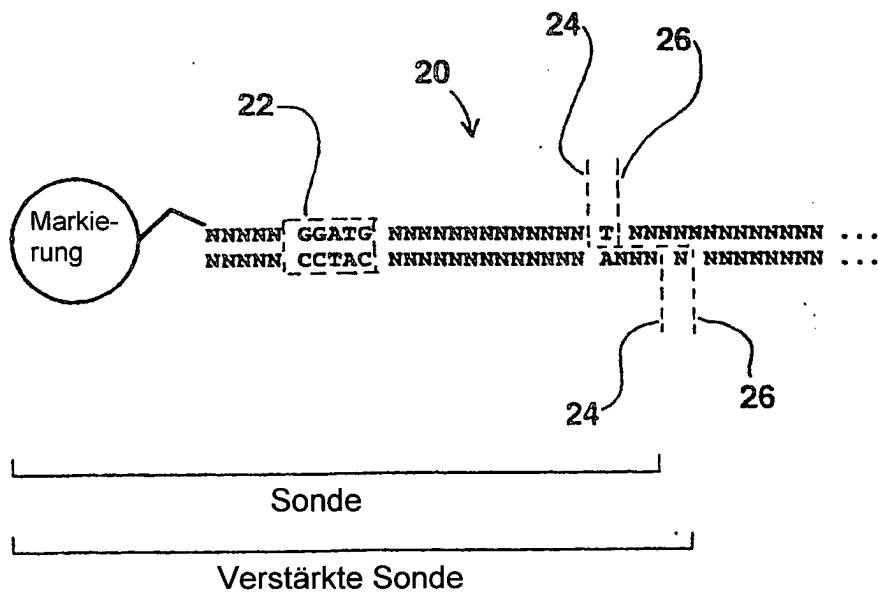


Fig. 2

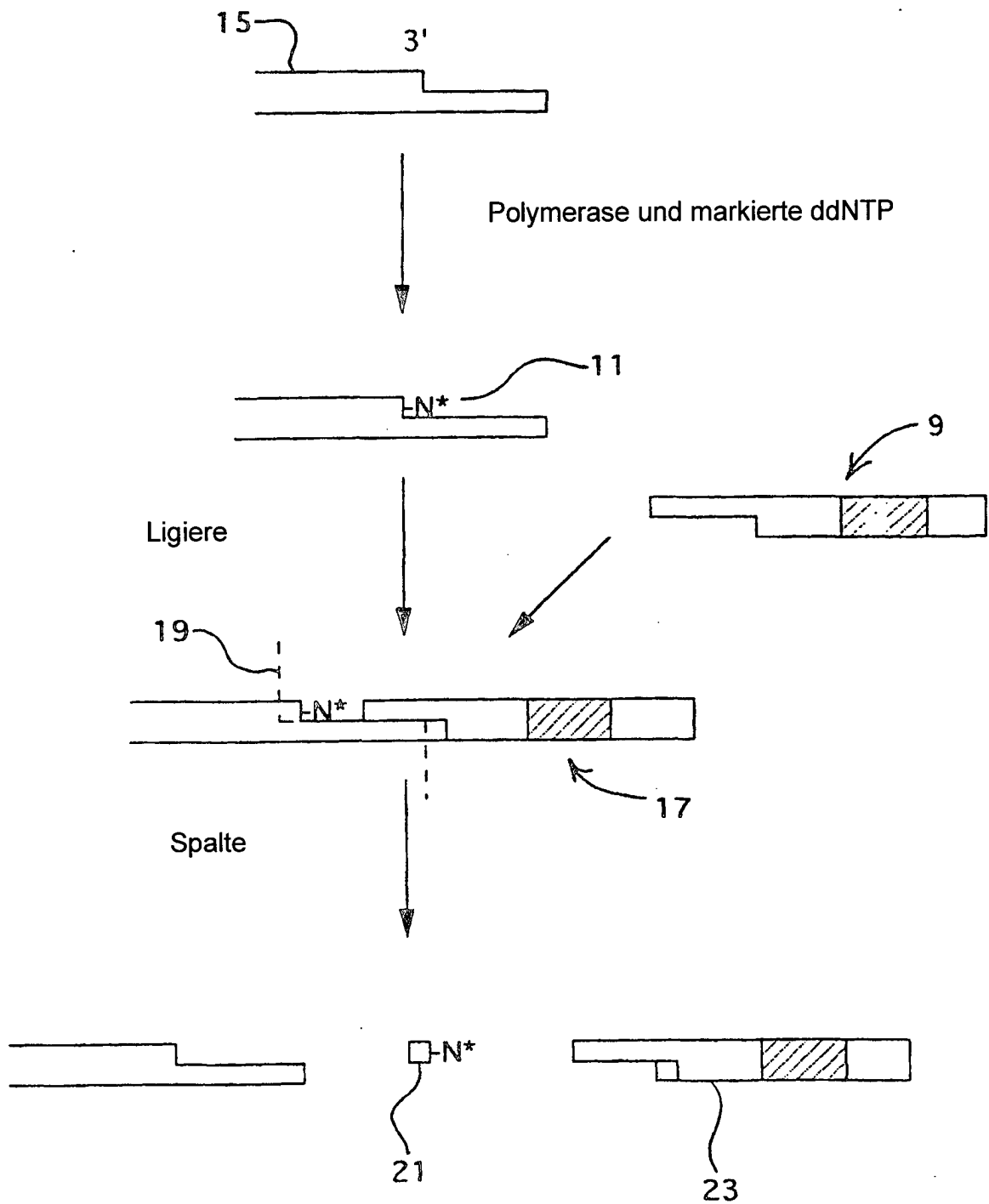
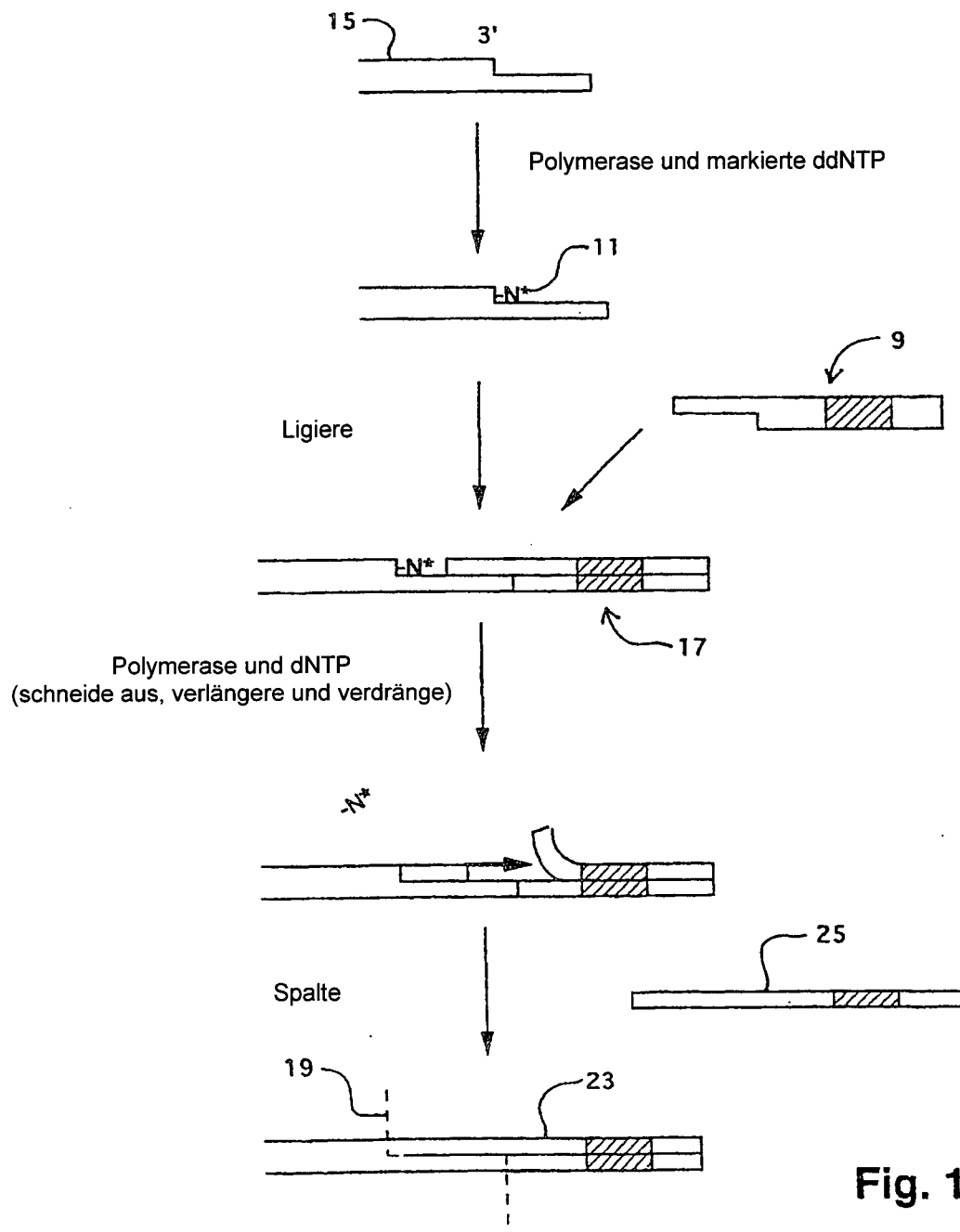


Fig. 1b



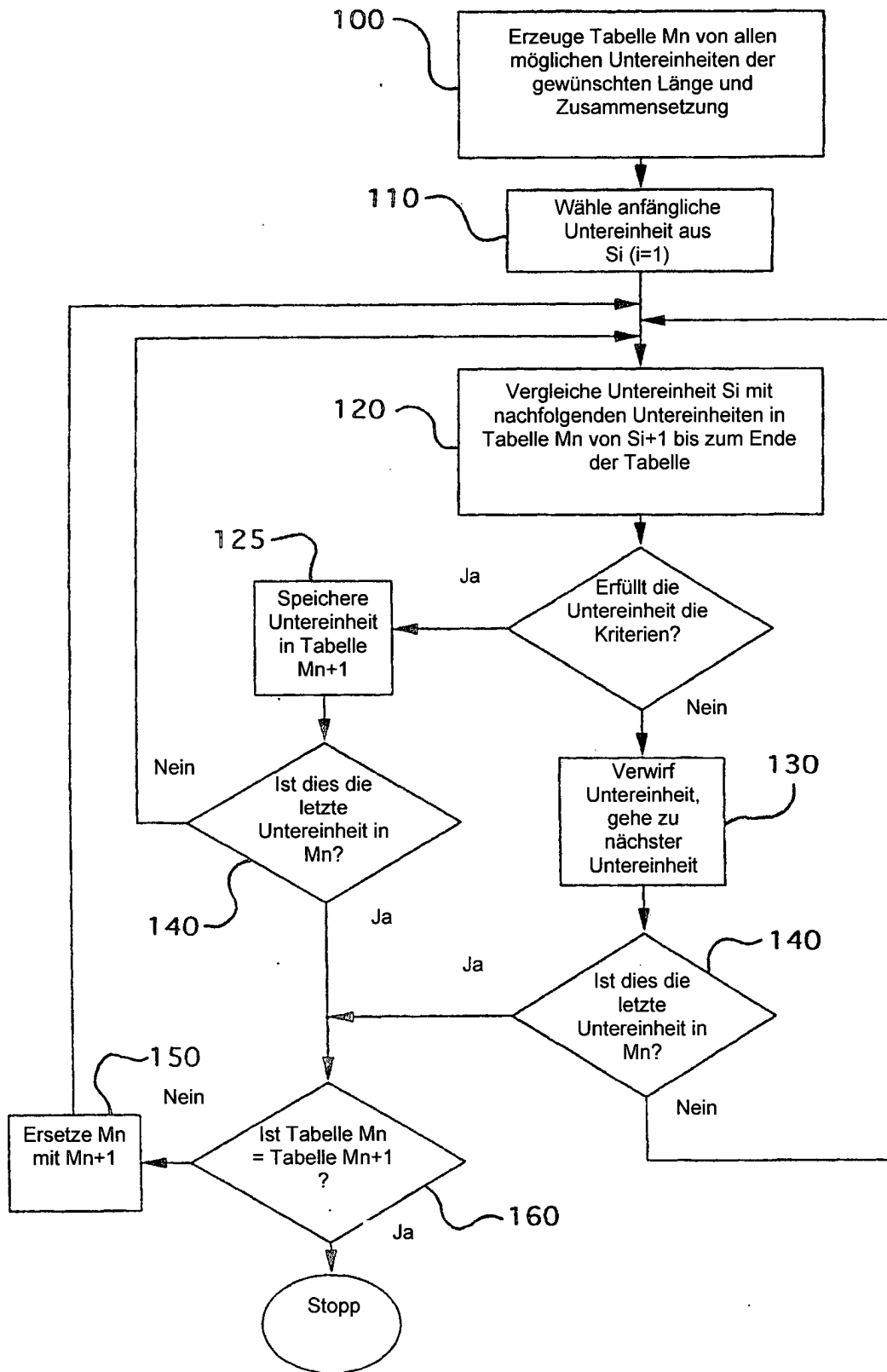


Fig. 3

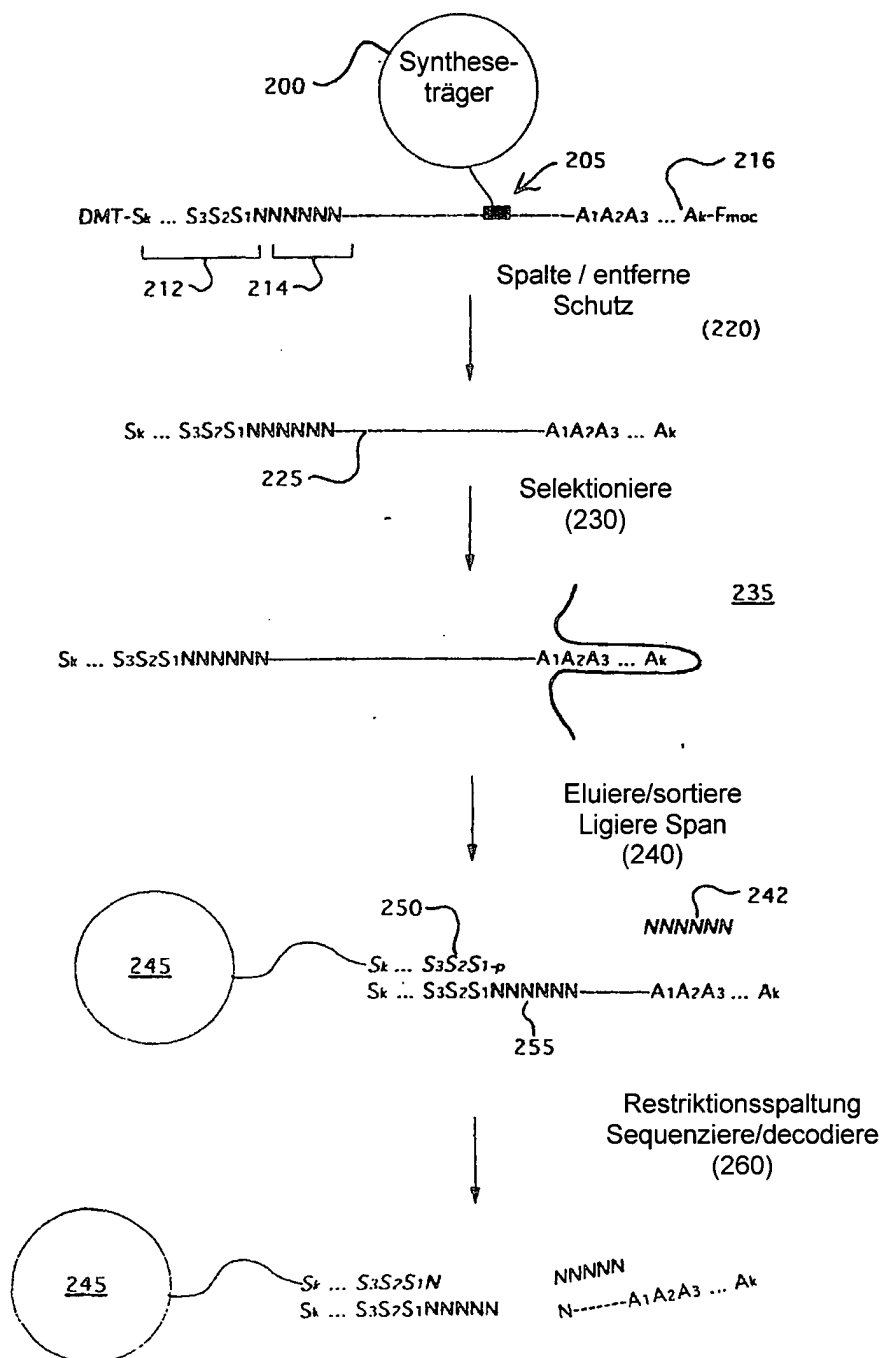


Fig. 4

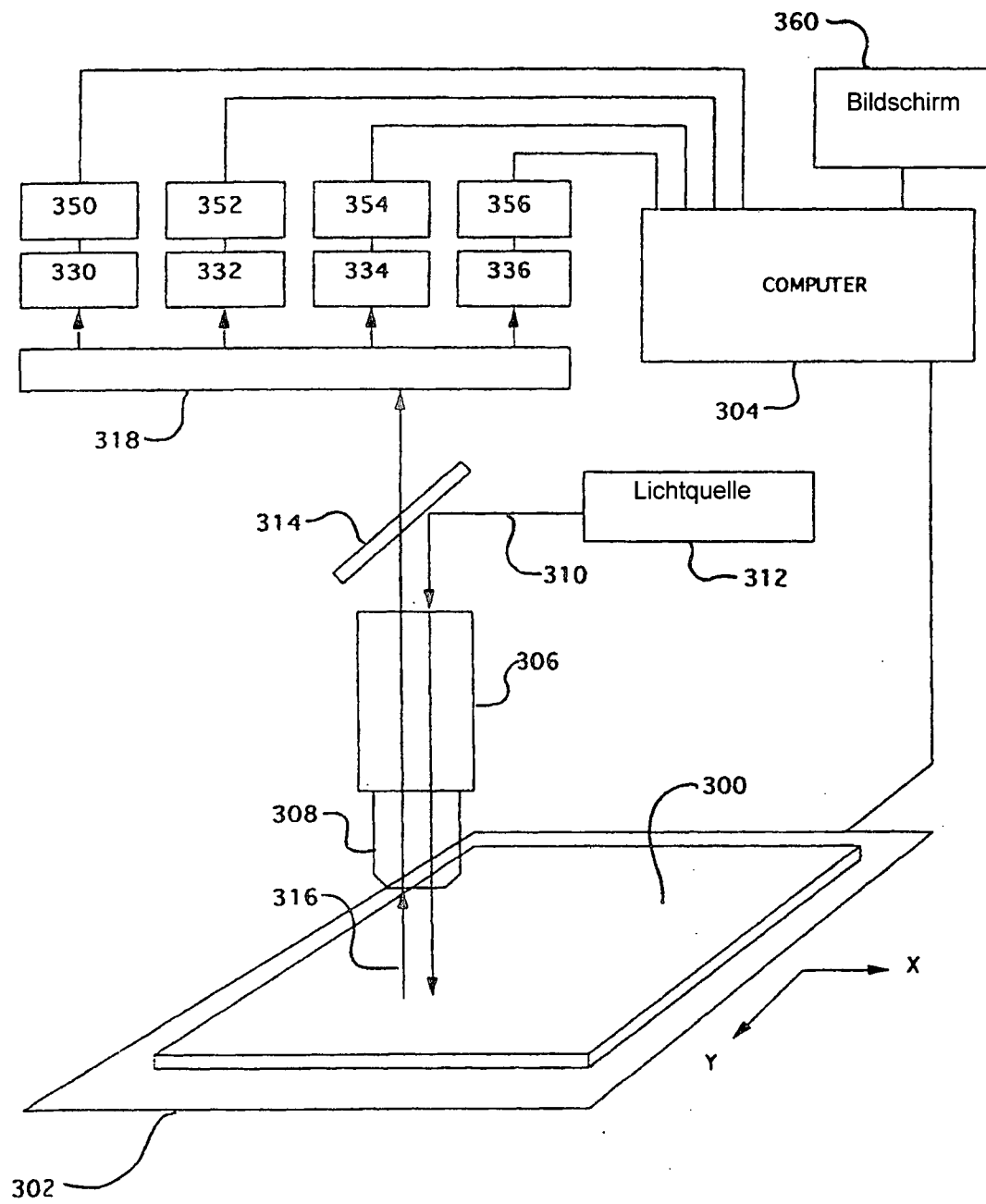


Fig. 5