



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 697 33 958 T2** 2006.05.18

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 935 674 B1**

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

(21) Deutsches Aktenzeichen: **697 33 958.0**

(86) PCT-Aktenzeichen: **PCT/FR97/01949**

(96) Europäisches Aktenzeichen: **97 912 288.4**

(87) PCT-Veröffentlichungs-Nr.: **WO 98/018959**

(86) PCT-Anmeldetag: **30.10.1997**

(87) Veröffentlichungstag
der PCT-Anmeldung: **07.05.1998**

(97) Erstveröffentlichung durch das EPA: **18.08.1999**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **10.08.2005**

(47) Veröffentlichungstag im Patentblatt: **18.05.2006**

(30) Unionspriorität:

9613276	30.10.1996	FR
9702999	13.03.1997	FR

(84) Benannte Vertragsstaaten:

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

(73) Patentinhaber:

**Institut Pasteur, Paris, FR; Centre National de la
Recherche Scientifique, Paris, FR**

(72) Erfinder:

**BENSIMON, Aaron, F-92160 Antony, FR;
BENSIMON, David, F-75013 Paris, FR; MICHALET,
Xavier, F-75018 Paris, FR**

(74) Vertreter:

Samson & Partner, Patentanwälte, 80538 München

(54) Bezeichnung: **VERFAHREN ZUR POSITIONIERUNG VON KLONEN MITTELS MOLEKULAREN KAEMMENS**

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

[0001] Die Erfindung betrifft ein Verfahren zur Positionsbestimmung von Polynukleotidsequenzen (welche Gene oder Teile von Genen enthalten können oder nicht) in einem Genom oder einem Genomabschnitt dank der sogenannten Technik des molekularen Kämmens.

[0002] Die Erfindung betrifft gleichfalls ein Verfahren zur Positionsbestimmung von Reagenzien biologischen, natürlichen oder synthetischen Ursprungs durch Kombination der Reagenzien mit der Gesamtheit oder einem Teil der gekämmten DNA.

[0003] Die Technik des molekularen Kämmens, wie sie in den nachfolgenden Referenzen beschrieben worden ist: WO 95/22056 und WO 95/21939 erlaubt, wenn sie auf Nukleinsäuren und insbesondere genomische DNA angewendet wird, die gleichförmige Ausdehnung bzw. Ausrichtung und Sichtbarmachung der DNA oder der RNA in Form von geradlinigen und praktisch schnurgerade ausgerichteten Filamenten.

[0004] Die Erfindung beruht auf dem Nachweis der Tatsache, dass man dank Sonden, d.h. Polynukleotiden, welche eine Abfolge von Nukleotidsequenzen enthalten, wie markierten DNA-Molekülen, die spezifisch Abschnitte der schnurgerade ausgerichteten DNA erkennen, die man mit der gekämmten DNA hybridisiert, direkt auf dem gekämmten Genom die Position der komplementären Sequenz sichtbar machen kann.

[0005] Unter diesen Bedingungen ist es möglich, beispielsweise unter Verwendung von zwei mit unterschiedlichen Chromophoren markierten Sonden, wie solchen, die mit einer beispielsweise roten und grünen Färbung sichtbar gemacht werden können, den Abstand, der diese trennt, zu messen. Es ist aber gleichfalls möglich, unter Verwendung von unterschiedlichen Sonden oder einer Reihe von aneinandergrenzenden Sonden (im Folgenden als „Contig“ bezeichnet) direkt die Länge der interessanten Zone zu messen und darin die gegebenenfalls vorhandenen Veränderungen im Falle eines abnormalen Genoms zu messen.

[0006] Die Erfindung betrifft ein Verfahren zur Positionsbestimmung von mehreren Klonen, wie in Anspruch 1 definiert. Verfahren zur Diagnose von genetisch bedingten Erkrankungen werden gleichfalls beschrieben, wobei die Verfahren (Erkrankungen) vorzugsweise durch bedeutende Veränderungen des Genoms entweder in seiner Struktur, beispielsweise Deletion, Translokation, oder in der Anzahl von Kopien von bestimmten Sequenzen (beispielsweise Trisomie, wobei die Sequenz die Gesamtheit eines Chromosoms darstellt) gekennzeichnet sind, wie auch Verfahren, welche erlauben, Gene schnell zu lokalisieren und zu kartieren.

[0007] Die Gendiagnostik kann in mehrere Bereiche unterteilt werden:

- pränatal,
- Pathologien mit genetischer Komponente,
- Krebs und Anfälligkeit für Krebs.

Pränataldiagnostik

[0008] Der Hauptteil (95%) der fötalen Anomalien ist auf Trisomien der Chromosomen 21, 18, 13, X oder Y zurückzuführen. Ihre unzweifelhafte Diagnose erfolgt mehr oder weniger spät (17. Woche Amenorrhoe, beispielsweise durch Amniozentese). Die Amniozentese erfordert eine bedeutende Punktion von Amnionflüssigkeit (einige zehn Milliliter), aus welcher in Suspension befindliche fötale Zellen extrahiert und mehrere Tage kultiviert werden (siehe die von S. Mercier und J.L. Bresson (1995) Ann. Génét., 38, 151-157, beschriebene Technik). Ein Karyotyp dieser Zellen wird durch mikroskopische Beobachtung und Zählung der Chromosomen durch ein hoch spezialisiertes Personal erstellt.

[0009] Eine Technik, die auf die Entnahme von Chorionzotten zurückgreift, erlaubt, auf den Schritt der Kultivierung zu verzichten, und vermeidet die Entnahme von Amnionflüssigkeit. Die Analyse des Karyotyps erfordert indessen den gleichen Arbeitsaufwand (siehe Médecine Périnatale, Biologie Clinique du Foetus, Andre Boué, Hrsg. Flammarion, 1989). Diese beiden Techniken können früher angewendet werden (bis zu 7 Wochen der Schwangerschaft für die Entnahme von Chorionzotten und 13-14 Wochen für die Amniozentese), aber mit einem leicht erhöhten Fehlgeburtsrisiko. Schließlich erlaubt eine direkte Entnahme von fötalem Blut im Bereich der Nabelschnur die Karyotypisierung ohne Kultur, setzt aber ein Team von auf diese Technik spezialisierten Klinikern voraus (Donner, C., et al., 1996, Fetal Diagn. Ther., 10, 192-199).

[0010] Andere Anomalien, wie Translokationen oder Deletionen/Insertionen von bedeutenden Teilen von Chromosomen können in diesem Stadium oder unter Einsatz von Techniken, wie der Fluoreszenz-in situ-Hy-

bridisierung (FISH), nachgewiesen werden. Indessen kann auch hier wieder diese Art der Diagnose lediglich durch ein hoch qualifiziertes Personal ausgeführt werden.

[0011] Untersuchungen zeigen außerdem, dass es noch keine immunologischen Methoden gibt, welche den Nachweis von fötalen Markern im mütterlichen Blut erlauben, welche eine unzweifelhafte Diagnose der Trisomie 21 oder von anderen Anomalien erlauben würden (siehe beispielsweise Wald, N.J., et al., 1996, Br. J. Obstet. Gynaecol., 103, 407-412, für das beispielsweise Wald, N.J., et al., 1996, Br. J. Obstet. Gynaecol., 103, 407-412, für das mit der Trisomie 21 verbundene Down-Syndrom).

[0012] Die aktuellen pränatalen Diagnosetechniken haben folglich zahlreiche Nachteile: sie können lediglich in einem relativ späten Stadium der Entwicklung des Embryos ausgeführt werden; sie sind nicht vollständig frei von Risiko für den Fötus oder für die Mutter; die Ergebnisse werden oftmals nach einer relativ langen Zeit (ungefähr 1 bis 3 Wochen je nach Technik) erhalten und sind kostspielig. Schließlich bleibt eine bestimmte Anzahl von chromosomalen Anomalien unbemerkt.

Diagnose von Pathologien mit genetischer Komponente

[0013] Zahlreiche Erkrankungen haben eine bekannte genetische Komponente (Diabetes, Bluthochdruck, Obesität u.s.w.), die das Ergebnis von chromosomalen Deletionen, Insertionen und/oder Umlagerungen von variablen Größen ist. Die Kultivierung von Zellen wirft in diesem Falle kein Problem auf, aber die FISH-Techniken, die von Lichter, G.D., et al. (1993), Genomics, 16, 320-324; Brandritt, B., et al. (1991), Genomics, 10, 75-82 und Van den Hengh, G., et al. (1992), Science, 257, 1410-1412) beschrieben worden sind, haben eine begrenzte Auflösung und bedürfen eines hoch qualifizierten Personals, was dieses Tests wenig erschwinglich macht.

[0014] Die Entwicklung eines leichter auszuführenden und preiswerten Tests würde das allgemeine Heranziehen von angepassten Therapien in einem frühen Stadium der beteiligten Pathologien, die deren Remission verbessern könnten, erlauben.

Krebsdiagnostik

[0015] Unter den Pathologien mit genetischer Komponente bilden die Krebserkrankungen eine wichtige Klasse, welche einen wachsenden Teil der Bevölkerung betrifft. Das aktuelle Verständnis des Prozesses des Auftretens einer Krebserkrankung lässt einen Proliferationsschritt von Protoonkogenen (Mutationen im Genom der Zellen), der der Transformation der Zelle zu einer Krebszelle vorangeht, auftreten. Dieser Proliferationsschritt ist bedauerlicherweise nicht nachweisbar, wohingegen die Möglichkeit, in diesem Stadium eine Behandlung vorzunehmen, sicherlich die Remissionschancen erhöhen und das Handicap der Patienten verringern würde.

[0016] Schließlich sind zahlreiche Tumore durch Chromosomenumlagerungen, wie Translokationen, Deletionen, partielle oder vollständige Trisomien u.s.w gekennzeichnet.

[0017] Auf jedem dieser Gebiete kann das molekulare Kämme ein bedeutenden Beitrag leisten entweder durch die Schnelligkeit und die geringe Menge von biologischem Material, die erforderlich ist, oder durch die quantitative Genauigkeit der Ergebnisse.

[0018] Der Vorteil der Technik erweist sich insbesondere in dem Falle, wo das genetische Material von Zellen stammt, die sich nicht mehr teilen oder die nicht in Kultur genommen werden können, ja sogar von toten Zellen, in welchen die DNA nicht signifikant abgebaut ist.

[0019] Im Falle der Pränataldiagnostik ist dies so nach Extraktion von fötalen Zellen, die im mütterlichen Blut zirkulieren (Cheung et al., 1996, Nature Genetics 14, 264-268). Dies gilt ebenso im Falle von Krebszellen, die von bestimmten Tumoren stammen.

[0020] Das molekulare Kämme erlaubt, die Möglichkeiten der Diagnose von genetisch bedingten Erkrankungen zu verbessern, es kann aber gleichfalls die Untersuchung und den Nachweis von genomischen Sequenzen, die für diese Erkrankungen verantwortlich sind, ermöglichen. Überdies beginnt gegenwärtig die Entwicklung eines Diagnose-„Kits“ oder -Koffers mit der Suche nach dem an der Pathologie beteiligten Gen.

[0021] Die Untersuchung von an (humanen oder andersartigen) Pathologien beteiligten Genen wird heutzutage im Allgemeinen in mehreren Schritten ausgeführt:

(i) Etablierung einer Zielpopulation von Individuen, die an den Pathologien leiden, von deren Nachkommen, Blutsverwandten in aufsteigender Linie und Seitenverwandten und Abnahme von Blut- und/oder Zellproben in Hinblick auf die Aufbewahrung von genetischem Material (in Form von DNA oder von Zellstämmen).

(ii) Genlokalisierung durch Cosegregations-Wahrscheinlichkeitsanalyse mit genetischen Markern (Linkage- oder Kopplungs-Analyse). In diesem Stadium der Untersuchung verfügt man über einige nahe gelegene Marker, die auf einem (oder mehreren) gegebenen Chromosomen lokalisiert sind, die erlauben, zu dem Schritt der physischen Lokalisierung überzugehen.

(iii) Physische Kartierung: ausgehend von den in dem vorangegangenen Schritt erhaltenen genetischen Markern wird ein Screening von Banken von humanen DNA-Klonen (YACs, BACs, Cosmiden oder anderen), die für die Region(en), die in dem vorangegangenen Schritt bestimmt wurde(n), spezifisch sind, ausgeführt. Man erhält so eine bestimmte Anzahl von Klonen, welche die vorangegangenen Marker enthalten. Der Bereich von Interesse kann dann genau kartiert werden, indem Klone von abnehmender Größe verwendet werden. Eine Klonierung des betreffenden Teils des Genoms kann gleichfalls erneut ausgehend von der humanen DNA ausgeführt werden.

(iv) Untersuchung des Gens: In diesem Stadium können mehrere Techniken eingesetzt werden: Exon-„Trapping“ (Verwendung von cDNA-Bibliotheken (komplementäre DNAs, erhalten ausgehend von mRNA), CpG-Inseln, Konservierung von interspezifischen Sequenzen u.s.w., die erlauben, eine kodierende Sequenz einem (oder mehreren) der in den vorangegangenen Schritten selektierten Klone zuzuweisen.

[0022] Die Gesamtheit dieser Strategie steht für einen bedeutenden Arbeitsaufwand (gegebenenfalls bis zu mehrere Jahre). Auch bildet jede Technik, die erlaubt, schneller bei dem Schritt (iv) anzukommen, einen Trumpf für die Untersuchung von Genen, aber gleichfalls bei der Diagnostik.

[0023] Im gegenwärtigen Stand der Technik wird, wenn das Gen lokalisiert worden ist, beispielsweise durch die vorangegangene Methode, dessen Nachweis im Allgemeinen ausgeführt dank spezifischer Sonden, die der fraglichen Sequenz entsprechen, wobei jene durch Methoden beispielsweise vom Typ PCR oder vom Typ LCR (wie in dem Patent EP 0 439 182 beschrieben) oder eine Technik vom Typ NASBA (von der Firma Organon Teknika vertriebener Kit) amplifiziert worden ist.

[0024] Gleichwohl sind die Amplifizierungstechniken nicht vollständig zufriedenstellend, insbesondere bei Heterozygoten, denn es existiert eine normale Kopie des Gens in dem Genom, ebenso wie in dem Falle einer großen Deletion oder von Erkrankungen mit wiederholten Sequenzen, wo die PCR nicht mehr zufriedenstellend funktioniert.

[0025] Die Diagnose einer großen Anzahl von genetisch bedingten Erkrankungen kann jetzt ins Auge gefasst werden unter Einsatz des molekularen Kämmens und der Markierung der DNA.

[0026] Desmaze und Aurias, Cellular and Molecular Biology, 1995, beschreiben eine FISH-Technik, die auf die Kartierung und die Charakterisierung von Genomumlagerungen angewendet wird durch eine Technik eines Streckens von Chromatin. Diese Technik besteht darin, eine DNA-Faser, welche mehrere Moleküle enthält, zu strecken, was deren Empfindlichkeit in bemerkenswerter Weise beeinträchtigt.

[0027] Das molekulare Kämmen ist eine Technik, die darin besteht, DNA-Moleküle unter wohl definierten physikalisch-chemischen Bedingungen durch ihre Enden auf Oberflächen zu verankern, gefolgt von deren Strecken mit Hilfe eines zurückweichenden Meniskus; man erhält so DNA-Moleküle, die schnurgerade parallel ausgerichtet sind. Die verwendete gereinigte DNA kann von jeglicher Größe sein und folglich insbesondere genomische DNA, die aus humanen Zellen extrahiert worden ist. Das Genom kann gleichfalls von einem genetischen Material stammen, welches wenigstens 80% genetisches Material fötalen Ursprungs umfasst.

[0028] Die so gekämmten DNA-Moleküle können denaturiert werden, bevor sie mit durch ein jegliches geeignetes Mittel (insbesondere mit Biotin-dUTP- oder Digoxigenin-dUTP-Nukleotiden) markierten Nukleinsäuresonden hybridisiert werden, die dann beispielsweise mit Hilfe von fluoreszierenden Antikörpersystemen sichtbar gemacht werden.

[0029] Da das molekulare Kämmen durch eine konstante Erstreckung der gekämmten Moleküle gekennzeichnet ist, liefert folglich das Messen der Längen der beobachteten fluoreszierenden Fragmente mit Hilfe (beispielsweise) eines Epifluoreszenz-Mikroskops direkt die Größe der Fragmente, an welchen Sonden anhybridisiert sind.

[0030] Das Erstreckungsausmaß hängt von dem Oberflächentyp ab, kann aber genau gemessen werden; es

beträgt beispielsweise 2 Kilobasen (kb) pro Mikrometer (μm) im Falle von silanisierten Oberflächen gemäß dem in der Referenz (1) beschriebenen und in den Beispielen ausgeführten Protokoll.

[0031] Wenn dies notwendig ist, ist es möglich, einen internen Standard vorzusehen, d.h. eine DNA von bekannter Länge, eine sogenannte Kalibrierungs-DNA, die erlauben wird, die Manipulation zu eichen, d.h. jede Messung zu kalibrieren.

[0032] Die Erfindung wird durch die Ansprüche 1-8 definiert und betrifft ein Verfahren, das dadurch gekennzeichnet ist, dass

- (a) man eine gewisse Menge des Genoms auf einer Oberfläche zum Kämmen fixiert und kämmt,
- (b) man das Produkt des Kämmens mit Sonden hybridisiert, die mit radioaktiven, fluoreszierenden oder anderen Elementen markiert sind, wie Kügelchen, Teilchen, welche jedem Klon entsprechen, derart, dass die Sonden spezifisch, insbesondere durch eine Farbe, nachgewiesen werden können,
- (c) man die Information, die dem Ort jedes Klons entspricht, sowie die entsprechenden Größen und Entfernungen auf dem Genom entnimmt,
- (d) man die Vorgehensweisen b) und c) n-mal wiederholt, indem man die Farbe, die Markierung oder die Nachweisart der Sonden modifiziert, wobei man weiß, dass man mit p verschiedenen Farben, Markierungen oder Nachweisarten am Ende von n Hybridisierungen die Position von p^n Klone bestimmen kann.

[0033] Im Rahmen der Kartierungs-Standardverfahren wächst die Anzahl I der Hybridisierungen, die erforderlich sind, um N Klone mit Hilfe von p Markierungen, Farben oder Nachweisarten zu kartieren, linear mit der Anzahl von Klone N.

[0034] So ist es bei 3 Farben heutzutage erforderlich, wenigstens 15 Hybridisierungen des vorangegangenen Typs auszuführen, um 30 Klone zu kartieren.

[0035] Diese Anzahl von Hybridisierungen ist bedeutend und die Anzahl von verfügbaren Farben ist in der Praxis begrenzt (sogar bei Verwendung von Kombinationen von Fluorophoren). Außerdem muss man, ist die Gesamtheit der Messungen einmal bewältigt, die Selektion der verschiedenen möglichen Positionen der Klone vornehmen, was nicht immer leicht sein kann, wenn man Messunsicherheiten berücksichtigt.

[0036] Das hier vorgeschlagene Verfahren erlaubt, eine exponentiell mit der Anzahl von ausgeführten Hybridisierungen ansteigende Anzahl von Klone zu kartieren.

[0037] Zur Veranschaulichung repräsentiert das Schema der [Fig. 10](#) das Ergebnis von zwei Hybridisierungen von 4 Klone, die mit 2 unterschiedlichen Farben von einer Hybridisierung zur anderen (für die Hälfte von diesen) nachgewiesen werden. Die 4 hybridisierten Klone bilden ein Schema, welches von einer Hybridisierung zur anderen unterschiedlich gefärbt ist, wobei die Klone über ihre Kodierung (in diesem Falle binär) markiert sind. In diesem Beispiel von 4 Klone reicht ein aus einer Aufeinanderfolge von 2 Farben gebildeter Code aus, um jeden Klon zu unterscheiden:

- A = Rot, dann Rot
- B = Grün, dann Grün
- C = Rot, dann Grün
- D = Grün, dann Rot.

[0038] Von 5 bis 8 Klone werden 3 Hybridisierungen erforderlich sein, um die Klone durch eine Aufeinanderfolge von 3 Farben zu unterscheiden.

[0039] Allgemeiner wird mit Hilfe von p Farben, um N Klone zu kartieren, eine Anzahl von Hybridisierungen I derart, dass : $N = p^I$, ausreichend sein.

[0040] Im Vergleich zu der Standardmethode können 30 Klone in 5 Hybridisierungen (anstelle von 30) kartiert werden, wenn man über lediglich 2 Farben verfügt, und in nur 4 Hybridisierungen (anstelle von 15), wenn man über 3 Farben verfügt.

[0041] Das hier präsentierte Kartierungsprinzip ist einfach. Indessen wird man, um bestimmte mögliche experimentelle Artefakte (Streuung der Größen der Signale, Variabilität der Sättigung, Bruch der Moleküle u.s.w.) zu überwinden, man vorteilhafterweise eine adäquate Software für die Bildbearbeitung und die statistische Analyse einsetzen.

[0042] Die nachfolgenden Beispiele werden erlauben, andere Merkmale und Vorteile der Erfindung besser zu verstehen.

[0043] Schließlich betrifft die Erfindung in einer vierten Ausführungsweise ein Verfahren, welches insbesondere die Detektion oder die Positionsbestimmung von Produkten, welche mit der gekämmten DNA reagieren können, erlaubt. Beispielsweise wird man die Transkription von Genen regulierende Proteine, die während des Zellzyklus an die DNA binden oder nicht, auf gekämmter DNA detektieren und deren präferentielle Bindungsstellen bezogen auf die Positionen von bekannten Sequenzen bestimmen und beispielsweise gemäß dem vorangegangenen Verfahren zur Ausführung der Erfindung ermitteln können.

[0044] Auf analoge Weise wird man Moleküle von therapeutischem Interesse, die mit der DNA reagieren können, auf gekämmter DNA detektieren können; ihre Wirkung auf andere Moleküle, die in der Lage sind, mit der DNA zu reagieren, könnte gleichfalls durch Vergleich untersucht werden.

[0045] Unter den Molekülen, die in der Lage sind, mit der gekämmten DNA zu reagieren, kann man die regulativen Proteine aufführen, wie sie beschrieben werden durch:

- Laughon und Matthew (1984), Nature, 310 : 25-30, für die regulativen Proteine, die sich an Drosophila-DNA anheften,
- Struhl, K., et al. (1987), Cell, 50 : 841-846, für regulative Proteine, die an die DNA in ihrer spezifischen Bindungsdomäne („binding-domain“) binden.
- Echols, H., et al. (1996), Science, 223 : 1050-1056, hinsichtlich der mehrfachen Wechselwirkungen der DNA, welche beispielsweise Transkriptionen induzieren;
- in dem Übersichtsartikel An. Rev. of Bioch. (1988), 57:159-167, beschreiben Gross und Ganard die Überempfindlichkeit der Nukleasestellen in Chromatin;
- Hanson et al. (1976), Science, 193 : 62-64, beschreiben Psoralen als photoaktives Mittel bei der selektiven Spaltung von Nukleotidsequenzen;
- Cartwright et. al. (1984), NAS, 10 : 5835-5852.

[0046] Im Rahmen der vorliegenden Beschreibung nimmt die Technik des Kämmens Bezug auf die Technologie, die in den zuvor erwähnten Dokumenten beschrieben worden ist, ebenso wie die Erwähnung einer „Oberfläche zum Kämmen“, die einer behandelten Oberfläche, welche die Verankerung der DNA und deren Strecken durch einen zurückweichenden Meniskus erlaubt, entspricht.

[0047] Es empfiehlt sich, anzumerken, dass die Oberfläche zum Kämmen vorzugsweise eine ebene Oberfläche ist, auf welcher die Ablesungen bequemer erfolgen.

[0048] Unter „Reaktion zwischen den markierten Sonden und der gekämmten DNA“ versteht man eine jegliche chemische oder biochemische Reaktion, insbesondere Reaktionen vom immunologischen Typ (beispielsweise gegen methylierte DNA gerichtete Antikörper), Protein/DNA- oder Nukleinsäure/DNA-Reaktionen (beispielsweise die Hybridisierung zwischen komplementären Abschnitten) oder Nukleinsäure/RNA-Reaktionen oder Nukleinsäure/RNA-DNA-Hybrid-Reaktionen. Man kann gleichfalls beispielsweise die chemischen DNA-DNA-Verknüpfungsreaktionen unter Verwendung von Psoralen-Molekülen oder die DNA-Polymerisationsreaktionen mit Hilfe eines Polymeraseenzymen aufführen.

[0049] Der Hybridisierung geht im Allgemeinen eine Denaturierung der fixierten und gekämmten DNA voran; diese Technik ist bekannt und wird hier nicht detailliert beschrieben.

[0050] Mit „Sonde“ sollen ebenso ein einzel- oder doppelsträngiges Polynukleotid, welches wenigstens 20 Nukleotide für dessen Aufbau umfasst, oder ein genomisches DNA-Fragment wie auch ein „Contig“, d.h. eine Gesamtheit von Sonden, die aneinander grenzen oder einander überlappen und die fragliche Zone überspannen, oder mehrere getrennte, markierte oder nicht-markierte Sonden bezeichnet werden. Unter „Sonde“ versteht man gleichfalls ein jegliches Molekül, das auf kovalente oder nicht kovalente Weise mit wenigstens einer der vorerwähnten Entitäten verknüpft ist, oder ein jegliches biologisches, natürliches oder synthetisches Molekül, welches mit der DNA reagieren kann, wobei der dem Begriff „Reaktion“ verliehene Sinn vorstehend präzisiert worden ist, oder ein jegliches Molekül, das auf kovalente oder nicht kovalente Weise mit einem jeglichen Molekül, das mit der DNA reagieren kann, verknüpft ist.

[0051] Im Rahmen der Erfindung sind die Sonden markiert. So könnten sie in dem Falle, wo die Sonden durch methylierte Cytosine markiert sind, sichtbar gemacht werden, nach Reaktion mit dem Produkt des Kämmens, durch fluoreszierende Antikörper, die gegen diese methylierten Cytosine gerichtet sind. Die die Markierung si-

herstellenden Elemente können radioaktiv sein, sind aber vorzugsweise kalte Markierungen, beispielsweise durch Fluoreszenz. Es kann sich gleichfalls um Nukleotidsonden handeln, bei denen bestimmte Atome ersetzt sind.

[0052] Die Größe der Sonden kann sich als ein jeglicher mit einer extensiven Einheit gemessener Wert verstehen, d.h., dass die Größe von zwei Sonden gleich der Summe der Größe der Sonden, einzeln genommen, ist. Ein Beispiel wird durch die Länge geliefert, aber es könnte beispielsweise auch eine Fluoreszenzintensität verwendet werden. Die Länge der eingesetzten Sonden liegt beispielsweise zwischen 5 kb und 40-50 kb, kann aber gleichfalls aus der Gesamtheit des gekämmten Genoms bestehen.

[0053] Die Reaktion der Sonde mit der gekämmten DNA wird vorzugsweise durch ein oder mehrere Moleküle, Lösemittel oder andere sachdienliche Parameter moduliert.

[0054] Schließlich muss man, wenn man im Folgenden von „Genom“ spricht, dies allgemein so auffassen, dass es sich um eine Vereinfachung handelt; eine jegliche Abfolge von DNA oder Nukleinsäuren, die in der Lage ist, auf einer Oberfläche zum Kämmen fixiert zu werden, wird von dieser Terminologie umfasst.

[0055] Außerdem wird der Begriff „Gen“ manchmal unterschiedslos verwendet, um einen „Genabschnitt“ genomischer Herkunft oder ebenso eine spezielle synthetische „Polynukleotidsequenz“ zu bezeichnen.

[0056] Ein in der Anmeldung beschriebenes Verfahren wird dazu eingesetzt, um die Diagnose von Brüchen in einem Genom zu ermöglichen, ebenso wie für die Positionsklonierung von solchen Brüchen. Es empfiehlt sich, anzumerken, dass der Begriff „Bruch“ eine große Anzahl von lokalen Modifikationen des Genoms, deren Liste weiter unten explizit erläutert werden wird, umfasst.

[0057] Das Verfahren besteht darin, die Position der etwaigen Bruchstellen, die an einer Pathologie von genetischer Ursache beteiligt sind, zu bestimmen durch Hybridisierung einer genomischen Sonde von bekannter Größe (kloniert oder andere), welche sich in der Region des untersuchten Gens befindet, mit gekämmter genomischer DNA von an der Pathologie leidenden Patienten. Diese Bruchstellen bestehen aus Stellen in der Gensequenz, deren Umgebung sich zwischen einem gesunden Individuum und einem kranken Individuum um mehrere Kilobasen (kb) unterscheidet.

[0058] Das Prinzip der Definition der Bruchstelle beruht auf der Möglichkeit, durch molekulares Kämmen eine lokale Modifikation des untersuchten Genoms bezogen auf ein bereits untersuchtes Genom im Bereich der betreffenden Regionen) zu detektieren.

[0059] Die Entwicklung von Verfahren zur Positionsbestimmung von lokalen Modifikationen des Genoms mit Größen unter 1 kb kann so mit Hilfe von Nahfeld-Beobachtungstechniken (AFM, STM, SNOM u.s.w.) oder solchen, welche eine intrinsisch höhere Auflösung aufweisen (beispielsweise Elektronenmikroskopie von Nanokügelchen aus Gold) ins Auge gefasst werden.

[0060] Insbesondere beschreibt die Anmeldung ein Verfahren zum Nachweis einer genetischen Anomalie in Form eines Bruchs in einem Genom, das dadurch gekennzeichnet ist, dass

- (a) man eine bestimmte Menge des Genoms auf einer Oberfläche zum Kämmen fixiert und kämmt,
- (b) man das aus dem Kämmen hervorgehende Produkt mit einer oder mehreren spezifischen markierten Sonden, welche der genomischen Sequenz, in welcher man die Anomalie sucht, entsprechen, hybridisiert,
- (c) man die Größe der den Hybridisierungssignalen entsprechenden Fragmente und gegebenenfalls ihre Wiederholung misst und
- (d) man daraus das Vorhandensein eines Bruches entweder durch direkte Messung oder durch Vergleich mit einem entsprechenden Standard mit einer Vergleichslänge ableitet.

[0061] Zur Veranschaulichung führt die Messung der Größe der Fragmente zu einem Histogramm, d.h. einer graphischen Darstellung der Längen der beobachteten Fragmente.

[0062] Um ein Histogramm von der Sonde zu erstellen, wertet man die Anzahl von Klonen aus, die eine bestimmte Sondenlänge aufweisen. Im Prinzip umfasst das Histogramm lediglich einen oder zwei Peaks je nach der Art des analysierten Bruchs, zwei Peaks, wenn die Sonde mit zwei getrennten Fragmenten hybridisiert, und einen Peak, wenn sie mit einem einzigen Fragment hybridisiert.

[0063] In dem Falle eines heterozygoten Genoms, bei dem eines der Allele hinsichtlich der betreffenden Re-

gion normal ist, überlagert die Signatur (das charakteristische Hybridisierungsmuster) des normalen Allels (das Fehlen eines Bruchs) jene des abnormalen Allels, sie kann aber aufgrund der Tatsache, dass sie bekannt ist, von dieser subtrahiert werden.

[0064] Dieses Verfahren kann gleichfalls verwendet werden, um eine Positionsklonierung auszuführen, d.h. um die Position von einem oder mehreren unbekannt Genen, die an einer Pathologie beteiligt sind, zu bestimmen. Das Prinzip besteht, wie zuvor, darin, Klone von humaner oder tierischer oder pflanzlicher DNA, die dann als Sonde dienen, mit der gekämmten genomischen DNA von einem oder mehreren an der untersuchten Pathologie leidenden Patienten zu hybridisieren. Das Sichtbarmachen dieser Hybridisierungen erlaubt, die Größe der hybridisierten Fragmente zu messen und ein Histogramm der beobachteten unterschiedlichen Größen zu konstruieren. Wenn der als Sonde verwendete Klon eine Bruchstelle des Gens überspannt, wird die Signatur dieses Phänomens an der (kleineren) Länge des hybridisierten Fragments ablesbar sein.

[0065] Die Verwendung einer begrenzten Anzahl von für die verantwortliche Region spezifischen Klonen, die man durch genetische Linkage- oder Kopplungsanalyse („linkage analysis“) wird abgeleitet haben können, wird so eine schnelle und genaue Bestimmung der Position einer Bruchstelle, einer Deletion oder einer jeglichen anderen genetischen Umlagerung von ausreichender Größe, um durch die mit dem molekularen Kämmen kombinierte Detektionstechnik aufgelöst werden zu können, erlauben.

[0066] In diesem Falle untersucht man selbstverständlich den Bruch, um diesen zu kartieren; in der Diagnostik ist der Bruch bekannt und es ist sein Vorhandensein oder sein Fehlen, die untersucht werden.

[0067] Es können zwei beispielhafte Fälle aufgeführt werden (unter der Hypothese, dass eine Bruchstelle in der an der Pathologie beteiligten Region des Genoms existiert):

- (i) die Sonde überlappt die Bruchstelle nicht,
- (ii) die Sonde überlappt die Bruchstelle.

[0068] In dem Falle (i) wird die Messung der Längen der fluoreszierenden Sonden vergleichbar sein mit jener, die mit der gleichen Sonde, die mit einer nicht-pathologischen genomischen DNA von gleicher Natur (d.h. im Wesentlichen von gleicher Größe und präpariert unter den gleichen Bedingungen) hybridisiert wird, erhalten wird.

[0069] In dem Falle (ii) unterscheidet sich im Gegensatz dazu, da die Sonde systematisch mit zwei (oder mehr) getrennten Stücken in der gekämmten genomischen DNA (durch Definition der Existenz einer Bruchstelle) hybridisiert wird, die Messung der Längen der hybridisierten fluoreszierenden Sonden von dem durch Hybridisierung mit einer nicht-pathologischen genomischen DNA erhaltenen Ergebnis. Überdies erlaubt die Größe der mit der pathologischen DNA hybridisierten Fragmente, die Position der Bruchstelle innerhalb des Klons mit einer Genauigkeit von einigen kb, ja sogar mehr, wenn eine stärker auflösende Technik eingesetzt wird, abzuschätzen.

[0070] Aus diesem Grund bleibt dann lediglich, das Gen in diesem Klon zu untersuchen. In der Praxis wird es sich darum handeln, diese Messungen für die Gesamtheit der Klone, die die dem Gen entsprechende Zone partiell überspannen können, zu wiederholen. Die Anzahl von Hybridisierungsplättchen könnte verringert werden, indem gleichzeitig mehrere unterschiedlich markierte Sonden hybridisiert werden oder unter Verwendung einer Kodierungsmethode durch Kombination von Farben, wie dies nachfolgend beschrieben wird.

[0071] Diese Technik erlaubt, die Position der gegebenenfalls vorhandenen Bruchstellen der an einer genetisch bedingten Pathologie beteiligten Region des Genoms durch Hybridisierung von klonierter genomischer DNA mit gekämmter genomischer DNA, welche von Patienten stammt, zu bestimmen. Diese Technik findet folglich Anwendung für die Untersuchung von Regionen des Genoms, die für Pathologien verantwortlich sind, welche zurückzuführen sind auf:

- die Deletion eines Teils oder der Gesamtheit dieser Region des Genoms,
- die Translokation der Gesamtheit oder eines Teils dieser Region des Genoms,
- die Verdoppelung oder das Vorhandensein von mehreren Kopien der Gesamtheit oder eines Teils dieser Region des Genoms innerhalb von dieser oder an einem jeglichen anderen Ort des Genoms,
- die Insertion einer jeglichen Gensequenz im Inneren dieser Region des Genoms.

[0072] In einem zweiten Fall und in bestimmten besonderen Fällen, insbesondere wenn die untersuchte genetische Anomalie bedeutende Deletionen oder Verdoppelungen (beispielsweise Fall der Trisomien) umfasst, kann ein in der vorliegenden Anmeldung beschriebenes Verfahren modifiziert werden, denn es handelt sich

dann darum, die Gene oder eine bestimmte Sequenz quantitativ zu bestimmen.

[0073] Insbesondere wird ein Verfahren zur quantitativen Bestimmung einer bestimmten genomischen Sequenz in einem Genom beschrieben, das dadurch gekennzeichnet ist, dass:

- (a) man eine gewisse Menge des Genoms auf einer Oberfläche zum Kämmen fixiert und kämmt,
- (b) man das Produkt des Kämmens mit einer markierten Vergleichssonde mit der Länge l_t , welche einer sogenannten genomischen Vergleichssequenz entspricht, d.h. von der man die Anzahl von Kopien in dem Genom kennt, und mit einer spezifischen markierten Sonde mit der Länge l_c , welche der quantitativ zu bestimmenden genomischen Sequenz entspricht, derart hybridisiert, dass die Sonden getrennt identifiziert werden können,
- (c) man dann die gesamte Länge der Hybridisierungssignale für die beiden Sonden, entsprechend L_c und L_t , misst,
- (d) man für jede die Anzahl von Kopien der entsprechenden Sequenz berechnet durch das Verhältnis

$$N_t = \frac{L_t}{l_t} \quad \text{und} \quad N_c = \frac{L_c}{l_c}$$

und man daraus die Anzahl von Kopien der quantitativ zu bestimmenden Sequenz bezogen auf die Vergleichssequenz ableitet.

[0074] In dem Falle der Pränataldiagnostik der Trisomie 21 könnte das Verfahren in der Hybridisierung einer für ein Vergleichschromosom spezifischen Cosmid-Sonde (beispielsweise Chromosom 1 – Sonde mit der Länge l_t), welche mit biotinylierten Nukleotiden markiert ist, und der Hybridisierung einer für das Chromosom 21 spezifischen Cosmid-Sonde (Sonde mit der Länge l_c), welche mit Digoxigenin markiert ist, mit aus Amnionproben oder einer jeglichen anderen Probe, welche Zellen fötalen Ursprungs enthält, extrahierter gekämmter genomischer DNA bestehen.

[0075] Man könnte beispielsweise ein Avidin-Texas Red (rote Farbe)-Nachweissystem für die Vergleichssonde und ein Antidigoxigenin-FITC-Nachweissystem (grüne Farbe) für die spezifische Sonde einsetzen: die gesamte Länge der in einer gegebenen Zone der Oberfläche beobachteten roten Hybridisierungssignale, L_t , und die gesamte Länge der in derselben Zone oder in einer äquivalenten Zone der Oberfläche beobachteten grünen Hybridisierungssignale, L_c , führen folglich zu den oben definierten Werten N_t und N_c .

[0076] Ein N_c/N_t -Verhältnis nahe 1 wird einen normalen Genotyp (2 Chromosomen 21 pro 2 Chromosomen 1) anzeigen, wohingegen ein Verhältnis nahe 1,5 einen Trisomie-Genotyp (3 Chromosomen 21 pro 2 Chromosomen 1) anzeigen wird.

[0077] Allgemein ist ein signifikanter Unterschied zwischen N_c und dem für die Anzahl von vorhandenen Genomen erwarteten Wert, der aus N_t abgeleitet wird, ein Anzeichen für das Vorhandensein einer Genanomalie.

[0078] In dem Falle der Diagnose von Onkogenen oder Protoonkogenen könnte die gleiche Methode eingesetzt werden: eine Vergleichssonde wird hybridisiert und beispielsweise in Rot nachgewiesen und eine dem Gen oder einem Teil des untersuchten Gens entsprechende Sonde wird hybridisiert und beispielsweise in Grün nachgewiesen. Nachdem die Messungen ausgeführt worden sind, wie oben, wird das Verhältnis N_c/N_t das relative Vorkommen des Gens bezogen auf die Häufigkeit von zwei Kopien pro diploidem Genom liefern.

[0079] Die aberrante Methylierung der GpC-Inseln, die bei zahlreichen Krebsarten häufig beobachtet wird (92% der Krebserkrankungen des Kolons), kann durch das erfindungsgemäße Verfahren ebenfalls nachgewiesen werden durch Reaktion zwischen der gekämmten DNA und fluoreszierenden Antikörpern, die gegen die methylierten Cytosine gerichtet sind.

[0080] Tatsächlich ist der Verlust der Heterozygotie bezüglich des Chromosoms 9p21 eine der Genveränderungen, die bei Krebserkrankungen beim Menschen am häufigsten identifiziert werden. Das Tumorsuppressorgen CDKN2/p16/MTS1, welches in dieser Region lokalisiert ist, ist bei vielen Krebserkrankungen beim Menschen durch homozygote Deletion häufig inaktiviert. Gleichwohl wurde über eine andere Inaktivierungsweise, an welcher der Verlust der Transkription, welcher mit einer de novo-Methylierung der 5'-GpC-Inseln von CDKN2/p16 verbunden ist, beteiligt ist, bei Krebserkrankungen der Lungen, Gliomen und den Karzinomen mit Desquamation des Kopfes und des Schwanzes berichtet. Diese aberranten Methylierungen der GpC-Inseln treten gleichfalls häufig in Zelllinien von Krebsarten der Brust (33%), der Prostata (60%), der Niere (23%) und

des Kolons (92%) auf (Herman, J.G., et al. (1995) Cancer Res., 15. Okt., 55(20) : 4525-30; Wales, M.M., et al. (1995), Nature Med., Jun., 1(6) : 570-607).

[0081] Die genaue Lokalisierung der Methylierungsbereiche auf einem Gen ist von sehr großer Bedeutung für das Verständnis des Mechanismus der Entwicklung der Krebserkrankung und für einen möglichen „Screening“-Test. Das molekulare Kämmen kann mit einer Genauigkeit von einigen kb die Lokalisierung von solchen GpC-Inseln, die an der Entwicklung der Krebserkrankung beteiligt sind, nachweisen.

[0082] Diese Technik, die erlaubt, die Anzahl von Kopien eines Gens in einem Genom zu bestimmen, kann gegebenenfalls eingesetzt werden, um die Abwesenheit eines Teils des Genoms nachzuweisen.

[0083] In dem Falle einer Pathologie, die durch die Deletion eines bedeutenden Teils eines Chromosoms gekennzeichnet ist, reicht es tatsächlich aus, als Zielsequenz einen Klon, der in der deletierten Zone enthalten ist, und als Vergleichssequenz einen Klon außerhalb dieser Zone zu nehmen. Es ist so möglich, Deletionen mit einer Größe eines Cosmid-Klons (30-50 kb) oder mehr nachzuweisen.

[0084] Wenn man über eine ausreichende Dichte von gekämmten Molekülen verfügt, kann in Betracht gezogen werden, kleinere Deletionen (einige kb), welche einem Teil der eingesetzten Zielsequenz entsprechen, zu detektieren. Dies ist der Grund, warum es besonders interessant ist, auf der gekämmten Oberfläche wenigstens etwa zehn Genomkopien anzuordnen.

[0085] Die statistische Unsicherheit bezüglich des Nc/Nt-Verhältnisses liegt in der Größenordnung $1/\sqrt{Nc} + 1/\sqrt{Nt}$. Es empfiehlt sich vorteilhafterweise, auf der gekämmten Oberfläche über eine ausreichende Anzahl von Signalen zu verfügen, um eine statistische Unsicherheit von weniger als 20% bezüglich des Nc/Nt-Verhältnisses zu haben. Es ist folglich wichtig, über eine große Anzahl von hybridisierten Sonden zu verfügen, typischerweise Nc, Nt > 100.

[0086] Indessen ist es in der Praxis auch möglich, die Genauigkeit dieser Messungen zu erhöhen, indem nicht mehr eine, sondern mehrere Arten von Vergleichssonden und von Zielsonden verwendet werden, ohne notwendigerweise anzustreben, zwischen diesen Arten von Sonden zu unterscheiden, d.h. indem sie alle auf die gleiche Weise nachgewiesen werden.

[0087] Die Möglichkeit, eine solche Anzahl von Signalen zu erhalten, ist gezeigt worden: es ist möglich, etwa hundert Signale auf einer silanisierter Glasoberfläche von 20 × 20 mm Nutzoberfläche zu zählen. Diese Dichte könnte beträchtlich erhöht werden, sobald man über eine bedeutende DNA-Menge verfügt.

[0088] Es erweist sich, dass eine ausreichende Anzahl von Genomen pro Oberflächen von 20 × 20 mm Nutzfläche sich im Bereich von 100 befindet, wenn eine einzige Sonde eingesetzt wird. In dem Falle der Verwendung von mehreren Sonden oder von größeren Sonden kann in Betracht gezogen werden, dass man verringern kann entweder:

- die Oberfläche zum Kämmen und folglich die analysierte Oberfläche
- die eingesetzte DNA-Dichte, folglich die Anzahl von gekämmten Genomen an der Oberfläche.

[0089] Je nach der hauptsächlichen Einschränkung (verlangte Schnelligkeit oder DNA in begrenzter Menge) könnte der eine oder der andere dieser beiden Wege eingesetzt werden.

[0090] Die dargelegte Technik beinhaltet die Verwendung von strikten Präparationsprotokollen, aber ohne besondere technische Schwierigkeiten. Auf der Ebene der Analyse der Signale ist keinerlei besondere Qualifikation erforderlich, was die Technik so für alle Laboratorien, die ein Personal mit minimalen Kompetenzen im Bereich der Molekularbiologie aufweisen, generell einsetzbar macht.

[0091] Einige hunderttausend Zellen müssen im Prinzip ausreichen, um eine Lösung von genomischer DNA herzustellen, die zu einer bedeutenden Dichte von gekämmten Molekülen auf den Analyseoberflächen führt. Es ist folglich im Prinzip notwendig, in den meisten Fällen Zellkulturen auszuführen. Die Gesamtheit aus Probennahme und Analyse muss folglich in einigen Tagen ausgeführt werden können.

[0092] Die Einfachheit der zu analysierenden Signale (parallel und von dem Hybridisierungshintergrund abgehoben) erlaubt es, eine vollständige Automatisierung des Prozesses der Analyse der Signale (Scan der Oberflächen, Erfassung und Auswertung der Messungen) ins Auge zu fassen. Die Ergänzung mit einem System zur Aufbewahrung von Oberflächen, welche verschiedenen Patienten entsprechen, erlaubt es, bedeuten-

de Ausbeuten in Betracht zu ziehen, was die Möglichkeit liefert, verschiedene Arten von Diagnosen in einigen Tagen zu liefern.

[0093] Das zuvor beschriebene Verfahren kann verschiedene Arten von Diagnosen erlauben: Zählung von Chromosomen (Trisomie, Monosomie u.s.w.), Zählung der Exemplare eines Gens, Nachweis von bekannten Deletionen oder andersartigen chromosomalen Modifikationen, die in einer Modifikation der mit einer gegebenen genomischen Sonde pro Genom hybridisierten Länge zum Ausdruck kommen.

[0094] Es ist festzuhalten, dass es auch möglich ist, eine partielle Deletion auf einem einzelnen Allel nachzuweisen.

[0095] Es ist ebenso möglich, die Hybridisierung von Klonen auf mehreren unterschiedlichen gekämmten Genomen auf ein und derselben Oberfläche auszuführen. Beispielsweise erlaubt das gleichzeitige Kämmen des Genoms eines Hauptorganismus und des Genoms von Wirtsorganismen (Parasiten, Bakterien, Viren..) und die Verwendung von einerseits für den Hauptorganismus und andererseits für die Wirtsorganismen spezifischen Sonden im Prinzip, das Verhältnis der Anzahl von Wirten/Anzahl von Zellen des Hauptorganismus zu bestimmen. In dem Falle eines durch ein Virus infizierten Organismus wird dies die Messung der Virusbelastung erlauben. Die weiter oben aufgeführten Zahlen begrenzen wahrscheinlich die Empfindlichkeit dieser Diagnosemethode in Situationen, wo man mehr als einen infektiösen Organismus pro ungefähr 100 Wirtszellen findet.

[0096] Diese unterschiedlichen Diagnosearten können kombiniert werden dank der Verwendung einer Mehrzahl von Nachweissystemen (mehrere Farben oder Kombinationen von Farben) oder einer jeglichen anderen Methode, welche die Unterscheidung zwischen den Hybridisierungssignalen, welche von unterschiedlichen Sonden stammen und für eine genaue Diagnose bestimmt sind, erlaubt.

[0097] Da das Prinzip der Technik auf dem Kämmen der DNA des Patienten beruht, erfordert dieser Schritt der Vorbereitung der DNA Protokolle für die Extraktion, das Kämmen und das entsprechende Material (behandelte Oberflächen, Apparatur für das molekulare Kämmen).

[0098] Die Diagnose selbst erfordert die Hybridisierung von spezifischen Nukleotidsonden und das Sichtbarmachen dieser Sonden, beispielsweise durch Antikörper-Systeme. Da außerdem im Falle von kombinierten Diagnostiken eine Kodierung durch Farbe ausgeführt werden kann, ist es folglich möglich, Gruppen von vor-markierten Sonden, welche einem Katalog von bestimmten Diagnosen entsprechen, vorzuschlagen.

[0099] Es wird gleichfalls ein Verfahren beschrieben, welches insbesondere die physische Kartierung eines Genoms erlaubt.

[0100] Da das Ziel der physischen Kartierung die Zuweisung eines Klons zu einem bestimmten Ort („ordonancement“) innerhalb eines Genoms ist, wird das molekulare Kämmen natürlicherweise zu diesem Zweck angewendet durch einfache Hybridisierung der Klone mit dem gekämmten Genom (beispielsweise kann im Falle eines YAC das gesamte Genom der Hefe gekämmt werden, um die Abtrennung des künstlichen Chromosoms von den natürlichen Chromosomen der Hefe wirtschaftlich auszuführen).

[0101] Die Position der Klone wird durch direkte Messung von deren Abstand zu einem Referenzklon oder zu einem jeglichen anderen Hybridisierungsreferenzsignal auf dem gekämmten Genom erhalten. Die konstante Erstreckung der gekämmten DNA erlaubt dann, direkt in Kilobasen (kb) die jeweilige Position der Klone wie auch deren Größe, wenn jene die Auflösung der Methode übersteigt, zu ermitteln. Insbesondere in dem Falle der klassischen Mikroskopie durch Epifluoreszenz, deren Auflösung eine halbe Wellenlänge ist, ist die genaue Kartierung von cDNA (zu den in der Zelle transkribierten RNAs komplementäre DNA) möglich, aber ohne die Möglichkeit zur genauen Messung der Größe der hybridisierten Fragmente (Exons), die in der Größenordnung von im Allgemeinen einigen hundert Basen liegt. Indessen kann durch diese Methode die genaue Lokalisierung von vollständigen cDNA-Fragmenten oder von deren Fragmenten innerhalb der genomischen DNA erhalten werden. Man kann beispielsweise das Vorhandensein oder das Fehlen und die Position der einem Protein von Interesse entsprechenden cDNAs bestimmen durch Hybridisierung der cDNAs mit der genomischen DNA oder mit einem Klon von genomischer DNA (Cosmid, BAC, YAC z.B.) zu der gleichen Zeit, wie ein Klon, der als Markierung dient.

[0102] Die Verwendung einer Mehrzahl von Fragmenten, welche ausgehend von einer cDNA erhalten worden sind, führt zu der Bestimmung des Vorhandenseins oder nicht von einer oder mehreren genomischen DNAs in

einem Vektor beispielsweise vom Cosmid- oder YAC-Typ.

[0103] Der Einsatz von höher auflösenden Methoden kann eine ergänzende Messung der Größe der Sonden (Nahfeld, Elektronenmikroskopie u.s.w.) erlauben, aber eine Messung der Fluoreszenzintensität, wenn es sich um die gewählte Beobachtungsweise handelt, kann diese Information gleichfalls liefern.

[0104] Die Methode, die wir beschreiben, erlaubt, die Anzahl von für die Ortsbestimmung einer gegebenen Anzahl von Klonen erforderlichen Hybridisierungen zu minimieren, wenn eine festgelegte Anzahl von zum Nachweis der Hybridisierungen dienenden Farben oder (allgemeiner) von unterschiedlichen Weisen für den Nachweis der Hybridisierungen gegeben ist.

Legenden der Figuren:

[0105] Fig. 1a: Veranschaulichung des Falles eines Klons (rot), welcher das untersuchte Gen (und folglich die Bruchstelle) überlappt. Die schraffierten Zonen muss man sich als fehlend vorstellen, was insbesondere impliziert, dass, wenn zwei Abschnitte von Sonden sich auf jeweils einer der beiden Seiten befinden, sie lediglich ein Segment mit einer Länge gleich der Summe von deren Längen bilden. Das (die) zugrundeliegende(n) Chromosomen) sind durch weiße Balken dargestellt.

(a) Situation bei einem gesunden Patienten.

(b1) Deletion(en) eines Teils des Gens. (b2) Deletion der Gesamtheit des Gens.

(c1) und (c2) Translokation eines Teils des Gens. (c3) Translokation der Gesamtheit des Gens.

(d1) Verdoppelung(en) eines Teils des Gens (gegebenenfalls mit Inversion). (d2) Verdoppelungen der Gesamtheit des Gens (gegebenenfalls mit Inversion). (d3) Wiederholung(en) eines Teils des Gens in einem anderen Abschnitt des Genoms. (e) Insertionen) eines unterschiedlichen Teils des Gens (und des klonierten Genoms) in das Gen.

[0106] Fig. 1b: Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Situationen entsprechen. In allen Fällen (ein oder zwei Peaks) ist die Situation klar unterscheidbar von jener des vollständigen Klons (a).

[0107] Die Histogramme dieser Figur entsprechen den Signalen des abnormalen Allels und werden folglich nach Subtraktion des auf das normale Allel zurückzuführenden Beitrags, dargestellt in (a), von den rohen Histogrammen, dargestellt in **Fig. 1c**, erhalten. Auf der Abszisse ist die Größe der hybridisierten Fragmente aufgetragen, wobei der (willkürliche) Maßstab mit jenem der **Fig. 1a** übereinstimmt. Auf der Ordinate ist die Anzahl von beobachteten Fragmenten in willkürlicher Einheit aufgetragen, wobei allein die Verhältnisse zwischen den verschiedenen Populationen von Bedeutung sind.

[0108] Fig. 1c: Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Situationen entsprechen und das normale Allel berücksichtigen. Der Beitrag des normalen Allels wird durch den Peak, welcher der Position des normalen Falls, dargestellt in (a), entspricht, dargestellt. Wenn das abnormale Allel dazu beiträgt, den Wert dieses Peaks zu erhöhen (Fall (d3)), kann der Beitrag des normalen Allels im Allgemeinen als äquivalent zu dem Wert eines abnormalen Peaks abgeschätzt werden.

[0109] Fig. 2a: Veranschaulichung des Falles eines Klons, welcher das untersuchte Gen teilweise überlappt.

(a) Situation bei einem gesunden Patienten

(b1) Deletion(en) eines Teils des Gens. Eine andere Situation wurde nicht dargestellt, wo der deletierte Abschnitt vollständig in dem Klon enthalten ist (siehe Fall, welcher dem Schema 1 a) entspricht).

(b2) Deletion der Gesamtheit des Gens

(c1) und (c2) Translokation(en) eines Teils des Gens. Eine andere Situation, welche dem Fall (c2) des Schemas 1a entspricht, wurde nicht dargestellt, wo der translozierte Abschnitt vollständig in dem Klon enthalten ist.

(c3) Translokation(en) der Gesamtheit des Gens.

(d1) Verdoppelung(en) eines Teils des Gens, gekennzeichnet durch einen doppelten Pfeil (der verdoppelte Abschnitt ist vollständig in dem Klon enthalten). (d2) Verdoppelung(en) eines Teils des Gens mit Inversion.

(d3) Verdoppelung(en) eines Teils des Gens (der verdoppelte Abschnitt ist teilweise in dem Klon enthalten).

(d4) Verdoppelung(en) der Gesamtheit des Gens (ohne Inversion).

(d5) Wiederholung(en) eines Teils des Gens in einem anderen Abschnitt des Genoms.

(e) Insertionen) eines unterschiedlichen Teils des Gens (oder des klonierten Genoms) in das Gen.

[0110] Fig. 2b: Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Si-

tuationen entsprechen. In allen Fällen (ein oder zwei Peaks) ist die Situation klar unterscheidbar von jener des vollständigen Klons (a).

[0111] Die Histogramme dieser Figur entsprechen den Signalen des abnormalen Allels und werden folglich nach Subtraktion des auf das normale Allel zurückzuführenden Beitrags, dargestellt in (a), von den rohen Histogrammen, dargestellt in [Fig. 2c](#), erhalten. Auf der Abszisse ist die Größe der hybridisierten Fragmente aufgetragen, wobei der (willkürliche) Maßstab mit jenem der [Fig. 2a](#) übereinstimmt. Auf der Ordinate ist die Anzahl von beobachteten Fragmenten in willkürlicher Einheit aufgetragen, wobei allein die Verhältnisse zwischen den verschiedenen Populationen von Bedeutung sind.

[0112] [Fig. 2c](#): Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Situationen entsprechen und das normale Allel berücksichtigen. Der Beitrag des normalen Allels wird durch den Peak, welcher der Position des normalen Falls, dargestellt in (a), entspricht, dargestellt. Wenn das abnormale Allel dazu beiträgt, den Wert dieses Peaks zu erhöhen (Fall (d4) und (d5)), kann der Beitrag des normalen Allels im Allgemeinen als äquivalent zu dem Wert eines abnormalen Peaks abgeschätzt werden.

[0113] [Fig. 3a](#): Veranschaulichung des Falles eines Klons, welcher in dem untersuchten Gen vollständig enthalten ist.

(a) Situation bei einem gesunden Patienten

(b1) Deletion(en) eines Teils des Gens. Eine andere Situation wurde nicht dargestellt, wo der deletierte Abschnitt vollständig in dem Klon enthalten ist (siehe Fall, welcher dem Schema 1 a) entspricht).

(b2) Deletion der Gesamtheit des Gens

(c1) und (c2) Translokation(en) eines Teils des Gens. Eine andere Situation, welche dem Fall (c2) des Schemas 1a entspricht, wurde nicht dargestellt, wo der translozierte Abschnitt vollständig in dem Klon enthalten ist.

(c3) Translokation(en) der Gesamtheit des Gens.

(d1) Verdoppelung(en) eines Teils des Gens, gekennzeichnet durch einen doppelten Pfeil (der verdoppelte Abschnitt ist vollständig in dem Klon enthalten). (d2) Verdoppelung(en) eines Teils des Gens mit Inversion.

(d3) Verdoppelung(en) eines Teils des Gens (der verdoppelte Abschnitt ist teilweise in dem Klon enthalten).

(d4) Verdoppelung(en) der Gesamtheit des Gens (ohne Inversion).

(d5) Wiederholung(en) eines Teils des Gens in einem anderen Abschnitt des Genoms.

(e) Insertionen) eines unterschiedlichen Teils des Gens (und des klonierten Genoms) in das Gen.

[0114] [Fig. 3b](#): Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Situationen entsprechen. In allen Fällen (ein oder zwei Peaks) ist die Situation klar unterscheidbar von jener des vollständigen Klons (a).

[0115] Die Histogramme dieser Figur entsprechen den Signalen des abnormalen Allels und werden folglich nach Subtraktion des auf das normale Allel zurückzuführenden Beitrags, dargestellt in (a), von den rohen Histogrammen, dargestellt in [Fig. 3c](#), erhalten. Auf der Abszisse ist die Größe der hybridisierten Fragmente aufgetragen, wobei der (willkürliche) Maßstab mit jenem der [Fig. 3a](#) übereinstimmt. Auf der Ordinate ist die Anzahl von beobachteten Fragmenten in willkürlicher Einheit aufgetragen, wobei allein die Verhältnisse zwischen den verschiedenen Populationen von Bedeutung sind.

[0116] [Fig. 3c](#): Ideale Histogramme der Längen von hybridisierten Klonen, welche den vorangegangenen Situationen entsprechen und das normale Allel berücksichtigen. Der Beitrag des normalen Allels wird durch den Peak, welcher der Position des normalen Falls, dargestellt in (a), entspricht, dargestellt. Wenn das abnormale Allel dazu beiträgt, den Wert dieses Peaks zu erhöhen (Fall (c3), (d2), (d3) und (d5)), kann der Beitrag des normalen Allels im Allgemeinen als äquivalent zu dem Wert eines abnormalen Peaks abgeschätzt werden.

[0117] [Fig. 4](#): Histogramme der Längen des Stufenmodells. Untersuchung des Einflusses der charakteristischen Größe.

[0118] [Fig. 5](#): Histogramme der Längen des Stufenmodells. Untersuchung des Einflusses des Bruchanteils.

[0119] [Fig. 6](#): Histogramme der Längen des Gauss-Modells. Untersuchung des Einflusses der charakteristischen Größe.

[0120] [Fig. 7](#): Histogramme der Längen des Gauss-Modells. Untersuchung des Einflusses des Bruchanteils.

[0121] Fig. 8: Simulation von Histogrammen in dem Fall der Hybridisierung eines BAC (125 kb) mit einer Bruchstelle:

(a) Fragmente von 35 und 90 kb; (b) Fragmente von 50 und 75 kb; (c) Fragmente von 60 und 65 kb; (d) Vergleichssituation; (e) reales Histogramm eines Contigs von Cosmid-Klonen, die mit humaner genomischer DNA hybridisiert worden sind (Länge des Contigs: ungefähr 77,5 µm, was 155 kb entspricht). Die Achse der Abszissen repräsentiert die Größe der Fragmente in Mikrometern. Die Achse der Ordinaten repräsentiert die Anzahl von kontinuierlichen Fragmenten von hybridisierten Sonden.

[0122] Fig. 9: Veranschaulichung des Prinzips der Kartierung ohne Farbkodierung. Jede Linie repräsentiert die mögliche Positionierung des neuen kartierten Subklons bezogen auf jenen, mit welchem er hybridisiert ist. Die Orientierung der beiden ersten Klone ist willkürlich. Bei jedem Schritt sind zwei Positionen auf beiden Seiten des älteren eingesetzten Klons für den neuen hybridisierten Klon möglich.

[0123] Fig. 10: Veranschaulichung des Prinzips der Kartierung mit Farbkodierung.

[0124] Bei jeder Hybridisierung werden die gleichen Subklone hybridisiert, was zu ein und demselben Hybridisierungsmotiv (-schema) führt, wenn man eine Abstraktion der Farbinformation vornimmt. Die beiden eingesetzten Farbpaletten erlauben es, ohne Zweideutigkeit jeden der Subklone zu markieren (Prinzip der Kodierung).

[0125] Fig. 11: Genomische Gesamt-DNA von Hefe, welche ein künstliches Chromosom enthielt (YAC774G4) wurde auf behandelten Oberflächen gekämmt. Cosmide, die zu zwei benachbarten Contigs, die durch ein unbekanntes Intervall voneinander getrennt sind, gehören, wurden jeweils zu zweit derart hybridisiert, dass man eine genaue Karte von deren Anordnung rekonstruieren konnte. Unter den eingesetzten Cosmiden enthält 1F11 den Hauptteil der Sequenz des Gens von Calpain 3, dessen Mutationen für die Gürtelmypopathie vom Typ 2A verantwortlich sind.

[0126] Die Figur zeigt charakteristische Bilder von jeder Hybridisierung (wobei links ein Cosmid hybridisiert und in rot sichtbar gemacht wurde, und das andere in grün rechts) wie auch die ausgehend von den Messungen mit ungefähr 3 kb erstellte Karte. Der Maßstab wird durch den Balken von 20 kb angegeben. Dieses Experiment hat insbesondere erlaubt, eine mit Hilfe von STSs erhaltene Karte der Region (Richard et al., 1995, Mammalian Genome, 6, 754-756) zu korrigieren und die chromosomale Orientierung des Gens zu ermitteln.

[0127] Fig. 12: Humane genomische Gesamt-DNA wurde auf behandelten Oberflächen gekämmt. Cosmide, die zu 3 benachbarten Contigs aus der Region des Chromosoms 9, welche das (nicht klonierte) „Tuberous Sclerosis 1“-Gen (TSC1) enthält, gehören, wurden jeweils zu zweit hybridisiert und in 2 unterschiedlichen Farben sichtbar gemacht, um die Größe der Intervalle, welche diese Contigs trennen, zu messen. Die Figur zeigt 3 Bilder, die für jede dieser Messungen charakteristisch sind, welche die Abstände zwischen den Cosmiden mit einer Genauigkeit von 1,5 bis 3 kb für 50 bis 80 Messungen angeben. Der Kode der eingesetzten Cosmide ist angegeben wie auch das Contig, zu welchem sie gehören (in Klammern). Die grüne Farbe ist weiß dargestellt und die rote Farbe in Grau.

MATERIAL UND METHODEN

Methode der Histogramme

[0128] Das Schema der [Fig. 1a](#) führt eine bestimmte Anzahl von Situationen auf in dem Falle einer Sonde, welche die gesuchte Region des Genoms (Region, die eine jegliche genomische Sequenz enthält, deren Modifikation das Auftreten der Pathologie zur Folge hat) vollständig überspannt. Die Sonde überspannt vorzugsweise einen Teil des Genoms, welcher an dem Bruch nicht beteiligt ist, d.h. beispielsweise auf beiden Seiten der Deletion.

[0129] Die entsprechenden theoretischen Histogramme (unter der Hypothese, dass die gekämmt genommene DNA intakt ist, was natürlich wenig realistisch ist, und was in der Folge noch diskutiert werden wird) sind in dem Schema der [Fig. 1b](#) angegeben. Die [Fig. 1c](#) repräsentiert die rohen Ergebnisse vor der Subtraktion des Beitrags des normalen Allels. Ähnliche Situationen sind in den Schemata der [Fig. 2a](#) und [Fig. 3a](#) repräsentiert in dem Fall, wo der eingesetzte Klon das untersuchte Gen (oder die untersuchte Region des Genoms) nicht vollständig überspannt.

[0130] In der Folge wird unterschiedslos der Ausdruck „diese Genomregion“ (stillschweigend einbezogen

eine solche, die an der Pathologie beteiligt ist) oder „das Gen“ verwendet im Rahmen dessen, wo die Technik nicht erlaubt, zu bestimmen, im eigentlichen Sinne von einem Gen, sondern von einer Region des Genoms, welche Veränderungen durchläuft (welche in einer „Bruchstelle“ zum Ausdruck kommen) zu sprechen.

[0131] Die Gesamtheit der in den Schemata dargestellten Situationen vermittelt eine Vorstellung von dem erwarteten Unterschied zwischen einer Situation, wo die eingesetzte klonierte Sonde mit einer Region des Genoms eines gesunden Individuums hybridisiert, die in dem Falle eines kranken Individuums modifiziert ist.

[0132] In allen aufgeführten Fällen (bis auf jenen einer Translokation der Gesamtheit oder eines Teils des Gens, enthaltend die Gesamtheit des Klons – siehe Schema 3a (c3)), unterscheidet sich das (theoretische) Histogramm von dem „normalen“ Histogramm:

- entweder durch das Vorhandensein eines einzigen Peaks, welcher sich aber bei einem unterschiedlichen Wert der Länge des Klons befindet (höher – beispielsweise Fall (d2), oder niedriger- Fall (b1)),
- oder durch das Vorhandensein von zwei Peaks oder mehr.

[0133] Die einfache Grundlage des Histogramms der Längen und der Länge L_c des Klons (erhalten beispielsweise mit Hilfe einer Hybridisierung mit der DNA eines gesunden Individuums und Ermittlung des Peaks des entsprechenden Histogramms) erlaubt folglich, zu wissen, ob der untersuchte Klon eine Bruchstelle überspannt (in dem hier verstandenen breiten Sinne).

[0134] Die auf der Ebene des Histogramms beobachteten Anomalien können in mehrere Kategorien unterteilt werden:

- (a) ein Peak für eine Länge $L_i < L_c$
- (b) ein Peak für eine Länge $L_s > L_c$
- (c) mehrere Peaks mit Länge(n) unter L_c
- (d) mehrere Peaks mit Länge(n) über L_c
- (e) ein Peak für eine Länge $L = L_c$ und ein oder mehrere Peaks mit Länge(n) unter L_c
- (f) ein Peak für eine Länge $L = L_c$ und ein oder mehrere Peaks mit Länge(n) über L_c
- (g) ein oder mehrere Peaks mit Längen $<$, $=$ und $> L_c$.

[0135] Die zuvor in einer Aufstellung aufgeführten Klassen von Anomalien reichen im Allgemeinen bedauerlicherweise nicht aus, um die zugrundeliegende genetische Anomalie zu identifizieren.

[0136] Das Schema der [Fig. 1b](#) zeigt tatsächlich, dass es a priori nicht möglich ist, zwischen einer partiellen Translokation (c1) oder (c2), vollständigen Translokation (c3) oder einer Insertion zu unterscheiden in dem Falle eines Klons, der die Gesamtheit des Gens überspannt. Das Gleiche gilt für den Fall, wo der Klon das Gen lediglich teilweise überspannt (Schema der [Fig. 2b](#) und [Fig. 3b](#)).

[0137] Im Gegensatz dazu erlaubt die Grundlage der Längen, die den Peaks des Histogramms entsprechen, in bestimmten Fällen die Position des Endes des deletierten, translozierten, verdoppelten oder in Kontakt mit einer Insertion befindlichen Abschnitts des Gens in dem Klon zu präzisieren.

[0138] In jedem fraglichen Fall signalisiert die Messung eines Histogramms, welches sich von jenem, welches im Falle eines normalen Genoms erwartet wird, unterscheidet, das Vorhandensein einer Anomalie auf der Ebene der in dem eingesetzten Klon enthaltenen Sequenz in dem untersuchten Genom.

- (a: ein Peak für eine Länge $L_i < L_c$): es handelt sich nahezu sicher um eine (oder um mehrere) Deletion(en). Indessen kann es sich, wie wir angemerkt haben, um einen Fall handeln, wo mehrere Peaks miteinander verschmolzen sind. Wir werden sehen, dass es im Allgemeinen möglich ist, zwischen diesen beiden Fällen zu unterscheiden durch Vergleich mit einer Sonde, welche einer unterschiedlichen Zone des Genoms entspricht (Referenzsonde).

[0139] In dem Falle einer Deletion ist es möglich, deren Größe durch Subtraktion der beiden restlichen Längen von der Länge des intakten Klons abzuleiten (einzig in dem Falle, wo der Klon die Gesamtheit des deletierten Abschnitts des Gens überspannt). Die Position bezogen auf den Klon kann im Gegensatz dazu nicht bekannt sein, da nur ergänzende Hybridisierungen mit anderen Klonen oder Subklonen diese Bestimmung erlauben können.

[0140] Indessen kann die gleiche Art von Histogramm einer Mehrzahl von Deletionen entsprechen, in welchem Falle es lediglich möglich ist, die gesamte Größe der Deletionen zu messen (stets unter der Hypothese, dass der Klon die Gesamtheit des deletierten Abschnitts des Gens überspannt).

[0141] In den Fällen, wo der Klon das Gen lediglich teilweise überspannt, kann die detektierte Deletion lediglich einen Teil der effektiven Deletion des Gens repräsentieren (siehe beispielsweise Schema 2a (b1)).

(b: ein Peak für eine Länge $L_s > L_c$): in diesem Falle handelt es sich um eine oder um mehrere Verdoppelungen im Inneren des Gens oder an seinen Enden eines Teils oder der Gesamtheit von jenem. Der Klon kann das Gen vollständig oder teilweise überspannen. Es kann darin gegebenenfalls eine Inversion der Sequenzen geben, aber dies kann in dem Falle, wo der Klon das Gen vollständig überspannt, nicht nachgewiesen werden. Die Position des oder der Verdoppelungen kann nicht bestimmt werden.

(c: mehrere Peaks mit Länge(n) unter L_c): es kann sich in diesem Fall um partielle Translokationen des Gens oder um die Insertion einer Sequenz im Inneren des Gens handeln. Unter der Hypothese einer Translokation ist es möglich, die beiden Enden des Gens innerhalb des Klons zu präzisieren (Fall, wo der Klon das Gen vollständig überspannt) oder zwei Möglichkeiten für diese Enden als Behauptung aufzustellen (in dem Falle, wo der Klon das Gen teilweise überspannt).

(d: mehrere Peaks mit Länge(n) über L_c): es handelt sich in diesem Falle um eine Mehrzahl von Verdoppelungen der Gesamtheit oder eines Teils des Gens in der Region des Genoms, welche durch den Klon überspannt wird.

(e: ein Peak für eine Länge $L = L_c$ und ein oder mehrere Peaks mit Länge(n) unter L_c): es handelt sich um eine (oder mehrere) Verdoppelungen der Gesamtheit oder eines Teils des Gens in einer Zone des Genoms, welche sich von der durch den Klon überspannten Zone unterscheidet, dies in dem Falle, wo der Klon die Gesamtheit oder einen Teil des Gens überspannt.

(f: ein Peak für eine Länge $L = L_c$ und einen oder mehrere Peaks mit Länge(n) über L_c): obgleich diese Situation in den Schemata 1 bis 3 nicht dargestellt worden ist, handelt es sich um einen Fall einer Verdoppelung des Typs (d), welcher mit einem normalen Gen verbunden ist.

(g: ein oder mehrere Peaks mit Längen $<, =$ und $> L_c$): Kombinationen von (e) und (f).

[0142] In einer bestimmten Anzahl von Situationen müsste die Analyse der Hybridisierungssignale und von deren Abstand in dem Falle, wo sie ausgerichtet sein sollten, außerdem ergänzende Informationen liefern.

[0143] Das Beispiel des Schemas 1a (e) wird diese Idee veranschaulichen: in diesem Falle einer Insertion muss man sich vorstellen, dass ein jegliches nicht hybridisiertes Segment, welches aber von zwei kolinearen Hybridisierungen (von irgendwelchen Größen) umgeben ist, eine Insertion darstellen kann: man wird folglich ein Histogramm von diesen Messungen konstruieren: wenn in der Realität eine Insertion existiert, müsste sie folglich als ein Peak auftreten genauso wie irgendeine Hybridisierung.

[0144] Diese Maßgabe wird gleichfalls in dem Falle von Deletionen oder Translokationen beispielsweise zutreffen.

[0145] Die zuvor dargelegte Technik setzt voraus, dass die hybridisierte gekämmte genomische DNA intakt ist. Tatsächlich wird die Referenzsituation, jene eines gesunden Individuums, dessen DNA, welche dem für die Hybridisierung eingesetzten Klon entspricht, nicht modifiziert ist, theoretisch durch einen Peak mit dem Wert der Länge des Klons repräsentiert.

[0146] In der Praxis wird die DNA während deren Präparation und gleichfalls während des Kämmens hydro-mechanischen Beanspruchungen ausgesetzt, was zu zahlreichen zufälligen Brüchen führt. Diese zufälligen Brüche können innerhalb der Zielsequenz der Sonde stattfinden, was nach dem Kämmen zu einer Auftrennung in mehrere Stücke der Sequenz führen wird. Während der Hybridisierung mit der Sonde werden diese Segmente hybridisiert und nach dem Nachweis oder der Sichtbarmachung als Fragmente mit einer geringere Größe als jene der Zielsequenz gemessen.

[0147] Diese Brüche beeinflussen nicht notwendigerweise die Gesamtheit der Zielsequenzen in der Gesamtheit der für das Kämmen eingesetzten Genome. Abhängig von der Anzahl von gekämmten humanen Genomen auf der untersuchten Oberfläche und abhängig von den Präparationsbedingungen wird der Anteil der Zielsequenzen mit Brüchen vernünftig niedrig bleiben. In diesem Falle wird das durch einen einzigen Peak mit der Länge der Sonde gebildete ideale Histogramm ersetzt werden durch einen weniger bedeutenden Peak, welcher von einer sich anschließenden Verteilung mit geringeren Längen begleitet wird.

[0148] Die Erfahrung und praktische elementare Erwägungen zeigen, dass unter den beschriebenen Versuchsbedingungen eine optimale Anzahl von gekämmten Genomen pro untersuchte Oberfläche wenigstens etwa zehn pro gekämmte Oberfläche ist.

[0149] Im Folgenden werden Simulationen von Bruchprozessen (in Folge der Präparation und des Kämmens

der DNA) abhängig von zwei Parametern untersucht, welche erlauben, den Einfluss dieser Phänomene auf das theoretische Histogramm der Längen auszuwerten.

[0150] Zwei Parameter scheinen bei dem Phänomen des Bruches von Bedeutung zu sein: die charakteristische Größe und der Bruchanteil.

[0151] Die charakteristische Größe definiert ungefähr die Grenzgröße, die durch Moleküle, die einer großen Anzahl von Manipulationen unterworfen werden, erreicht wird. Ein Beispiel für eine mögliche Manipulation ist das Pipettieren: abhängig von dem Durchmesser des Pipettenkegels wird die maximale Grenzgröße der Moleküle mehr oder weniger bedeutend sein. Ein anderes Beispiel für eine Manipulation ist das Vortexen der DNA-Lösung.

[0152] Der Bruchanteil ist sinnvoll, um die Anzahl von Manipulationen im Modell nachzuempfinden: beispielsweise riskieren, je mehr die Anzahl von Pipettierungen zunimmt, umso mehr Moleküle, gebrochen zu werden.

A. Treppenstufen-Modell.

[0153] Die einfachste Weise, DNA-Manipulationen in Lösung im Modell nachzuempfinden, besteht darin, eine Bruchwahrscheinlichkeit von Null den Fragmenten mit einer Größe unter der charakteristischen Größe L_0 und eine Bruchwahrscheinlichkeit gleich 1 den Fragmenten mit einer Größe über L_0 zuzuweisen. Der Bruch eines Fragments mit der Größe L erfolgt außerdem zufällig in zwei beliebige Fragmente:

$$L < L_0 \Rightarrow P_0(L) = 0$$

$$L > L_0 \Rightarrow P_0(L) = 1$$

[0154] Um einen Bruchanteil in dieses Modell einzuführen, legt ein Parameter die Anzahl von Bruchprozessen, denen ein Molekül und seine nachfolgenden Fragmente unterworfen werden, fest.

[0155] Dieses Modell erlaubt, die während einer Hybridisierung einer Sonde mit der bekannten Länge L_c erwarteten Histogramme abhängig von den Parametern des Modells zu simulieren. Die gewählten Bedingungen sind: Sonde, die mit einem Chromosom von 50 Mb hybridisiert, 200 gekämmte Genome, $L_c = 50$ kb.

A.1. Einfluss der charakteristischen Größe auf das Histogramm

[0156] Die [Fig. 4](#) repräsentiert für unterschiedliche Bruchanteile (oder Anzahl von Bruchschritten) den Verlauf der theoretischen Histogramme, die für Werte der charakteristischen Größe (L_0 , L_{break}) des 1-, 2-, 4- oder 8-fachen der Größe der Sonde erhalten werden. Es erscheint klar, dass, je näher die charakteristische Größe der Größe der Sonde kommt, sich das Histogramm umso mehr von dem einzigen Peak, der in dem Teil 2 beschrieben wurde, unterscheidet, wobei die sich anschließende Verteilung mit Längen unter der Länge L_c der Sonde an Bedeutung zunimmt, bis sie den Peak für den erwarteten Wert L_c praktisch verschwinden lässt.

A.2. Einfluss des Bruchanteils auf das Histogramm

[0157] Die [Fig. 5](#) präsentiert für unterschiedliche charakteristische Größen die Entwicklung des theoretischen Histogramms abhängig von dem Bruchanteil, d.h. der Anzahl von zufälligen Bruchschritten (N-fache Schritte). Es erweist sich natürlich, dass der Peak mit der erwarteten Länge L_c abnimmt, wenn der Bruchanteil zunimmt.

[0158] In diesem Modell verschwindet indessen der Peak nicht vollständig, sogar wenn der Bruchanteil unendlich anwächst, denn es findet keinerlei Bruch statt, wenn alle restlichen Fragmente eine Größe unter L_c aufweisen.

A.3. Vorteil des Modells

[0159] Das vorangegangene Modell weist als Hauptvorteil auf, ein schnelles Mittel bereitzustellen, um den wahrscheinlichen Verlauf des Histogramms der Längen der hybridisierten Sondenfragmente derart auszuwerten, dass man die Ergebnisse der Beobachtungen interpretieren kann.

B. Gauss-Modell

[0160] Ein ein wenig realistischeres Modell besteht in einem ähnlichen Modell von aufeinanderfolgenden Bruchschritten, aber mit einem Bruchgesetz, welches sich von dem „alles oder nichts“-Gesetz des Treppenstufen-Modells unterscheidet. Eine einfache Weise, um dieses Gesetz „abzurunden“, besteht darin, die vorangegangene Wahrscheinlichkeit durch ein Gaussches Bruch-Wahrscheinlichkeitsgesetz zu ersetzen:

$$P_G(L) = 1 - e^{-(L/L_0)^2}$$

[0161] B.1. Einfluss der charakteristischen Größe auf das Histogramm

[0162] Die [Fig. 6](#) zeigt für unterschiedliche Bruchanteile den Verlauf der theoretischen Histogramme, die für Werte der charakteristischen Größe (L_0 , L_{break}) des 1-, 2-, 4- oder 8-fachen der Größe der Sonde erhalten werden, die identisch sind mit jenen, die bei der im Rahmen des Treppenstufen-Modells ausgeführten Simulation eingesetzt worden sind. In gleicher Weise unterscheidet sich das Histogramm von dem einzigen Peak, der in dem Abschnitt 2 beschrieben wird, umso mehr, je näher die charakteristische Größe der Größe der Sonde kommt, wobei die sich anschließende Verteilung mit Längen unter der Länge L_c der Sonde an Bedeutung zunimmt, bis sie den Peak für den erwarteten Wert L_c praktisch verschwinden lässt.

B.2. Einfluss des Bruchanteils auf das Histogramm

[0163] Die [Fig. 5](#) zeigt für unterschiedliche charakteristische Größen die Entwicklung des theoretischen Histogramms abhängig von dem Bruchanteil, d.h. der Anzahl von zufälligen Bruchschritten (N-fache Schritte). Es erweist sich natürlich, dass der Peak mit der erwarteten Länge L_c abnimmt, wenn der Bruchanteil zunimmt.

[0164] Gleichwohl weist dieses Modell ein realistischeres Verhalten als das Vorangegangene auf, denn, wenn der Bruchanteil wächst, nimmt der L_c entsprechende Peak ab unabhängig von dem Wert der charakteristischen Größe, um vollständig zu verschwinden, wenn der Bruchanteil ins Unendliche wächst.

B.3. Vorteil des Modells

[0165] Das Gauss-Modell, welches realistischer ist, wird jenes sein, das wir in der Folge verwenden werden, um Histogramme von Längen von hybridisierten Sondenfragmenten zu simulieren.

C. Simulation einer Hybridisierung mit einer Bruchstelle

[0166] Um die Bedingungen für die Anwendbarkeit unserer Technik zu untersuchen, werden wir die Simulation von drei Arten von charakteristischen Situationen der Hybridisierung eines Klons mit einer Bruchstelle durchführen. Wir werden als Beispiel den Fall eines BAC („Bacterial Artificial Chromosome“, künstliches Bakterienchromosom) von 125 kb, welches eine Bruchstelle, die an einer Translokation beteiligt ist (siehe beispielsweise Schema 2a (c1)), überspannt, heranziehen.

[0167] Die in der [Fig. 8d](#) dargestellten Simulationen entsprechen einer Situation, wo der Klon mit einer normalen genomischen DNA hybridisiert wird: unter diesen Bedingungen haben die meisten Hybridisierungen eine Länge von $L_c = 125$ kb abgesehen von den aufgrund der Manipulationen gebrochenen Fragmente. Wir haben zwei Werte für die charakteristische Größe L_0 ausgewählt: $L_0 = 100$ ($< L_c$) und $L_0 = 200$ ($> L_c$). Der Bruchanteil ist einheitlich und gleich $N = 10$.

[0168] Diese Histogramme zeigen, dass es möglich ist, einen Peak bei 125 kb in den meisten Fällen, sogar ziemlich ungünstigen Fällen, wie in dem Falle $L_0 = 100$, $N = 10$, zu unterscheiden.

C.1. Fragmente von 35 und 90 kb

[0169] Die [Fig. 8a](#) zeigt die Ergebnisse der Simulation eines Histogramms der Längen von hybridisierten Fragmenten in dem Falle, wo das BAC von 125 kb in zwei nicht zusammenhängenden Abschnitten des Genoms hybridisiert, mit sehr unterschiedlichen Größen (35 und 90 kb) wie in dem in dem Schema 2a (1c) beschriebenen Falle.

[0170] In den beiden Fällen ($L_0 = 100$, $L_0 = 200$) sind die beiden Peaks bei 35 und 90 kb erkennbar, selbst wenn die Situation insgesamt günstiger ist in dem Falle, wo die charakteristische Größe deutlich größer als das

größte Fragment ist.

[0171] Wir merken außerdem an, dass sogar in dem Falle einer kleineren Probennahme, die zu einem weniger deutlichen Peak bei 90 kb führt, der Peak bei 35 kb bestehen bleiben wird, welcher eine Anomalie bezogen auf die normale Größe des Klons signalisiert.

C.2. Fragmente von 50 und 75 kb

[0172] Die [Fig. 8b](#) repräsentiert das Ergebnis von Simulationen in dem Falle von zwei Fragmenten mit bezogen aufeinander wenig unterschiedlichen Größen. Auch hier wieder sind die entsprechenden Peaks deutlich und voneinander unterschieden. In diesem Falle kann man vorhersehen, dass mit einer geringeren Probennahme es möglich bleiben wird, diese beiden Peaks zu identifizieren.

C.3. Fragmente von 60 und 65 kb

[0173] Die [Fig. 8c](#) repräsentiert das Ergebnis von Simulationen in dem Falle von zwei Fragmenten von sehr ähnlicher Größe. In den beiden Fällen treten die Peaks deutlich aus dem Histogramm der kleinen Fragmente hervor, aber es ist zu befürchten, dass die begrenzte Genauigkeit der Messungen diese zu einem einzigen verschmelzen lässt.

[0174] Auch hier wieder wird eine solche Information indessen ausreichen, um die Existenz einer Bruchstelle, welche sich in halbem Abstand von dem hybridisierten BAC befindet, sicherzustellen.

C.4. Schlussfolgerungen

[0175] Die vorangegangenen Simulationen zeigen, dass unter Bedingungen, die vermeiden, die DNA vor und während des Kämmens zu oft zu brechen, es möglich ist, unzweideutig zwischen einer Vergleichssituation, wo die Sonde mit einem normalen humanen Genom hybridisiert, und einer Situation, wo die Sonde mit einem durch das Vorhandensein einer Bruchstelle modifizierten Genom hybridisiert, zu unterscheiden.

[0176] Die heuristischen Parameter, welche das Kämmen charakterisieren, zeigen, dass es manchmal zu bevorzugen ist, eine charakteristische Größe zu haben, die über der maximalen Länge der Hybridisierung liegt oder gleich dieser ist. Wenn jene unbekannt ist, ist ein einfaches Kriterium folglich eine charakteristische Größe in der Größenordnung von oder über der Länge des als Sonde eingesetzten Klons.

[0177] Um experimentell die Realisation dieses Kriteriums zu bestimmen, scheint es folglich wünschenswert, zusätzlich zu den Sonden, die für die Untersuchung einer Anomalie in dem Genom bestimmt sind, eine Vergleichssonde mit einer vergleichbaren Größe, die mit einem unterschiedlichen Abschnitt des Genoms hybridisiert, zu verwenden. Das entsprechende Histogramm wird so erlauben, qualitativ die Parameter des Modells, welche die experimentelle Situation reflektieren, zu bestimmen.

[0178] Die [Fig. 8e](#) veranschaulicht eine experimentelle Situation des zuvor in der [Fig. 8a](#) beschriebenen Beispiels. Humane genomische DNA wurde gemäß dem nachfolgend beschriebenen Protokoll gekämmt. Fünf aneinandergrenzende Cosmid-Sonden des Chromosoms 9 (280A6, 37A1, 149B8, 134A11, 99B4, welche zu dem Contig des TSC1-Gens gehören), welche insgesamt ungefähr 155 kb ausmachten, wurden mit dieser DNA hybridisiert, dann mit fluoreszierenden Antikörpern sichtbar gemacht. Es wurde ein Histogramm der gemessenen hybridisierten Längen gezeichnet, das mit dem Histogramm der [Fig. 8d](#) für $L_{\text{break}} = 100$ sehr stark vergleichbar ist. Es wird außerdem festgehalten, dass das reale Histogramm mittels eines Äquivalents von weniger als 40 diploiden Genomen erhalten worden ist (wobei die Simulationen 100 diploiden Genomen entsprechen).

Oberflächenbehandlung

[0179] Plättchen von 22 × 22 mm werden vorbereitet und mit Silan bedeckt unter Verwendung des in den Patentanmeldungen WO 95/22056 und WO 95/21939 beschriebenen Verfahrens.

Herstellung der Cosmid-Sonden

[0180] Die Cosmid-DNA wird durch eine Verlängerung ausgehend von einem statistischen Primer mit Nucleotiden, die mit Digoxigenin modifiziert oder die biotinyliert sind, markiert. Für die mit Biotin markierten Sonden verwendet man den DNA Bioprime™-Markierungs-„Kit“ (Gibco-BRL), welcher statistische octamere Primer

und 14-dCTP-Biotin enthält. Für die Markierung mit Digoxigenin wird eine unterschiedliche Mischung von dNTPs verwendet, die dCTP, dATP und dGTP (0,1 mM), dTTP (0,065 mM) und dig-11-dUTP (0,033 mM) enthält.

[0181] Die Größe und die Konzentration der markierten DNA-Fragmente werden durch Elektrophorese auf einem 0,6%-igen Agarosegel und durch Densitometrie der Banden bestätigt. Die Wirksamkeit der Markierung wird ausgehend von auf Nylon-Membranen aufgetragenen Flecken unter Verwendung von aufeinanderfolgenden Verdünnungen von Digoxigenin-Sonden und von Biotin-Sonden ausgewertet. Nach Inkubation mit einem alkalische Phosphataseanti-dig (AP) (Boehringer Mannheim) oder Streptavidin AP (Gibco-BRL) werden die Flecken mit N BT (NitroBlue Tetrazolium) und BCIP (5-Brom-4-chlor-3-indolylphosphat) von Gibco-BRL sichtbar gemacht. Die positiven Flecken werden mit den ausgehend von DNA-Kontrollproben, die durch dig oder durch Biotin markiert sind, verglichen.

[0182] Um den Hintergrund zu verringern, werden die Stämme abschließend an Bio-Spin 6-Säulen (Biorad) durch fünfminütige Zentrifugation bei 1500 g gereinigt.

Lösung von Hefe-DNA

[0183] Die Hefe-DNA, welche das YAC 774G4 (1600 kb) enthält, wird in Blöcken von 100 µl Agarose LMP 0,8% (1 µg/Block) unter Verwendung eines Protokolls zur Herstellung der PFGE-Standardblöcke hergestellt. Die Blöcke werden in 0,5 M EDTA bei 4°C aufbewahrt und mit 15 ml TE-Puffer (10 mM Tris/1 mM EDTA, pH 8) 2 h gespült, bevor sie eingesetzt werden. Jeder Block wird mit 3,3 µM YOYO-I (Molecular Probes) in 100 µl T₄₀E₂ (40 mM Tris/2 mM EDTA, pH 8) 1 h bei Umgebungstemperatur (RT) gefärbt. Die Agarose wird dann 45 min bei 68°C geschmolzen und 2 h bei 40°C unter Verwendung von 2 E β-Agarase I in einem NEB × 1 Agarose-Puffer (Biolabs) pro Block verdaut.

[0184] Die Verdünnung der DNA (0,25 µg/ml) in 50 mM MES (pH 5,5) erfolgt sehr vorsichtig derart, dass der Bruch der DNA-Stränge vermieden wird. Die Lösung wird dann in ein 4 ml-Teflon™-Reservoir gegossen, welches das Einbringen von 3 silanisierten 22 × 22 mm-Plättchen für das Kämmen erlaubt. Die DNA-Lösung kann gleichfalls bei 4°C mehrere Tage aufbewahrt werden.

Molekulares Kämmen

[0185] Die silanisierten Plättchen werden in das Teflon™-Reservoir, welches eine DNA-Lösung von Hefe-Gesamt-DNA (0,25 g/ml in 50 mM MES, pH 5,5) enthält, getaucht, bei RT inkubiert und nach 10 min Inkubation aus dem Reservoir unter Verwendung einer einfachen mechanischen Vorrichtung herausgezogen. Während der Inkubation verankern sich die DNA-Moleküle auf der Oberfläche durch ihre Enden. Indem die Oberfläche aus dem Reservoir herausgezogen wird, hat dies die gleiche Wirkung wie das Verdampfen, das in der „Tropfen-Methode“ („methode de la goutte“) vorgesehen wird; der Meniskus verschiebt sich bezogen auf die Oberfläche und übt eine konstante Zugkraft auf die Moleküle, die in dem Reservoir zurückbleiben, aus.

[0186] Die mit der gekämmten DNA bedeckten Oberflächen werden dann mittels eines Mikroskops in Epifluoreszenz beobachtet derart, dass die Charakteristiken des Kämmens kontrolliert werden. Eine Registrierung der repräsentativen Betrachtungs- oder Sichtfelder erfolgt für jedes Plättchen für die Posthybridisierungskontrolle.

[0187] Die Oberflächen werden dann auf Mikroskop-Träger geklebt (Cyanacrylat) und eine Nacht bei 60°C erwärmt. Sie können bei -20°C oder bei Umgebungstemperatur mehrere Monate, wenn man sie vor Feuchtigkeit schützt, aufbewahrt werden. Die Oberflächen werden dann vor der Denaturierung dehydratisiert, indem ein Bad, welches steigende Ethanol-Konzentrationen (70%, 90%, 100%) enthält, verwendet wird.

Denaturierung und Hybridisierung

[0188] Die Oberflächen werden in 70% entionisiertem Formamid/30% 2 × SSC, pH 7,0) während 4 min bei 70°C denaturiert und unverzüglich 5 min in ein kaltes Ethanol-Bad (0°C) mit steigenden Konzentrationen (70%, 90%, 100%) eingetaucht. Man lässt die Oberflächen dann trocknen.

[0189] 50 ng mit Biotin markierte Sonden und 50 ng mit Digoxigenin markierte Sonden werden dann mit 3 µg humaner Cot1-DNA und 10 µg Heringssperma-DNA in 10 µl Hybridisierungspuffer (50% entionisiertes Formamid/10% Dextransulfat/2 × SSC/1 % Tween 20, pH 7) gemischt. Die Sonden werden 5 min bei 80°C denaturiert und unverzüglich auf 0°C abgekühlt.

[0190] 10 µl der Sondenlösung werden pro gekämmtes 22 × 22 mm-Plättchen zugesetzt, dann mit einem nicht behandelten Plättchen bedeckt und mit einem Polymer vom Typ „Rubbercement“, welches von Sanford, USA, vertrieben wird, versiegelt. Die Hybridisierung wird eine Nacht bei 37°C in einer feuchten Kammer (HC) ausgeführt.

Sichtbarmachung/Nachweis mit fluoreszierenden Sonden

[0191] Nach der Hybridisierung werden die Plättchen 5 min bei RT in 3 Bädern (50% entionisiertes Formamid/2 × SSC, pH 7) und 5 min bei 5' bei RT in 3 Bädern mit 2 × SSC gewaschen. Die Plättchen werden 30 min bei 37°C (HC) mit 50 µl pro Plättchen einer Blockierungslösung (1,5% (Gew./Vol.) Reagens (Boehringer Mannheim) in 4 × SSC/Tween 20, 0,05%, pH 7,2) inkubiert.

[0192] Die Detektionen der mit Biotin und mit Digoxigenin markierten Sonden werden gleichzeitig ausgeführt unter Verwendung des gleichen Protokolls für jede Detektionsschicht, wobei jede Antikörperschicht 30 min bei 37°C inkubiert wird, wobei die Plättchen nach jeder Schicht gewaschen werden (dreimal 5 min bei RT in 4 × SSC/Tween 20, 0,05%) Für die mit Biotin markierten Sonden werden die folgenden Schichten eingesetzt (50 µl Hybridisierungspuffer pro Plättchen): (1) 40 mg/ml Avidin-Texas Red (Vector). (2) 5 mg/ml biotinyliertes Anti-Avidin von der Ziege (Vector). (3) 40 mg/ml Avidin-Texas Red (Vector).

[0193] Für die mit Digoxigenin markierten Sonden: (1) 34 mg/ml eines anti-dig-Konjugats mit FITC (Jackson). (2) 28 mg/ml FITC-anti-Maus vom Esel (Jackson). (3) 30 mg/ml FITC-anti-Kaninchen von der Maus (Jackson).

[0194] Nach der Detektion werden die Plättchen schnell in 1 × PBS gespült und vor der Beobachtung mit einem Reagens eingedeckt (Vectashield, Vector). Die Plättchen können bei 4°C im Dunkeln monatelang aufbewahrt werden.

BEISPIEL 1

[0195] In dem Falle des Kämmens von humaner genomischer DNA führt man gleichzeitige Hybridisierungen von zwei Cosmiden, die durch einen Zwischenbereich („gap“) von mehreren zehn kb voneinander getrennt sind (180F1 und 50D9) aus: auf einem einzigen 22 × 22 mm-Plättchen konnten ungefähr 80 gekoppelte Signale (Cosmid 1 – Zwischenbereich – Cosmid 2) gemessen werden, was einem Gesamtabstand von ungefähr 120 kb entspricht (siehe Beispiel 5).

[0196] Dieses Ergebnis erlaubt, sicherzustellen, dass es möglich ist, ungefähr einhundert Hybridisierungen eines vollständigen BAC mit einem normalen humanen Genom zu beobachten. Diese experimentelle Situation ist zu vergleichen mit den theoretischen Situationen, die in der **Fig. 8** ins Auge gefasst wurden: die Histogramme 8d zeigen tatsächlich, dass bei 100 diploiden Genomen (folglich 200 Auftreten einer klonierten Sequenz) und mit besonderen Parametern ungefähr 10 intakte Hybridisierungssignale eines BACs von 125 kb (Fall $L_0 = 100$ kb) oder 16 intakte Hybridisierungssignale (Fall $L_0 = 200$ kb) erwartet werden. Anders gesagt, sind die bei den Simulationen eingesetzten theoretischen Parameter viel pessimistischer als die experimentellen Bedingungen, die man gegenwärtig in der Lage ist, zu realisieren.

[0197] Die Gesamtheit dieser theoretischen und experimentellen Analyse erlaubt folglich eine erste Validierung der Erfindung.

[0198] Außer der Untersuchung von Bruchstellen mit Hilfe von BACs (oder anderen Sonden von äquivalenter oder geringerer Größe), die zuvor erwähnt worden ist, kann gleichfalls ins Auge gefasst werden, YACs (verfügbar für die Gesamtheit des menschlichen Genoms und für andere Genome) einzusetzen. Ähnliche Simulationen zu jenen, die oben präsentiert worden sind, zeigen, dass in Betracht gezogen werden kann, die Hybridisierung eines YACs von 1600 kb an zwei unterschiedliche Fragmente von 400 und 1200 kb (beispielsweise) auf der Grundlage von 100 gekämmten Genomen unter Bedingungen, die $L_0 = 600$, $N = 10$ in dem Gauss-Modell entsprechen, zu detektieren.

[0199] Indessen erfordert im Unterschied zu den Beobachtungen von BACs oder von Sonden von ähnlicher oder geringerer Größe, die ohne Schwierigkeiten mit Hilfe eines Epifluoreszenz-Mikroskops, ausgerüstet mit einem x100- oder x63-Objektiv und einer Kamera (maximale Größe eines Blickfelds: ungefähr 123 µm bzw. 195 µm) ausgeführt werden können, die Messung von Fragmenten von mehreren hundert kb die Verwendung von Objektiven mit geringerer Vergrößerung, für welche sich gegebenenfalls Probleme der Nachweisbarkeit der Signale stellen können:

x 40: Sichtfeld mit maximaler Größe von ungefähr 307 μm .

x 20: Sichtfeld mit maximaler Größe von ungefähr 614 μm .

[0200] Der Vorteil einer Verwendung von Sonden von so bedeutender Größe besteht selbstverständlich darin, die Anzahl von Hybridisierungen, welche notwendig sind, um einen interessanten Klon zu finden, zu verringern.

BEISPIEL 2

Quantitative Bestimmung von Phage λ -DNA im Genom von E. coli

[0201] Genomische DNA von E. coli ($l_c = 4,7 \text{ Mb}$), welche ein Exemplar des Genoms des Phagen λ enthält (Linie 5243, $l_t = 49 \text{ kb}$), wurde auf silanisierten Oberflächen, nachdem sie mit einem fluoreszierenden Molekül (YOYO-1) gegengefärbt worden war, gekämmt.

[0202] Die gesamte Länge der gekämmteten DNA pro Sichtfeld wurde ausgehend von 30 bis 50 Sichtfeldern, die gleichförmig auf der Gesamtheit von jeder Oberfläche verteilt waren, abgeschätzt.

[0203] Mit Biotin-dUTP markierte Phage λ -DNA wurde dann hybridisiert und mit Hilfe eines mit FITC (grün) gekoppelten Antikörpersystems sichtbar gemacht. Die gesamte Länge von hybridisierter DNA pro Sichtfeld wurde ausgehend von 100 Sichtfeldern, die gleichförmig auf der Gesamtheit von jeder Oberfläche verteilt waren, abgeschätzt.

[0204] Die Ergebnisse von 4 Experimenten sind in der nachfolgenden Tabelle aufgeführt, in welcher

$N_b \text{ E. coli} = \frac{L_c}{l_c}$ und $N_b = \frac{L_t}{l_t}$.

	l_c	l_t			
	$N_b \text{ E. Coli}$	$N_b \lambda$	$R = \lambda / \text{E. Coli}$	ΔR	
43_4	0,19	0,19	1,0	0,6	
43_6	0,25	0,22	0,9	0,5	
43_7	0,31	0,26	0,8	0,4	
43_8	0,20	0,24	1,2	0,7	
43_9	0,26	0,39	1,5	0,8	
43_13	0,57	0,74	1,3	0,4	
43_14	0,61	0,56	0,9	0,3	

[0205] Diese Ergebnisse haben, außer für das Plättchen 43_9, ein Verhältnis, das ausreichend nahe bei dem erwarteten Wert von 1 liegt, gezeigt. Indessen ist das eingesetzte Prinzip, welches darin besteht, die Gesamtheit der Signale vor der Hybridisierung zu messen, für größere Genome wenig praktikabel (die Unsicherheit ΔR ist hier hauptsächlich auf die geringe Anzahl von gemessenen E. coli-Genomen, in der Größenordnung von 6 bis 24, zurückzuführen).

BEISPIEL 3

Quantitative Bestimmung der Anzahl von Amplicons im Genom von Hamster-Zellen

[0206] Die Untersuchung der Durchführbarkeit wurde auf die genomische DNA eines Säugetiers ausgedehnt. Das gewählte System ist die DNA von 2 Linien von Fibroblasten aus der Lunge eines Hamsters (Linien 618 und GMA32), welche 1 bzw. 2 Kopien des Zielgens AMP1 pro haploides Genom enthalten.

[0207] Genomische DNA von bereitgestellten Zellen in Lösung wurde auf silanisierten Oberflächen mit einer abgeschätzten Dichte von 100 diploiden Genomen pro Oberfläche von $22 \times 22 \text{ mm}$ gekämmt.

[0208] Eine Cosmid-Sonde (D3S1, $\sim 40 \text{ kb}$), die für die Region des Zielgens spezifisch war, wurde mit dig-dUTP markiert. Eine andere Cosmid-Sonde (565.5A1, $\sim 40 \text{ kb}$), welche für eine ungefähr 1 Mb von dem Zielgen entfernt gelegene Region spezifisch war, wurde als Vergleich eingesetzt und mit Biotin-dUTP markiert.

[0209] Die Sonden wurden mit der gekämmten, zuvor denaturierten genomischen DNA hybridisiert und in rot (biotinylierte Sonden) und grün (digoxigenierte Sonden) in einem der Fälle und mit umgekehrten Farben in dem anderen sichtbar gemacht. Die Hybridisierungssignale von jeder Farbe wurden an einer Anzahl von Sichtfeldern, die für die Oberfläche repräsentativ war, gemessen.

[0210] In dem Falle der Linie A32 repräsentiert die gesamte, für die Sonde D3S1 erhaltene Größe (ungefähr 1400 µm) 70 Kopien des Gens. Die bei denselben Sichtfeldern für die Vergleichssonde erhaltene Größe (ungefähr 1180 µm) repräsentiert ihrerseits 60 haploide Genome. Diese Messung liefert folglich ein Ziel/Vergleichs-Verhältnis von $1,2 \pm 0,3$, welches mit dem Vorhandensein einer Kopie des Gens pro haploidem Genom verträglich ist.

[0211] Diese Experimente zeigen folglich die Durchführbarkeit der quantitativen Genbestimmung ausgehend allein von den Hybridisierungssignalen von Ziel- und Vergleichssonden, sofern eine ausreichende Anzahl von Signalen beobachtet werden kann.

BEISPIEL 4

Paarweise Kartierung von 6 Cosmiden auf einem YAC von 1600 bp, welches in der genomischen DNA von Hefe enthalten ist

[0212] Das YAC 774G4 enthält einen Klon von humaner genomischer DNA des Chromosoms 15, welcher das Calpain-Gen (CANP3), dessen Mutation für eine Dystrophie der Gürtel (LGMD 2A) verantwortlich ist, enthält. In Zusammenarbeit mit der Gruppe von J. Beckman vom G n thon (Evry) haben wir genomische Hefe-DNA, die in Low Melting-Agarose-Bl cken enthalten ist, (1 Block, 1 µg DNA pro Block) auf silanisierten Oberfl chen gek mmt.

[0213] Die sechs Cosmide wurden jeweils zu zweit hybridisiert und die Messung von deren Gr ssen und jeweiligen Abst nden erfolgte, indem Histogramme der Gr ssen und Abst nde erstellt wurden. Der Mittelwert und die Standardabweichung des Hauptpeaks von jedem Histogramm wurde daraus mit Hilfe einer speziellen Software extrahiert. Es erweist sich, dass die Standardabweichung der Messungen in der Gr ssenordnung von 2 bis 4 kb liegt ([Fig. 11](#)).

BEISPIEL 5

Kartierung von Cosmid-Paaren auf humaner genomischer DNA

[0214] 4 Cosmide, die zu 3 benachbarten Contigs, die durch Zwischenbereiche („gaps“) voneinander getrennt sind, geh ren, wurden in 2 Reihen von Hybridisierungen verwendet, die in Zusammenarbeit mit der Gruppe von S. Povey vom MRC (London) ausgef hrt wurden. Die Contigs  berspannen die Region des TSC 1-Gens, welches an einer der Formen der tuber sen Sklerose beteiligt ist (Chromosom 9). Die Cosmid-Sonden wurden gem   dem obigen Protokoll hergestellt.

[0215] Die eingesetzte genomische DNA wurde aus Zellkulturen extrahiert und in Low Melting-Agarose-Bl cke in einer Menge von 10^6 Zellen pro Block gegossen. 3 gem   dem vorangehenden Protokoll (Reservoir von 4 ml, Verwendung einer Endmolarit t von MES, pH 5,5, von 150 mM) behandelte Bl cke wurden f r das K mmen der genomischen DNA eingesetzt.

[0216] Die Cosmide wurden jeweils zu zweit auf gleichen Pl ttchen hybridisiert, wobei im Mittel eine Gesamtzahl von mehreren zehn doppelten Signalen (rot/gr n ausgerichtet) pro Pl ttchen beobachtet wurde. Es wurde das gleiche Messprotokoll wie zuvor eingehalten, was zu Endwerten mit der gleichen Genauigkeit wie in dem vorangegangenen Experiment f hrte.

[0217] Die [Fig. 12](#) pr sentiert einige typische Bilder, die die Messung der Gr ssen der zwei unter suchten Zwischenbereiche („gaps“) erlaubt haben.

BEISPIEL 6

Kartierung von Restriktionssegmenten

[0218] Diese Kartierungstechnik ist nat rlich auf jede andere Art von gek mmter DNA oder von Subklonen

anwendbar. Eine mögliche Erweiterung der Technik würde beispielsweise darin bestehen, nicht mehr Subklone eines Klon zu kartieren, sondern direkt Restriktionsfragmente der gekämmten DNA (beispielsweise ein Klon vom BAC-Typ).

[0219] Dies würde die Konstruktion von intermediären Subklonen vor der Sequenzierung vermeiden, da die physische Karte der Restriktionsfragmente mit ausreichender Genauigkeit erhalten werden würde, um eine Wiederherstellung der endgültigen Sequenz zu erlauben. Die Anwendbarkeit dieser Technik beruht auf der guten Auftrennung der Restriktionsbanden und der ausreichenden Größe (> 10 kb) der Hauptfragmente wie auch auf der späteren Subklonierung der DNA dieser Banden (Subklonierung in Vektoren von geringer Größe nach ergänzender enzymatischer Restriktion für die Sequenzierung).

Patentansprüche

1. Verfahren zur Positionsbestimmung von mehreren Klonen, die aus einem Genom hervorgegangen sind, **dadurch gekennzeichnet**, dass man:

- (a) eine gewisse Menge des Genoms auf einer Oberfläche zum Kämmen fixiert und kämmt,
- (b) das Produkt des Kämmens mit Sonden hybridisiert, die mit radioaktiven, fluoreszierenden oder anderen Elementen, wie Kügelchen, Teilchen, markiert sind, welche jedem Klon entsprechen, derart, dass die Sonden spezifisch nachgewiesen werden können,
- (c) die Information, die dem Ort jedes Klon entspricht, sowie die entsprechenden Größen und Entfernungen auf dem Genom entnimmt,
- (d) die Vorgehensweisen (b) und (c) n-mal wiederholt, indem man die Farbe, die Markierung oder die Nachweisart der Sonden modifiziert, wobei man weiß, dass man mit p verschiedenen Farben, Markierungen oder Nachweisarten am Ende von n Hybridisierungen die Position von p^n Klonen bestimmen kann.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man Sonden verwendet, welche mit modifizierten Nukleotiden markiert sind, die gegebenenfalls auf unterschiedliche Weisen nachgewiesen werden.

3. Verfahren nach einem der Ansprüche 1 bis 2, dadurch gekennzeichnet, dass man Sonden verwendet, die durch Einbau von Nukleotiden markiert sind, welche durch Biotinylierung, durch DIG oder andere Haptene modifiziert sind, die durch ein System von Schichten von Antikörpern oder spezifischen Molekülen nachgewiesen werden.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass man auch Sonden verwendet, die durch fluoreszierende Nukleotide markiert sind.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, dass man auf der gekämmten Oberfläche mindestens zehn Kopien des Genoms anordnet.

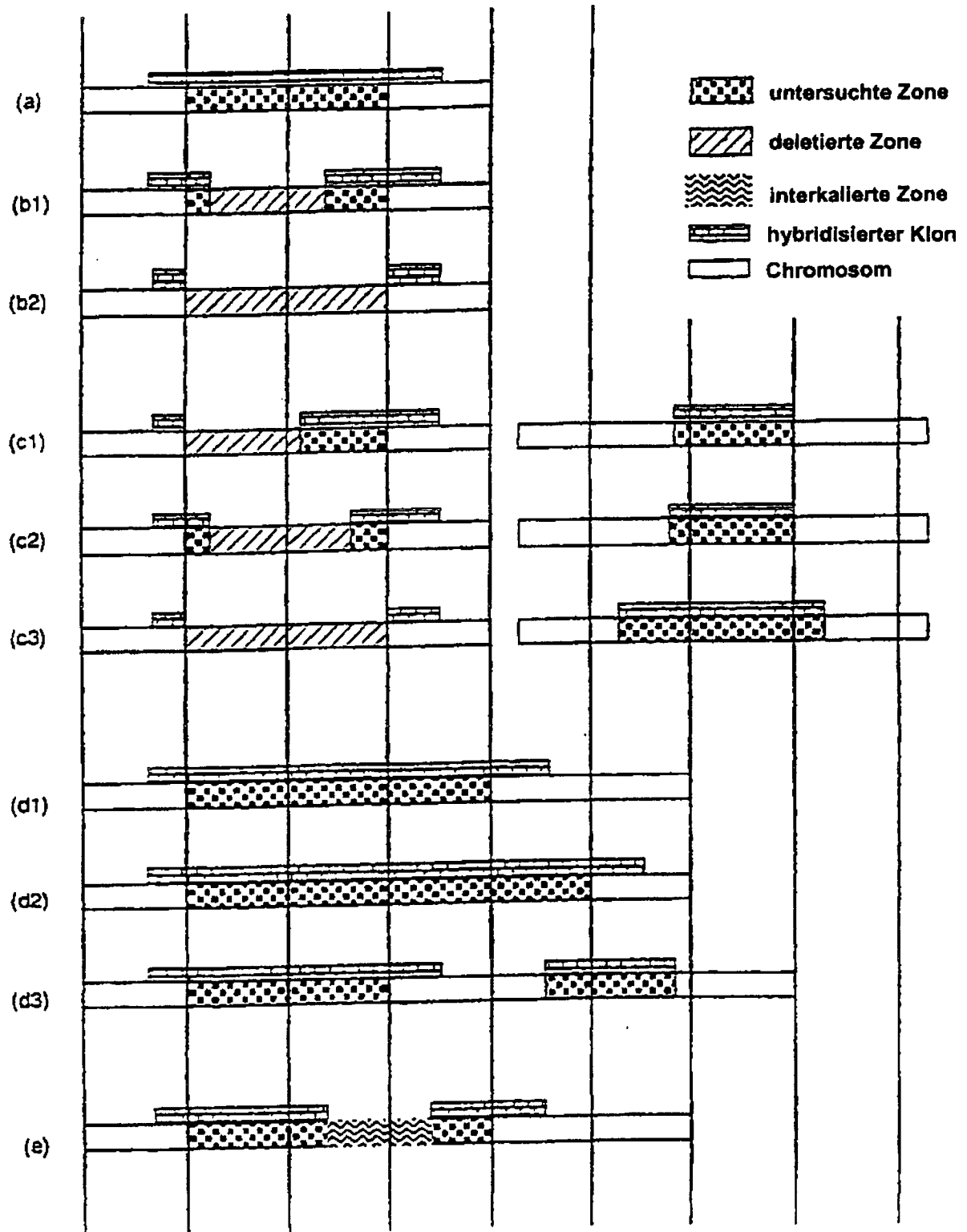
6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, dass die Oberfläche eine Kalibrierungs-DNA umfasst, welche es ermöglicht, jede Messung zu kalibrieren.

7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, dass das Genom aus einer Probe von biologischer Flüssigkeit oder einem Gewebe biologischen Ursprungs hervorgeht.

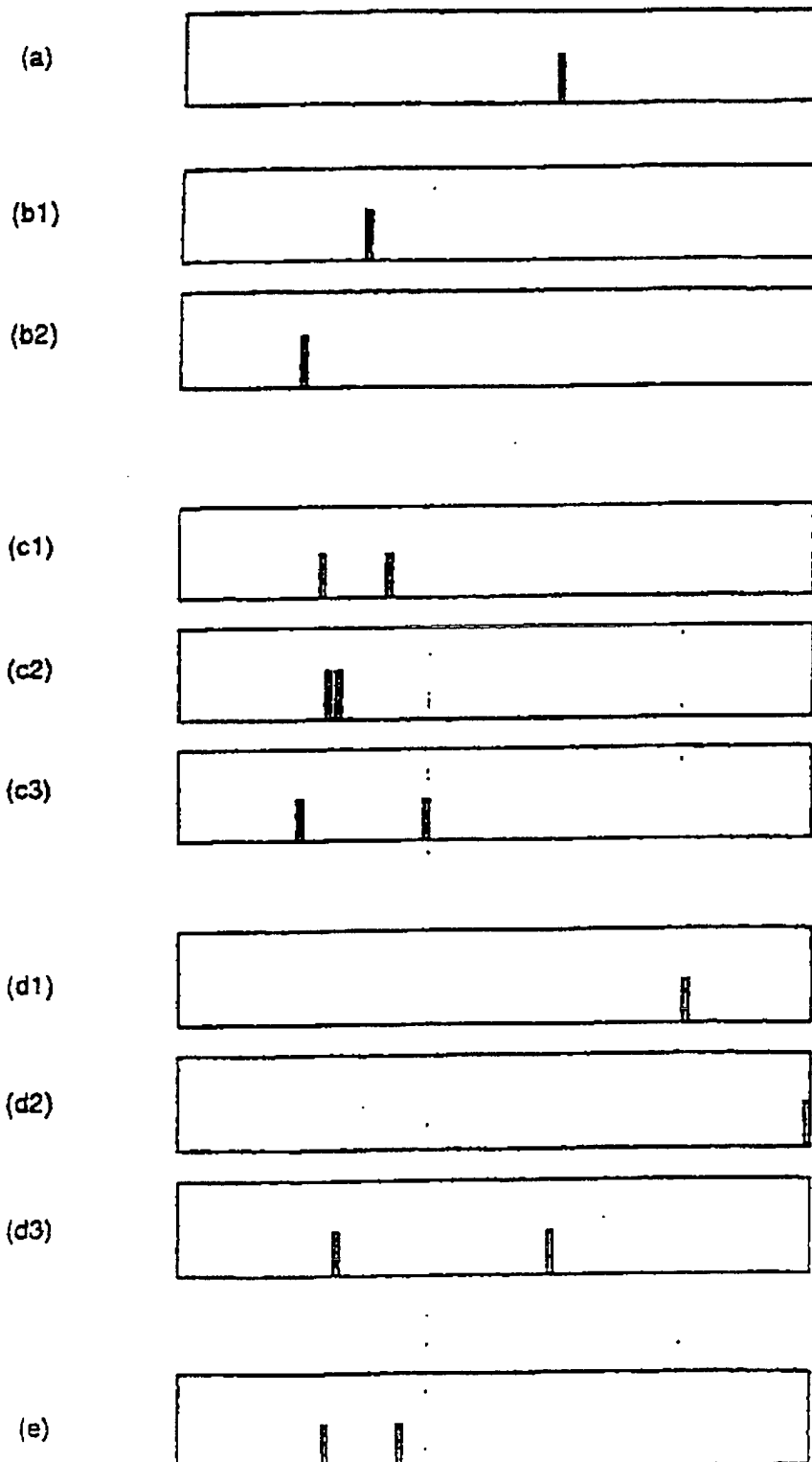
8. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, dass das Genom aus einem biologischen Material hervorgeht, das mindestens 80% genetisches Material fötalen Ursprungs umfasst.

Es folgen 22 Blatt Zeichnungen

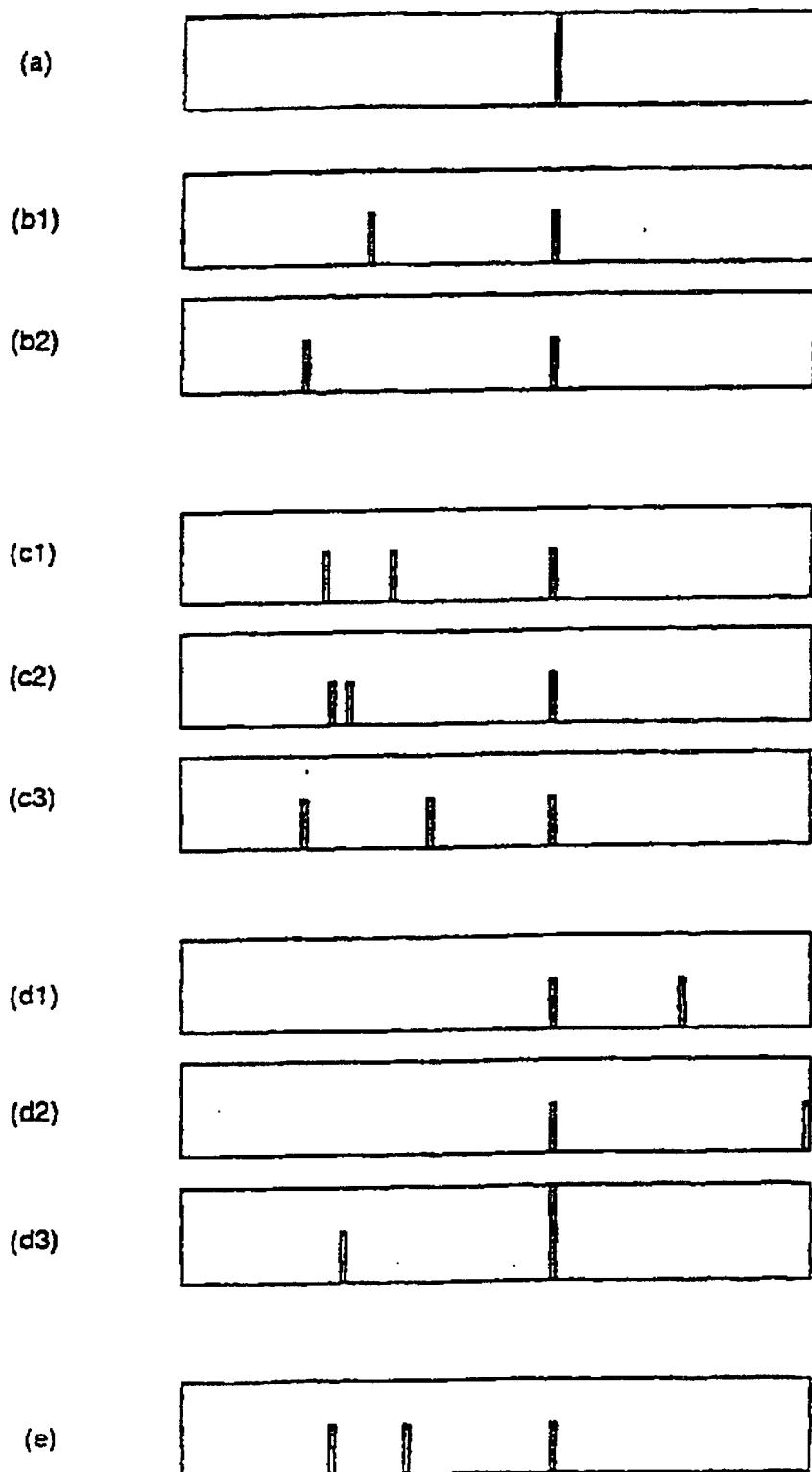
Anhängende Zeichnungen



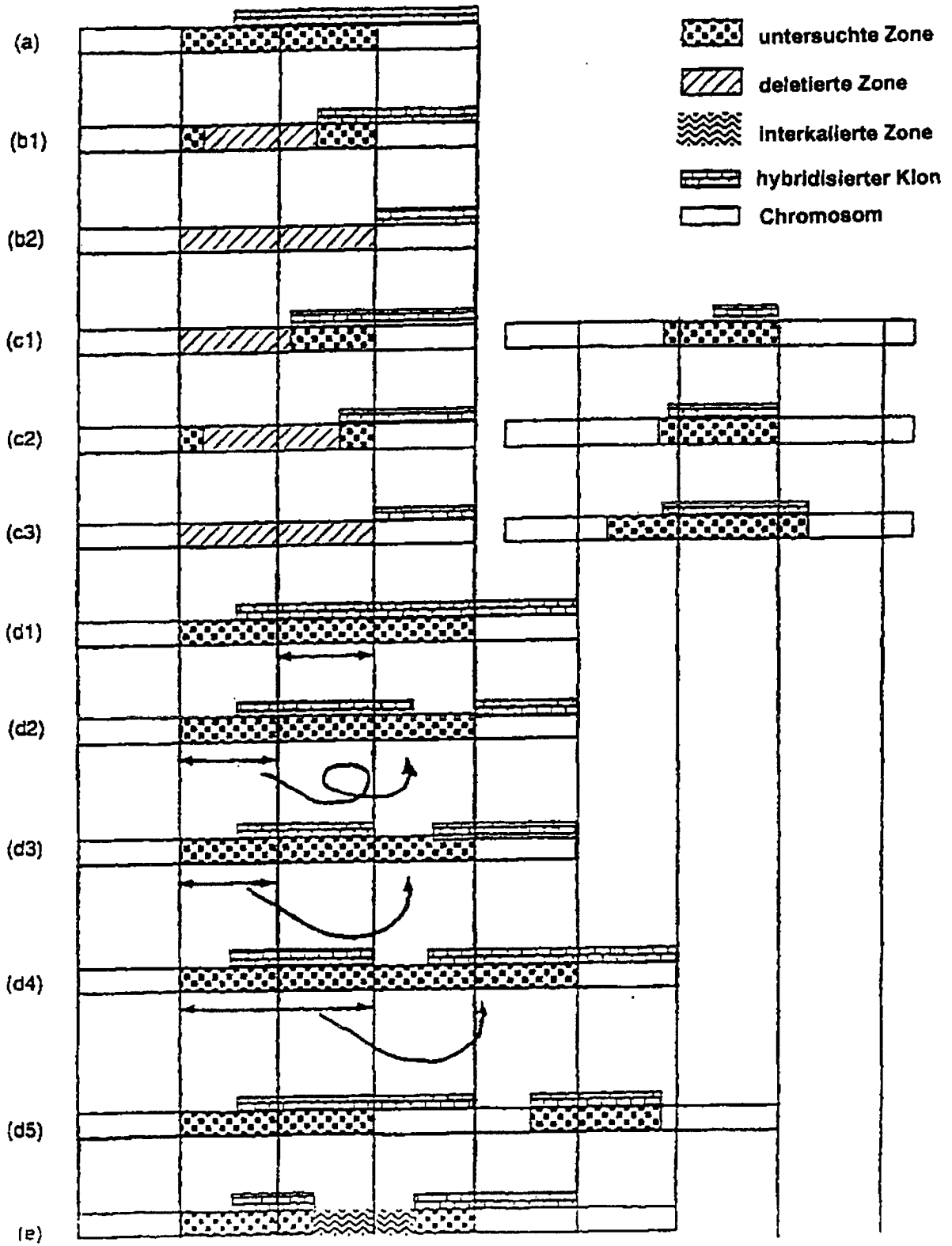
Figur 1a



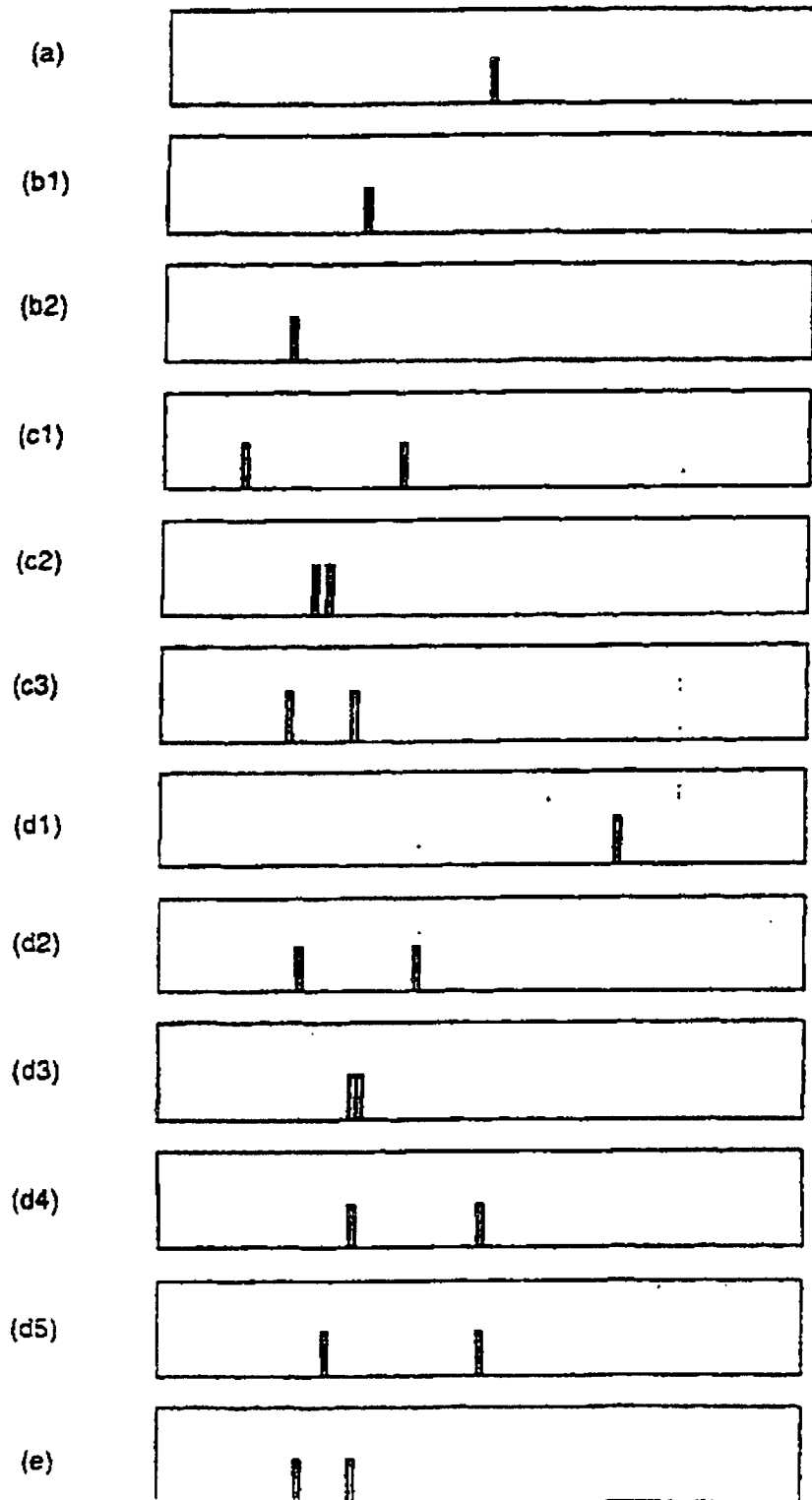
Figur 1b



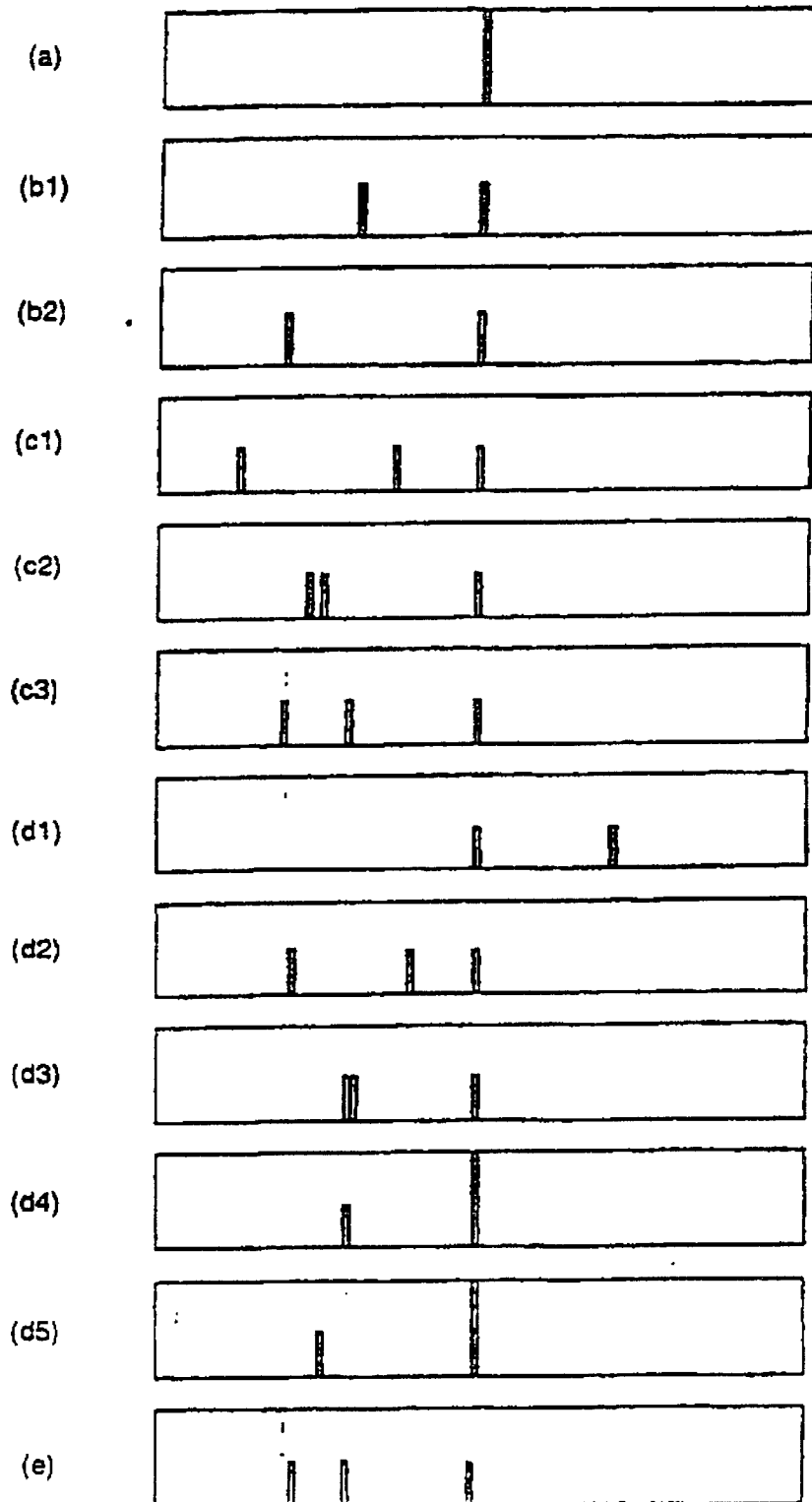
Figur 1c



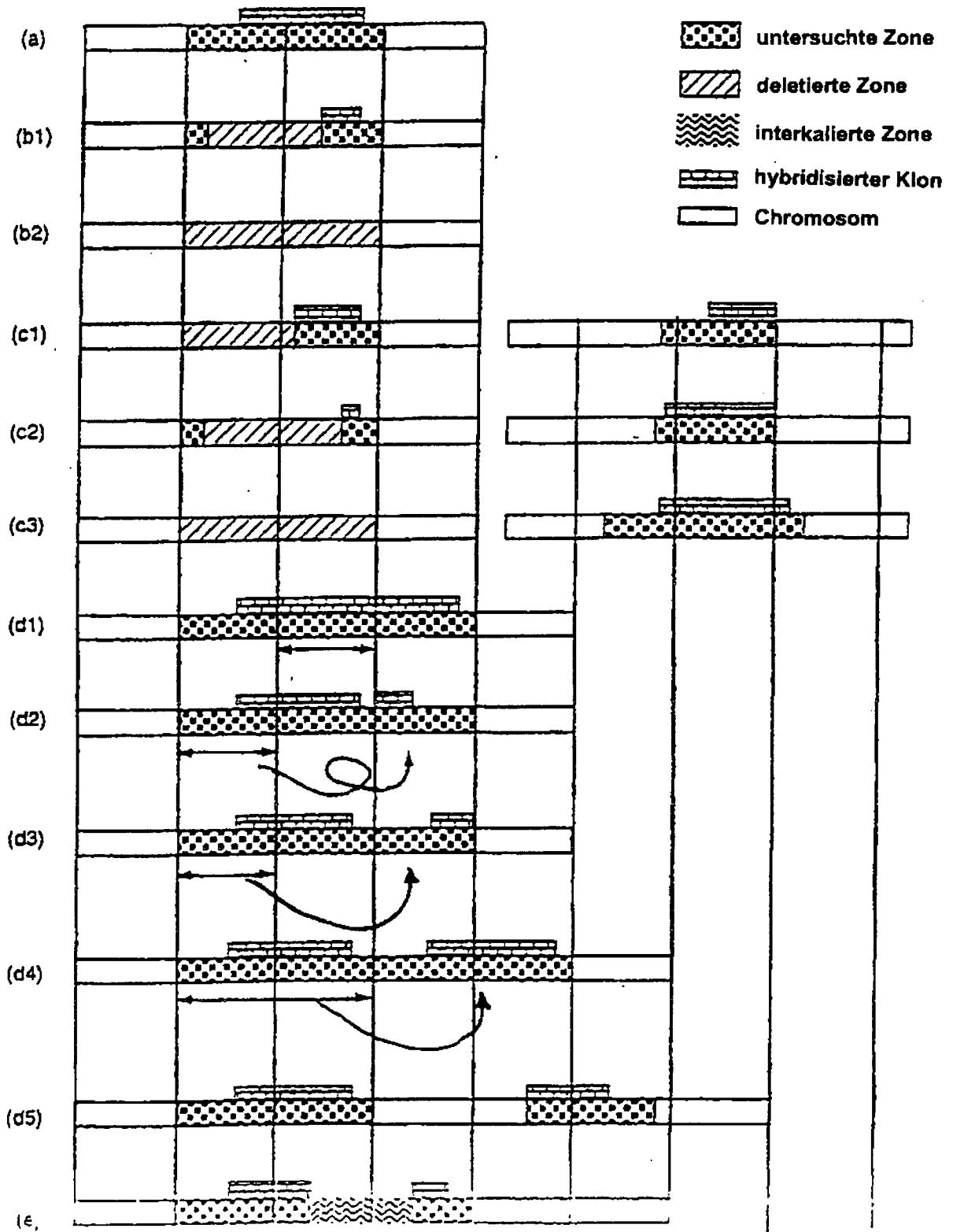
Figur 2a



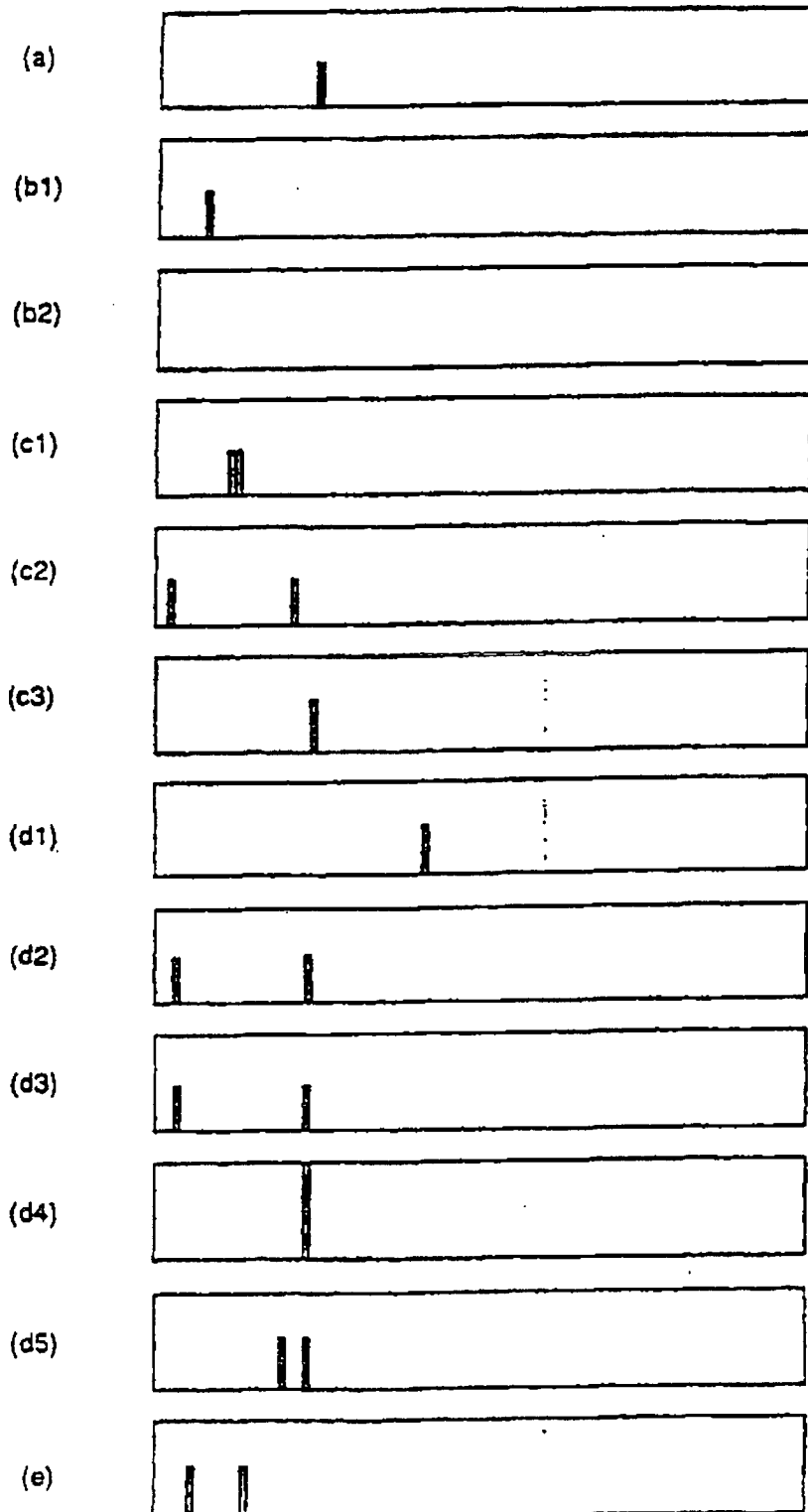
Figur 2b



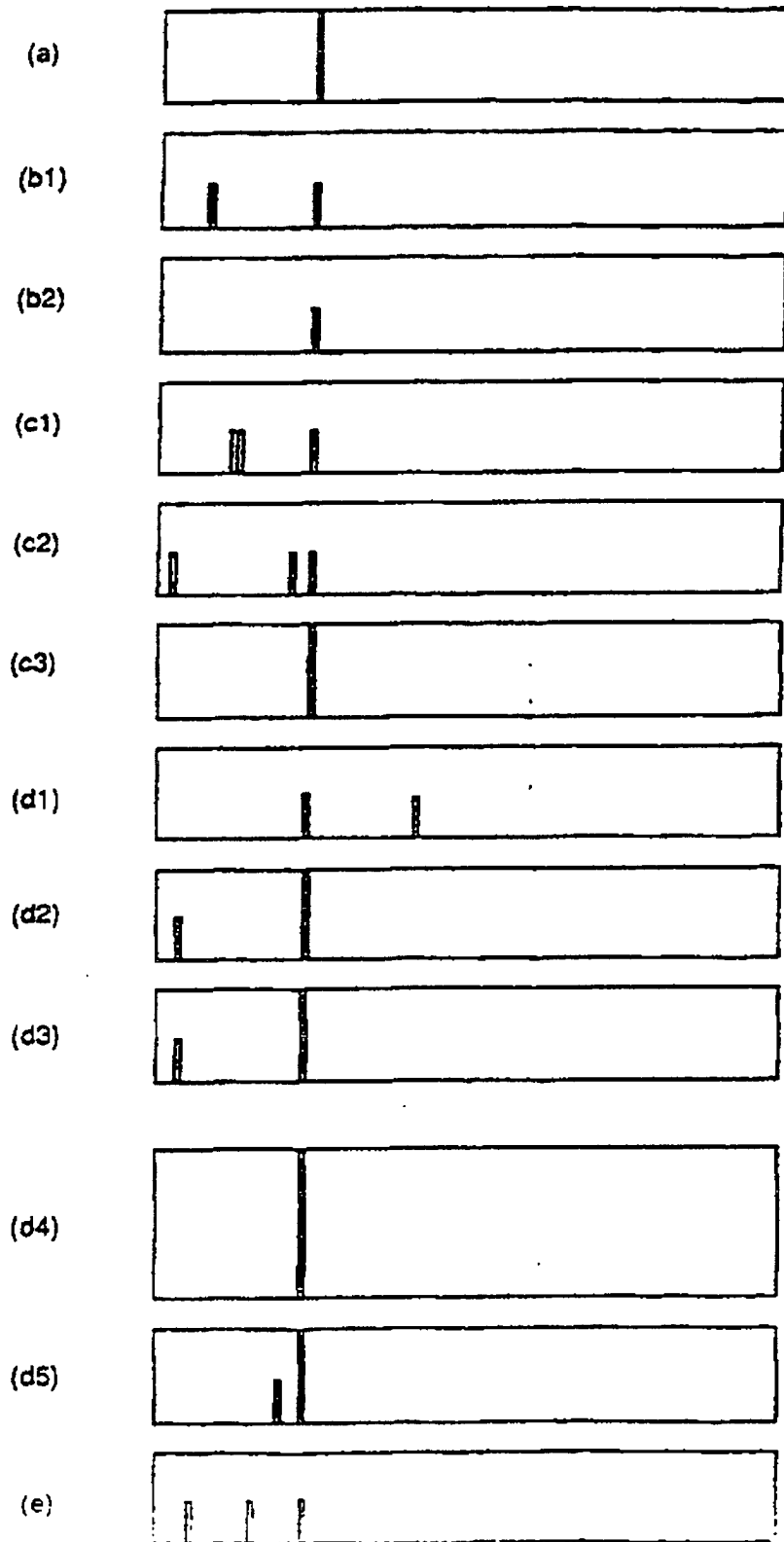
Figur 2c



Figur 3a

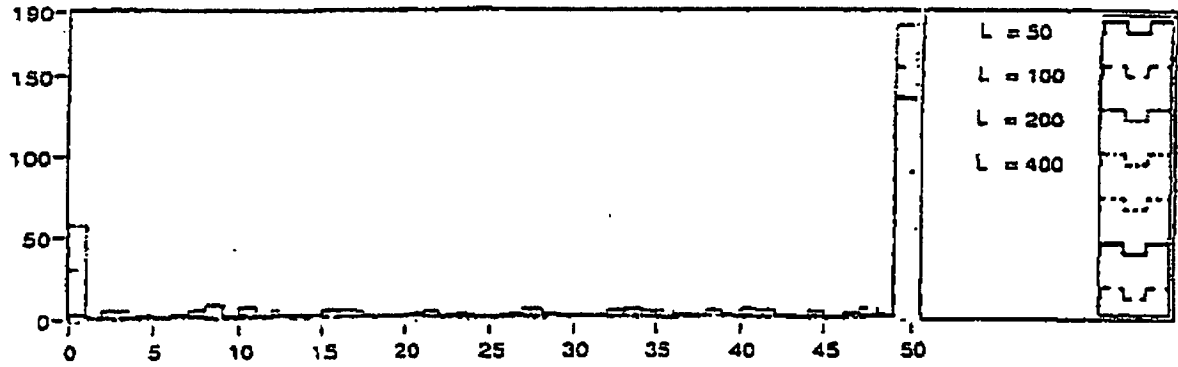


Figur 3b

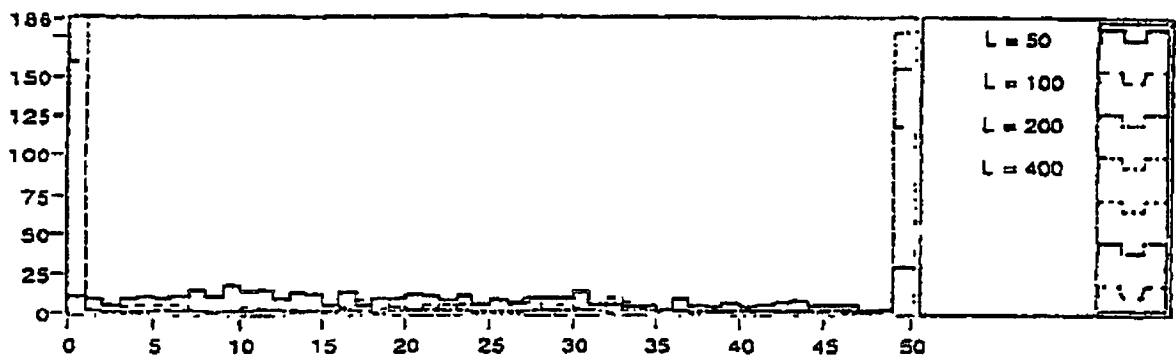


Figur 3c

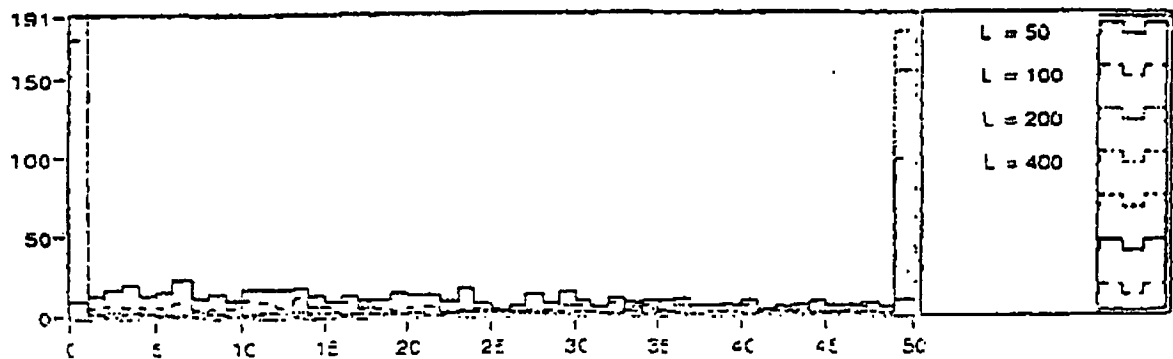
$N = 5$



$N = 10$

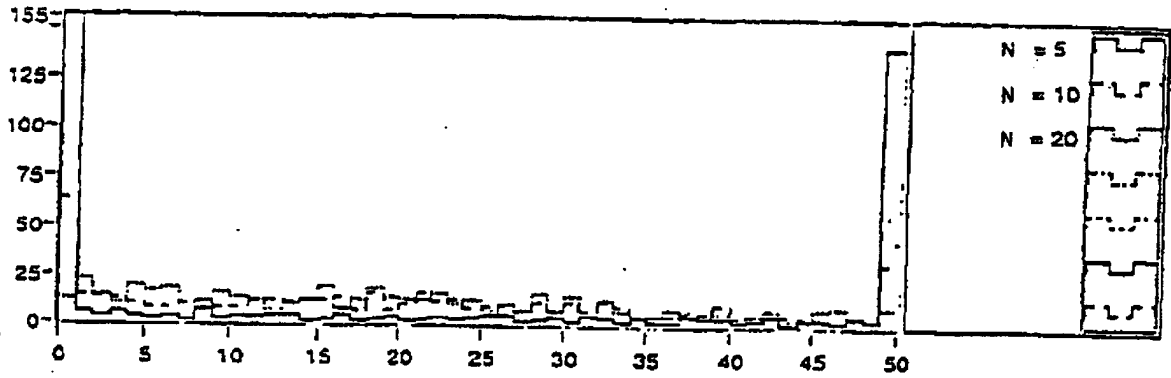


$N = 20$

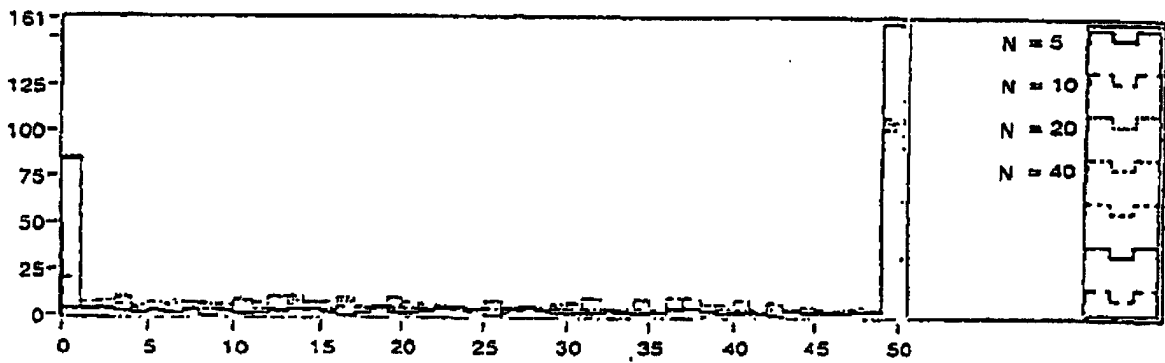


Figur 4

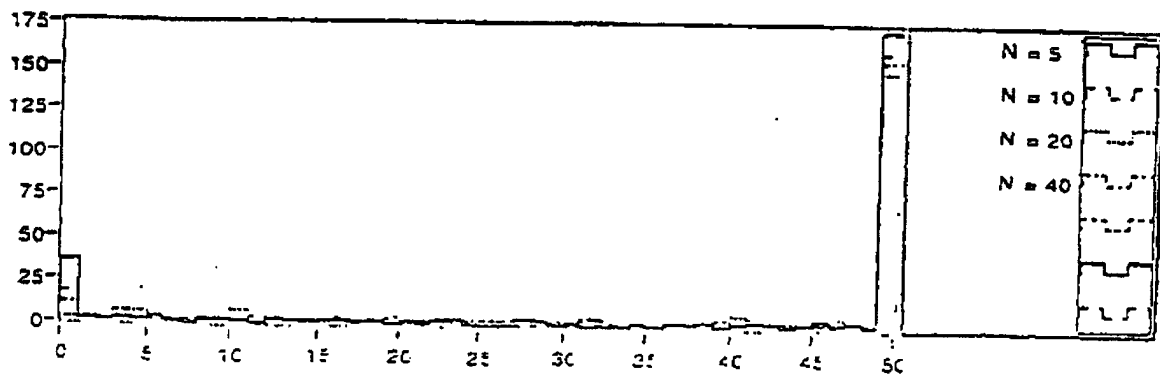
L = 50



L = 100

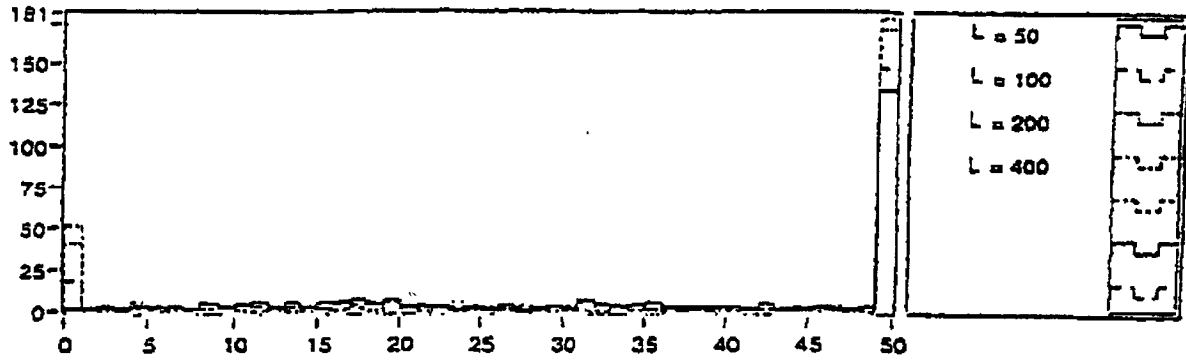


L = 200

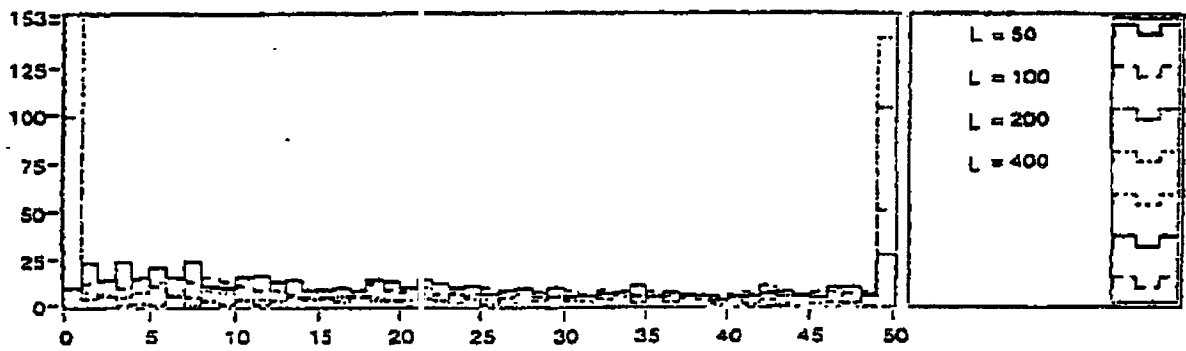


Figur 5

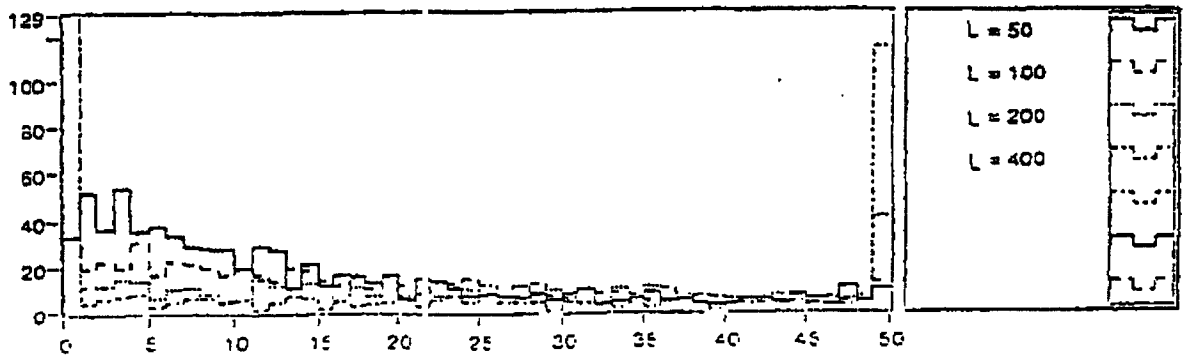
$N = 5$



$N = 10$

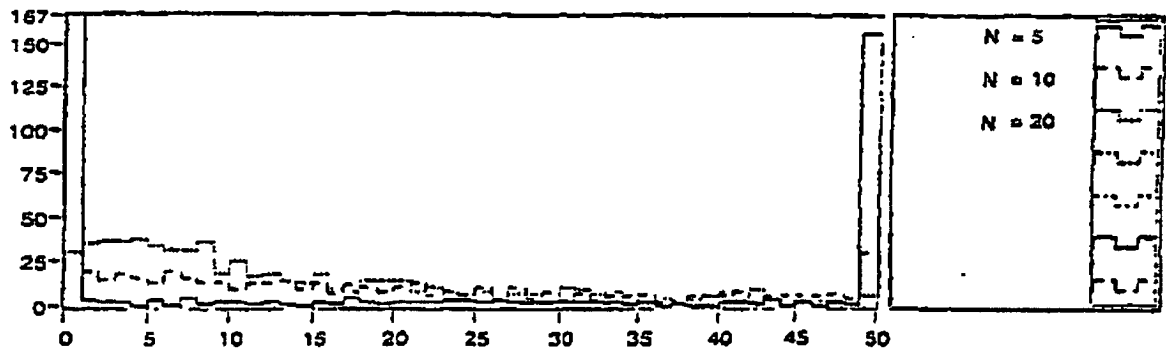


$N = 20$



Figur 6

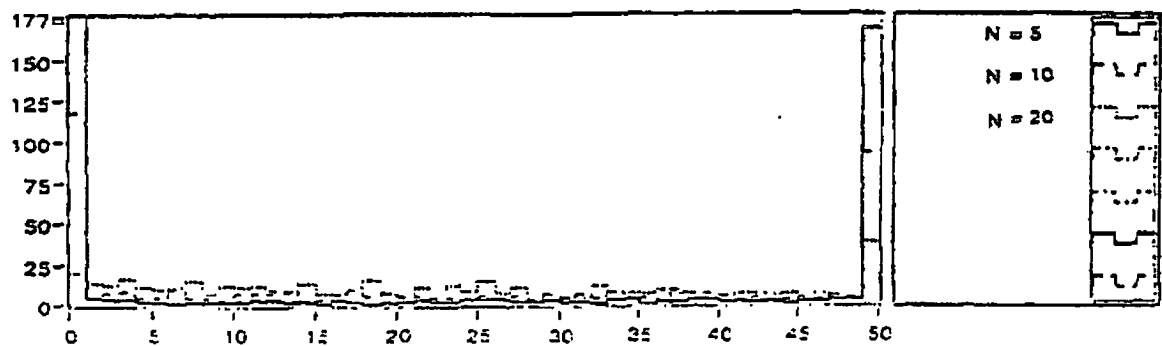
$L = 50$



$L = 100$

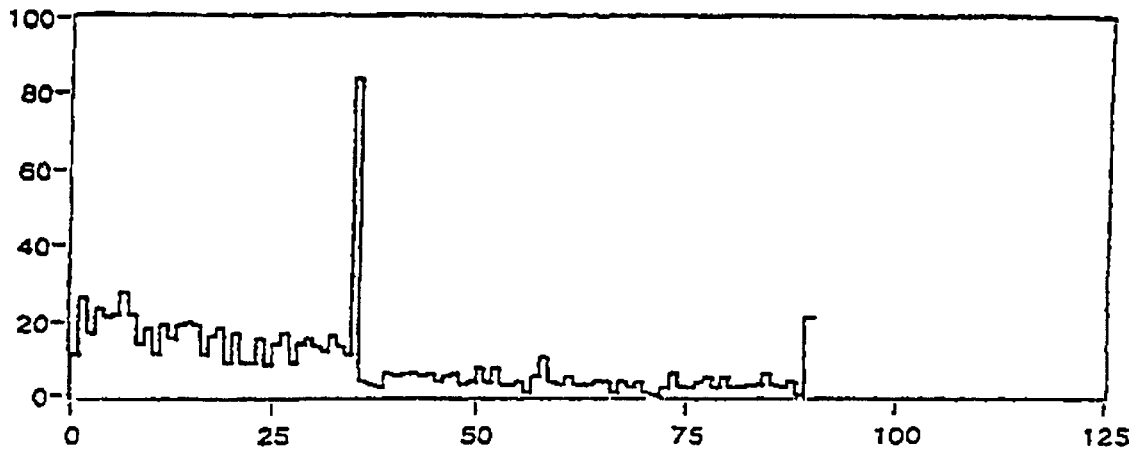


$L = 200$

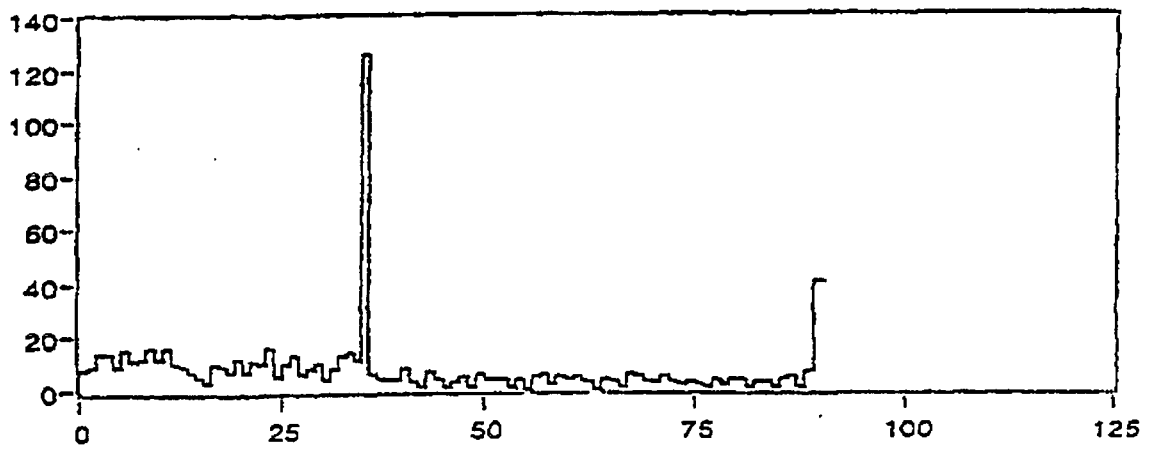


Figur 7

$L_1 = 35, L_2 = 90, L = 100$

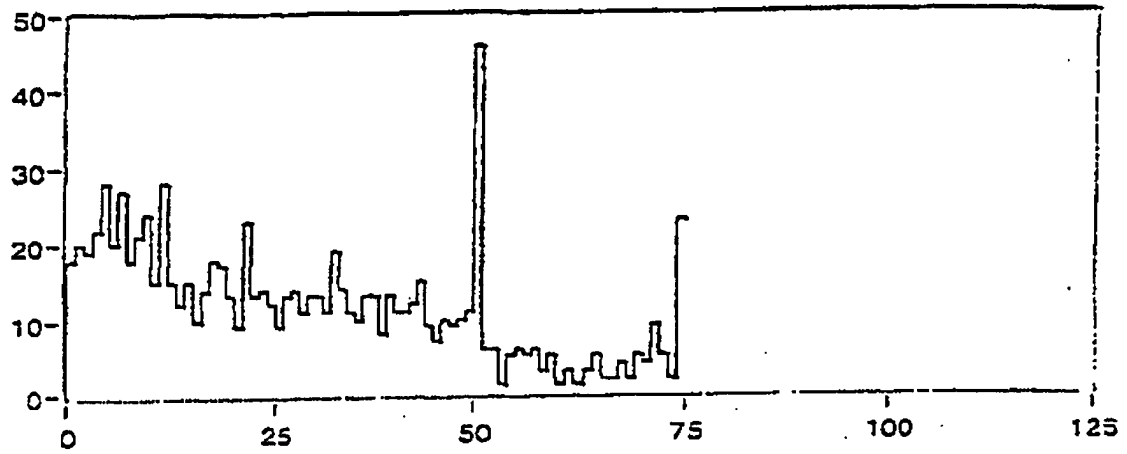


$L_1 = 35, L_2 = 90, L = 200$

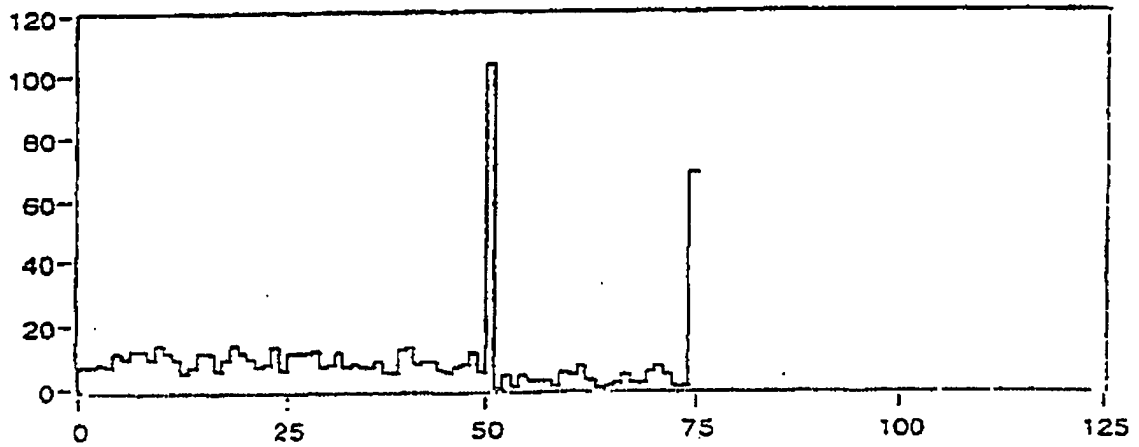


Figur 8a

$L_1 = 50, L_2 = 75, L = 100$

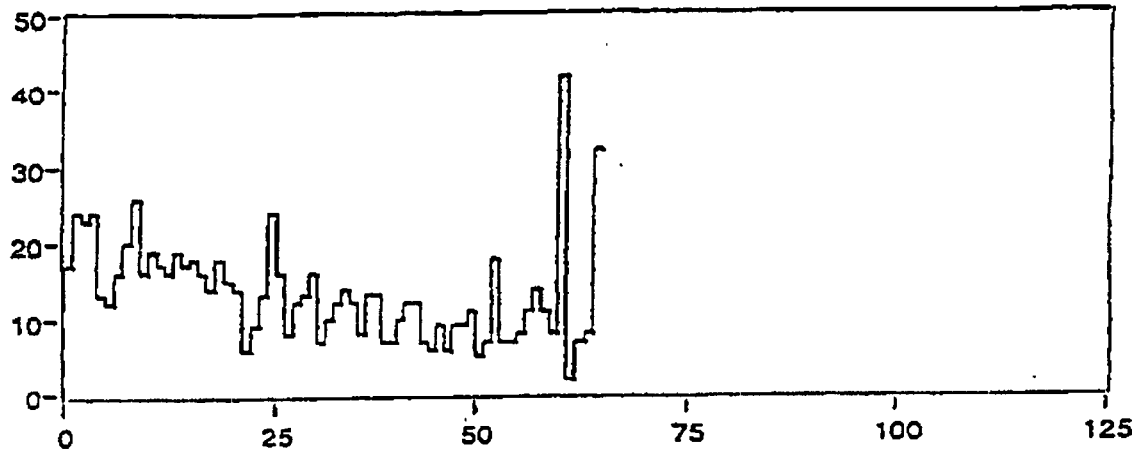


$L_1 = 50, L_2 = 75, L = 200$

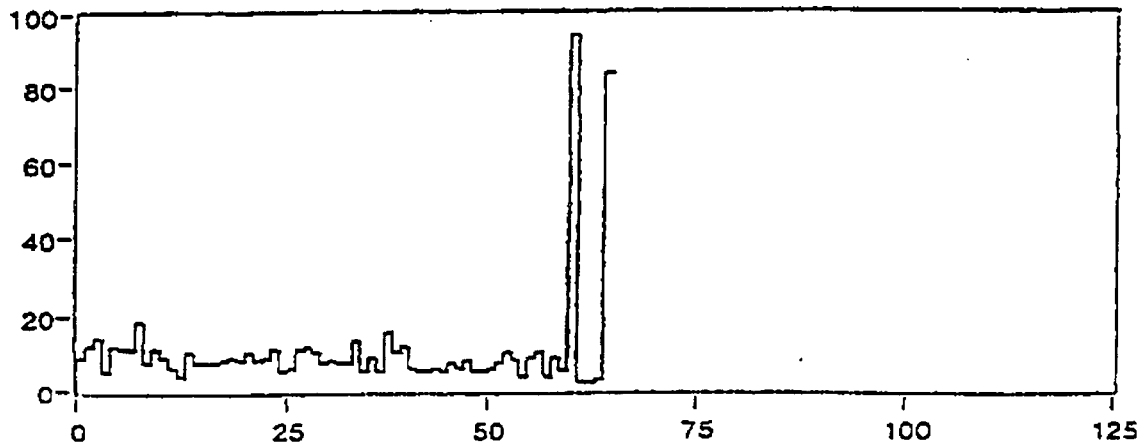


Figur 8b

$L_1 = 60, L_2 = 65, L = 100$

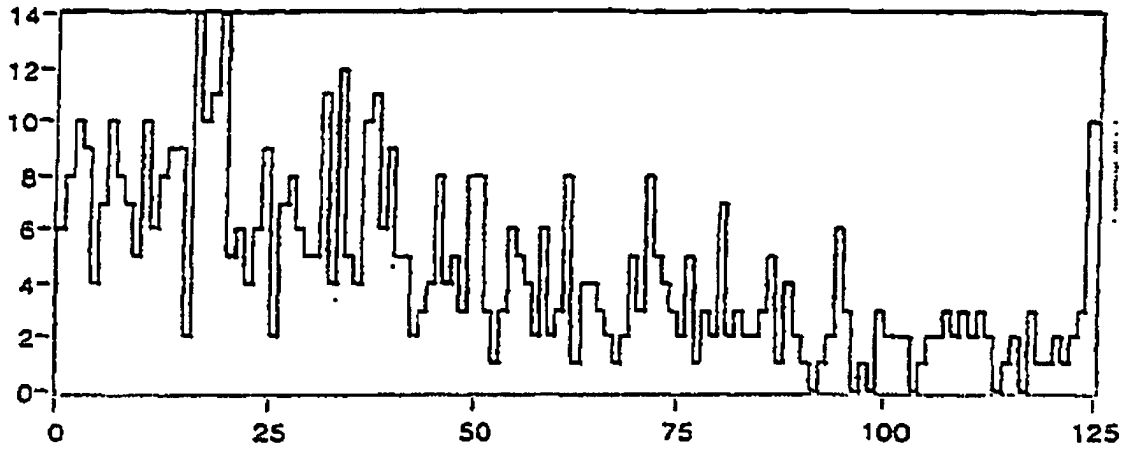


$L_1 = 60, L_2 = 65, L = 200$

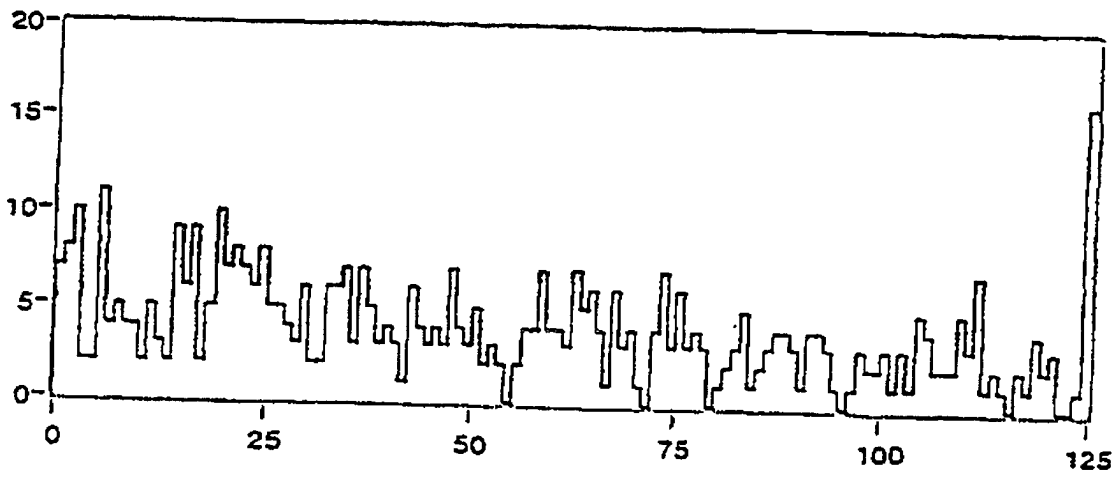


Figur 8c

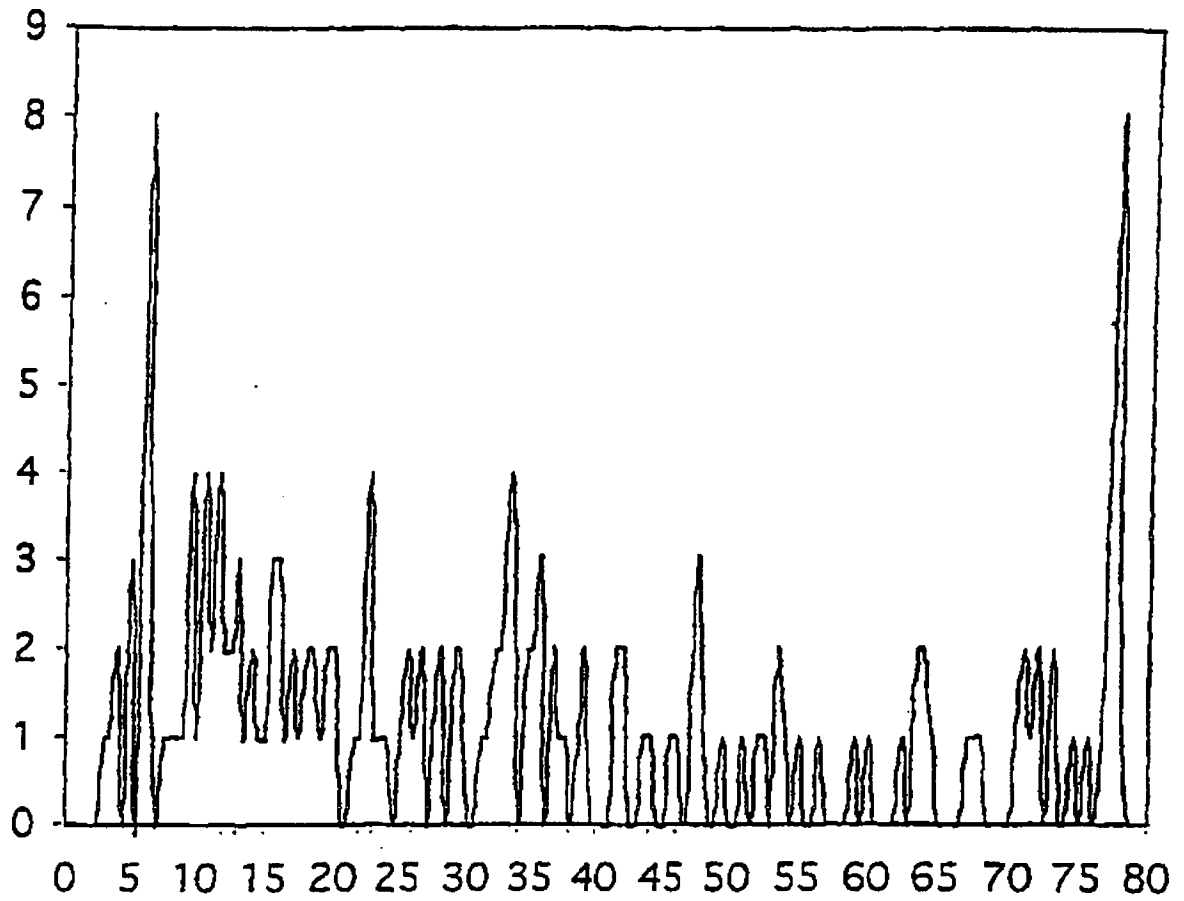
$L_1 = 125, L = 100$



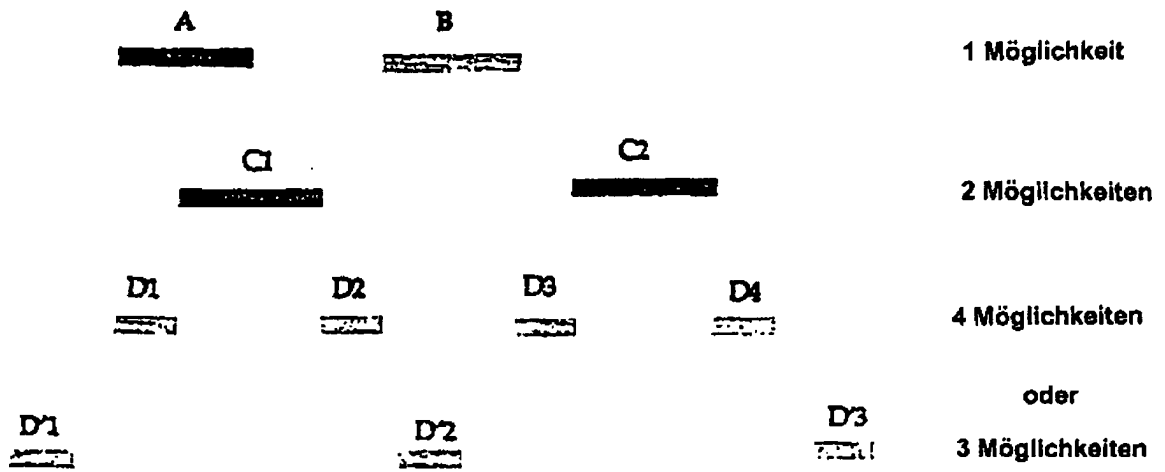
$L_1 = 125, L = 200$



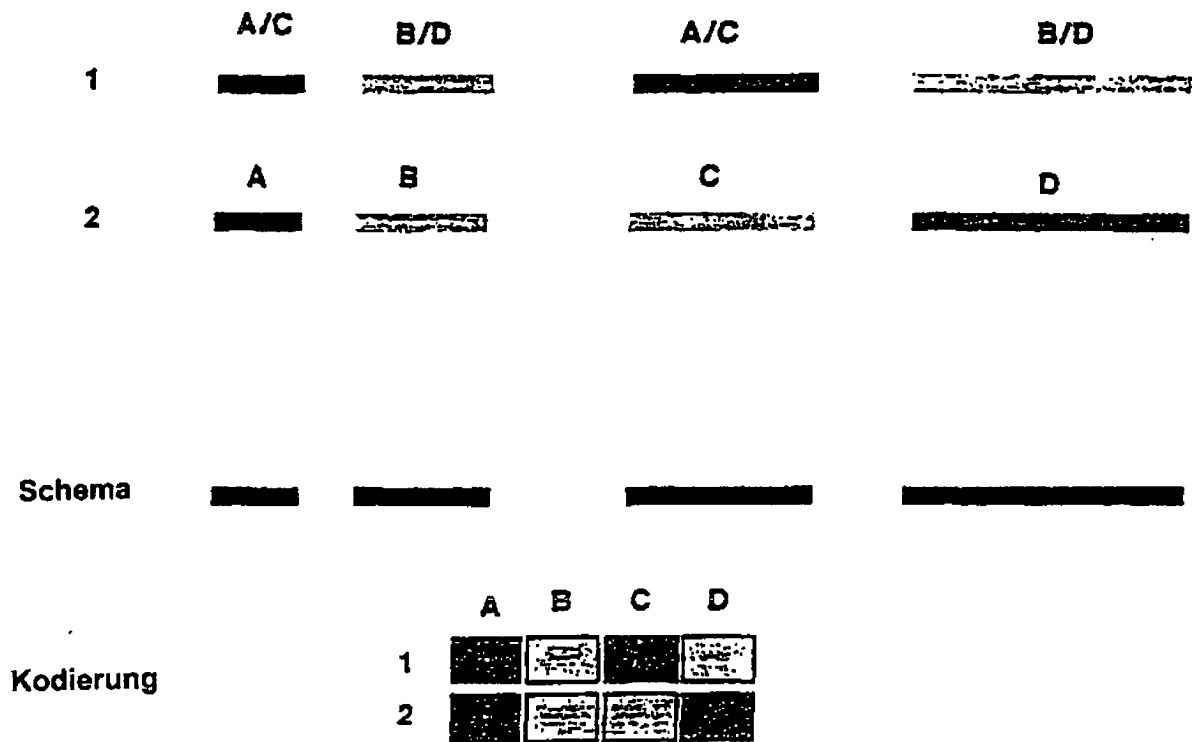
Figur 8d



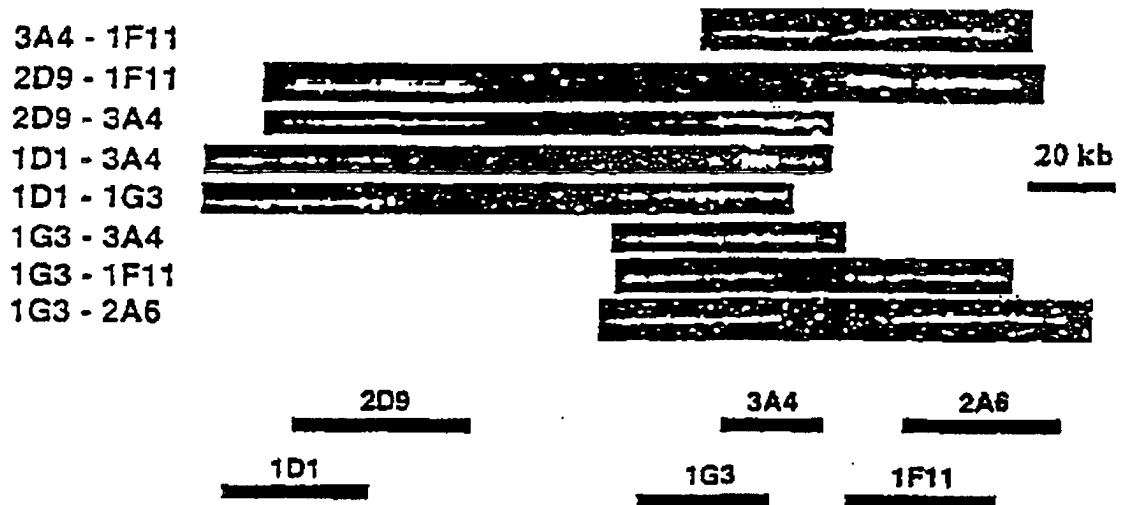
Figur 8e



Figur 9

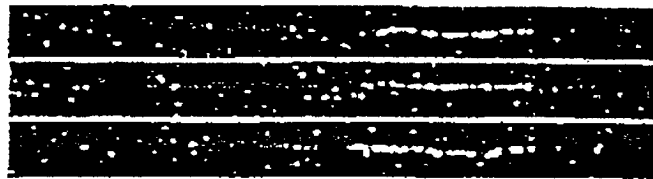


Figur 10



Figur 11

255A6 (SURF ctg) ■ & 220F3 (ctg 37) ■



20 kb

180F1 (ctg 37) ■ & 50D9 (DBH ctg) ■



20 kb

Figur 12