

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 643 712**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.01.2015 PCT/EP2015/050414**

87 Fecha y número de publicación internacional: **16.07.2015 WO15104411**

96 Fecha de presentación y número de la solicitud europea: **12.01.2015 E 15701486 (1)**

97 Fecha y número de publicación de la concesión europea: **30.08.2017 EP 3094745**

54 Título: **Procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos**

30 Prioridad:

13.01.2014 DE 102014200446

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

24.11.2017

73 Titular/es:

**TECHNISCHE UNIVERSITÄT DRESDEN (100.0%)
Mommсенstrasse 13
01069 Dresden, DE**

72 Inventor/es:

**ZHANG, YIXIN;
HERRMANN, JANA;
WIEDUWILD, ROBERT y
BERTHOLD, ANNETT**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 643 712 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos

- 5 La invención se refiere a un procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos con el uso de secuencias de ácido nucleico diana generadas sintéticamente.

10 Los ácidos nucleicos sirven en la naturaleza para la codificación de informaciones genéticas. Los procedimientos para el reconocimiento e interpretación de secuencias de ácido nucleico, a partir de las cuales se forman ácidos nucleicos, son por tanto de gran interés para muchos campos de investigación. Con los procedimientos según Maxam y Gilbert o según Sanger, pudieron asentarse los fundamentos pioneros para la secuenciación de ácidos nucleicos. También se han establecido ya procedimientos para la síntesis de secuencias nucleotídicas cortas (síntesis de oligonucleótidos) como, por ejemplo, el procedimiento de triéster fosfito, y se cuentan en el estado de la técnica. A causa de estas capacidades, se han desarrollado otros procedimientos que aprovechan las secuencias nucleotídicas, particularmente secuencias de ADN, como portadores de información. Una tecnología que se sirva de moléculas de ADN para el almacenamiento de informaciones se denomina código de barras del ADN. El fin es a este respecto sintetizar secuencias de ADN cortas, los denominados códigos de barras de ADN, para asignar entonces a secuencias nucleotídicas conocidas (habitualmente mayores) o sustancias o para acoplar con la secuencia nucleotídica o sustancia que representan. La identificación de las secuencias nucleotídicas o sustancias así preparadas es entonces posible de modo sencillo mediante el código de barras de ADN respectivo, cuyas secuencias cortas se secuencian en menos tiempo y/o se amplifican con procedimientos correspondientes (PCR) y de este modo pueden enriquecerse. A causa de la amplificabilidad de las secuencias nucleotídicas, los procedimientos basados en ácidos nucleicos se cuentan en el campo de la química analítica y bioquímica entre los procedimientos de detección más sensibles.

25 Se encuentra un campo de aplicación adicional en la investigación química, biológica y médica. Un objetivo básico consiste aquí en el descubrimiento de estructuras moleculares con afinidades de unión específicas por proteínas. Para ello, las colecciones de moléculas químicas codificadas por ADN sirven como una herramienta eficaz para la localización de ligandos para proteínas farmacéuticamente relevantes. Así, las moléculas codificadas por ADN pueden enriquecerse por ejemplo mediante una selección basada en la afinidad y a continuación descodificarse a causa de su codificación de ADN definida (código de barras de ADN). Habitualmente, se obtienen en tales experimentos de selección (cribados) mezclas de sustancias codificadas por ADN. Tales mezclas contienen habitualmente una pluralidad de sustancias codificadas por ADN.

35 A pesar de esto, por razones de costes en el análisis de mezclas de sustancias de los experimentos de selección, se ha renunciado ampliamente a etapas de aislamiento y purificación. Los datos recogidos a este respecto se basan convenientemente en la suposición de que las sustancias enriquecidas en códigos de barras de ADN se presentan en una de tales mezclas con una alta probabilidad y por consiguiente se secuencian también con una elevada probabilidad. Esta correlación no se aplica sin embargo necesariamente. Así, el resultado puede estar afectado por varios factores, como por ejemplo a causa de la transformación de diferentes plásmidos en bacterias (en la preparación para secuenciación de Sanger) o mediante procesos de reasociación y amplificación con microestructuras/nanoestructuras (procedimientos de secuenciación profunda). Esta circunstancia hace necesaria otra etapa de procedimiento prolongada en que debe comprobarse si en la sustancia presuntamente identificada se trata realmente de la sustancia enriquecida en la mezcla.

45 Respecto a la baja paralelización de los procedimientos de secuenciación habituales (Maxam y Gilbert o procedimientos de dideoxi según Sanger), es inevitable por tanto una costosa preparación de muestra, particularmente en mezclas de sustancias que no presentan un enriquecimiento significativo de un ácido nucleico o secuencia nucleotídica buscado. Además, los procedimientos de secuenciación de nueva generación, como por ejemplo la pirosecuenciación, requieren también un aislamiento y purificación de una mezcla de muestras antes de poder empezar la verdadera secuenciación.

50 Faircloth, B.C. y Glenn, T.C. ("Large sets of edit-metric sequence identification tags to facilitate large-scale multiplexing of reads from massively parallel sequencing", disponible en Nature Precedings, <http://hdl.handle.net/10101/npre.2011.5672.1>; S. 1-15, 2011), Krishnan, A.R. et al. ("Barcodes for DNA sequencing with guaranteed error correction capability", Electronic Letters, The Institution of Engineering and Technology, vol. 47, nº 4, S. 236-237, 2011), Bystrykh, L.V. et al. ("Generalized DNA barcode design based on hamming codes", Plos One, vol. 7, nº 1, e36852, 2013) divulgan respectivamente un procedimiento para la generación y deconvolución de secuencias nucleotídicas en el que, según un algoritmo predeterminado, se generan secuencias nucleotídicas diana distintas entre sí que se acoplan con una sustancia.

60 Por tanto, debe ser objetivo de la invención proponer un procedimiento con el que puedan identificarse secuencias nucleotídicas individuales en mezclas de sustancias que contienen ácidos nucleicos en un tiempo corto y económicamente.

65 Se consigue este objetivo mediante un procedimiento según la reivindicación 1. Las configuraciones y realizaciones

adicionales ventajosas del procedimiento según la invención pueden concretarse con los rasgos designados en las reivindicaciones dependientes.

5 Según la presente invención, se propone para conseguir el objetivo un procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos en que, en una primera etapa, se generan a partir de varios nucleótidos (A, C, G, T/U), según un algoritmo predeterminado, varias secuencias nucleotídicas diana (TNS) diferentes entre sí ($A_1-A_n, B_1-B_n, \dots, Z_n$) con posiciones de secuencia N_0-N_n . En una etapa adicional, al menos una de las TNS generadas ($A_1-A_n, B_1-B_n, \dots, Z_n$) se asigna respectivamente a al menos una sustancia o combinación de sustancias y se acopla químicamente con esta. Además, en el procedimiento según la invención se proporciona al menos una mezcla de sustancias para analizar con al menos dos TNS contenidas y/o sustancias acopladas a TNS diferentes entre sí que se secuencian según un procedimiento de secuenciación, en el que simultáneamente se recogen todas las TNS contenidas en la mezcla de sustancias ($A_1-A_n, B_1-B_n, \dots, Z_n$) u otros ácidos nucleicos o secuencias nucleotídicas en un espectro de secuencia común. Para facilitar la deconvolución y por tanto para la identificación de las TNS enriquecidas, los espectros de secuencia de una mezcla de sustancias deben restarse/sustraerse entre sí antes y después del experimento de selección.

20 Simultáneamente o a continuación de ello, se deconvolucionan las secuencias superpuestas en el espectro de secuencia mediante cribado de las posiciones de secuencia N_0-N_n según el algoritmo predeterminado, y según su asignación, se identifican como una sustancia o combinación de sustancias. A este respecto, puede plantearse por ejemplo así que las posiciones de secuencia N_0-N_n que presentan en el espectro de secuencia una intensidad de señal significativamente elevada de un nucleótido particular (A, C, G, T/U), se criban según el algoritmo predeterminado.

25 La intensidad de señal de un nucleótido (A, C, G, T/U) en una posición de secuencia N_0-N_n corresponde a la frecuencia del nucleótido (A, C, G, T/U) en la posición de secuencia N_0-N_n contemplada. En la señal, puede tratarse preferiblemente de una señal luminosa, como por ejemplo fluorescencia o quimioluminiscencia excitada externamente. En consecuencia, el límite de detección para la detección de un nucleótido (A, C, G, T/U) depende del ruido de fondo o de la sensibilidad del procedimiento usado y/o del detector.

30 Se evalúa como significativa una intensidad de señal de un nucleótido (A, C, G, T/U) cuando en una posición de secuencia N_0-N_n está elevada frente a al menos un nucleótido, preferiblemente frente a dos nucleótidos, con especial preferencia frente a tres nucleótidos al menos un 5 %, preferiblemente al menos un 30 %. A este respecto, pueden presentarse también en una posición de secuencia N_0-N_n intensidades de señal significativamente elevadas para dos o tres nucleótidos (A, C, G, T/U).

35 El espectro de secuencia registrado en la secuenciación debería reflejar al menos las TNS para identificar de posiciones de secuencia N_0-N_n . Es especialmente ventajoso cuando en las posiciones de secuencia respectivas en el espectro de secuencia puede representarse una frecuencia relativa de nucleótidos individuales (A, C, G, T/U). Tal distribución de frecuencia puede determinarse mediante comparación de las intensidades de señal de los nucleótidos individuales en las posiciones de secuencia respectivas. La determinación de la frecuencia puede realizarse mediante al menos una TNS patrón que se añade a la mezcla de sustancias para analizar a concentración conocida, antes de realizar la etapa de secuenciación.

45 Así, puede llevarse a cabo la deconvolución del espectro de secuencia adicionalmente o como alternativa también de modo que en el espectro de secuencia se sustraigan las intensidades de señal significativamente elevadas de nucleótidos individuales (A, C, G, T/U) por etapas, hasta alcanzar en cada posición de secuencia N_0-N_n como máximo un nucleótido (A, C, G, T/U) de intensidad de secuencia mínima (legible), y los espectros de sustracción obtenidos a este respecto, que presentan respectivamente al menos una secuencia o al menos segmentos de secuencia en las posiciones de secuencia N_0-N_n , se criban según el algoritmo predeterminado.

50 La ventaja esencial del procedimiento según la invención se basa en la aplicación del algoritmo predeterminado, con cuya base pueden generarse una pluralidad de TNS con alto grado de diferencia. El alto grado de diferencia influye favorablemente particularmente en la identificación de las TNS individuales, que puede realizarse de este modo con sensibilidad más elevada. Así, la identidad de una TNS o un candidato a TNS en consideración puede establecerse ya mediante pocas posiciones de secuencia de una serie de nucleótidos cribadas por el algoritmo predeterminable. De esta manera, pueden deconvolucionarse también espectros de secuencia con secuencias individuales o fragmentos de secuencia superpuestos repetidos, en que según el algoritmo predeterminado se criban preferiblemente las posiciones de secuencia N_0-N_n con diferencia de secuencia conocida. La deconvolución de una mezcla de sustancias que contiene ácidos nucleicos puede realizarse por tanto respecto a una TNS para identificar sin etapas de amplificación ni aislamiento.

60 Habitualmente, se realiza la descodificación después de experimentos de selección realizados mediante matrices de ADN y secuenciación de alto rendimiento/secuenciación profunda/secuenciación de la nueva generación. Esto es caro y complejo. Mediante el procedimiento según la invención, es posible deconvolucionar, después de realizar experimentos de selección, una mezcla que contiene ácidos nucleicos, como por ejemplo una mezcla de ADN, de modo rápido, económico y sencillo, por ejemplo mediante secuenciación de Sanger.

En la identificación de TNS enriquecidas (iguales), es suficiente con cribar únicamente las posiciones de secuencia del espectro de secuencia que presenten una intensidad de señal significativamente elevada para nucleótidos individuales (A, C, G, T/U). Las TNS clarificadas sirven entonces como prueba de la presencia de la sustancia
5 asignada respectivamente. Debe entenderse aquí por el término sustancia preferiblemente moléculas, componentes moleculares y particularmente sus grupos funcionales y/o estructurales. Puede tratarse también en el término sustancia, según el tipo de aplicación, de partículas de hollín, humo de tabaco, polución, humo de aceite, cenizas volátiles, cemento en suspensión, metal, óxido metálico, plástico, polen, bacterias o virus.

10 Se establece la ocupación nucleotídica de los nucleótidos (A, C, G, T/U) en las posiciones de secuencia N_0-N_n de una TNS para formar $(A_1-A_n, B_1-B_n, \dots, Z_n)$ mediante el algoritmo predeterminado. Para ello, se predetermina para cada posición de secuencia N_0-N_n una condición que está ligada a una ocupación nucleotídica de al menos una posición de secuencia adicional. Así, puede plantearse por ejemplo en la formación de la TNS según el algoritmo
15 que, para cada posición de secuencia $N_{0+1}-N_n$ de una TNS para formar, se predetermine una limitación referida a un nucleótido (A, C, G, T/U) de una posición de secuencia precedente respectivamente para al menos un nucleótido (A, C, G, T/U).

Para una identificación sencilla preferida de las TNS, es ventajoso cuando las TNS formadas se diferencian entre sí en al menos dos posiciones de secuencia y/o mediante al menos una serie nucleotídica consistente en al menos 5
20 posiciones de secuencia. Apropiadamente, las TNS que presentan una diferencia de secuencia de al menos un 75 %, preferiblemente más de un 80 %, con especial preferencia más de un 90 %, deberían asignarse respectivamente a sustancias que presenten la mayor diferencia estructural y/o funcional entre sí.

Preferiblemente, las TNS que representan respectivamente sustancias individuales deberían ser respectivamente de
25 la misma longitud, es decir presentar respectivamente el mismo número de posiciones de secuencia N_0-N_n . Así, puede simplificarse la identificación ya solo con limitar la deconvolución a una longitud de secuencia predeterminada. Esto posibilita también la comparación directa en las posiciones de secuencia respectivas N_0-N_n de TNS superpuestas. En este contexto, es además ventajoso cuando todas las TNS presentan un segmento de secuencia común a causa del cual pueden identificarse como tales. Este segmento de secuencia debería
30 concretarse preferiblemente en la región de inicio o final de una TNS.

Además, las TNS formadas pueden presentar al menos un segmento de secuencia que codifique un grupo de sustancias, tamaño de sustancia, coordenadas geográficas de una variedad de exposición o una fecha. Las propiedades de las sustancias pueden codificarse también por la longitud de la TNS, es decir, por el número de
35 posiciones de secuencia. Las propiedades de las sustancias pueden codificarse también en forma de distintos sitios de unión a cebador. Por ejemplo, una TNS puede codificar coordenadas geográficas. Con un segmento de secuencia adicional que funcione como sitio de unión a cebador durante la secuenciación, puede codificarse un grupo de sustancias. Por tanto, mediante el empleo de diferentes cebadores durante la reacción de secuenciación, puede establecerse la secuencia de la TNS respectiva y así las coordenadas geográficas. Por medio del cebador
40 utilizado en la secuenciación respectivo, es conocido qué grupo de sustratos contemplar/considerar en este caso. En la elaboración de la secuencia de sitios de unión a cebador, no se utiliza el algoritmo anteriormente descrito. En lugar de ello, esta secuencia debería diseñarse mediante parámetros que permitan una unión a cebador/secuenciación exitosas. Estos pueden ser, por ejemplo, el contenido de G/C, la longitud de cebador y la temperatura de fusión del cebador.

45 Las sustancias combinadas, y particularmente aquellas sustancias que se combinan entre sí como resultado de un experimento de selección o experimento de afinidad, pueden acoplarse químicamente con las correspondientes TNS combinadas $(A_1-A_n, B_1-B_n, \dots, Z_n)$. Adecuadamente, las TNS $(A_1-A_n, B_1-B_n, \dots, Z_n)$ y/o segmentos de secuencia de TNS $(A_1-A_n, B_1-B_n, \dots, Z_n)$ pueden combinarse entre sí.
50

A causa de la variedad de combinaciones de secuencia posibles, la formación de las TNS según el algoritmo predeterminado puede simularse *in silico*. A este respecto, pueden comprobarse las TNS formadas respecto a posibles colisiones con secuencias nucleotídicas ya conocidas.

La síntesis química de las TNS puede realizarse entonces preferiblemente según el procedimiento de triéster fosfito.
55

Ya que el número de combinaciones de secuencia posibles se guía por el número de posiciones de secuencia que están disponibles, deberían formarse TNS con suficiente longitud, es decir suficiente número de posiciones de secuencia N_0-N_n . Por consiguiente, deberían formarse o sintetizarse TNS con una longitud de más de 5 posiciones de secuencia.
60

En las TNS, puede tratarse de moléculas de ARN o ADN monocatenarias o bicatenarias, prefiriéndose moléculas de ADN bicatenarias. El acoplamiento químico de las TNS con las sustancias que representan puede realizarse preferiblemente mediante unión covalente.

65 En el transcurso del procedimiento según la invención, puede disponerse al menos una etapa de procedimiento para la selección de al menos una TNS $(A_1-A_n, B_1-B_n, \dots, Z_n)$ o al menos una sustancia acoplada a TNS. Así, puede

llevarse a cabo una primera etapa de selección, por ejemplo después de la formación/síntesis de las TNS, para retirar las TNS incorrectas. Puede realizarse una etapa de selección adicional después del acoplamiento de las TNS con sus sustancias asignadas, de modo que se establezcan las sustancias acopladas incorrectamente y por tanto pueda evaluarse la calidad de la colección. A este respecto, puede plantearse respectivamente así que los componentes para seleccionar (sustancia o TNS) que se encuentren en una fase móvil líquida se unan a una fase estacionaria en la que se inmovilizan los correspondientes copartícipes de unión o conjugados de los componentes para seleccionar. Así, puede proporcionarse una fase estacionaria respectivamente a una mezcla de sustancias para analizar de una fase móvil líquida y/o de una elución líquida.

Como alternativa o adicionalmente, puede establecerse la proporción de sustancias indeseadas o acopladas incorrectamente con TNS en una etapa de selección adicional. A este respecto, las sustancias indeseadas se llevan a una reacción de degradación mediante el empleo de un reactivo de terminación que contiene el mismo grupo reactivo que los componentes de la sustancia de la etapa de reacción precedente. Por tanto, el reactivo de terminación puede reaccionar con moléculas precursoras todavía no reaccionadas y marcar estas, de modo que el reactivo de terminación pueda designarse también como sustancia marcadora. Una sustancia marcadora puede presentar, por ejemplo para posibilitar el acoplamiento con una fase estacionaria, una secuencia de ADN o ADN, una molécula de biotina o estreptavidina/avidina y/azida/alquilo.

La mezcla de sustancias así preparada puede asociarse entonces en forma de una fase móvil con una fase estacionaria en la que se inmovilizan los correspondientes dominios de captura, conjugados, secuencias de ARN y/o ADN para acoplamiento de la sustancia marcadora, en la que pueden establecerse y/o cuantificarse las sustancias indeseadas unidas a la fase estacionaria. Como sistema marcador/de captura, se tienen en cuenta por ejemplo la reacción de biotina-estreptavidina/avidina, ADN/ADN, ARN/ARN o azida-alquilo de Huisgen Klick.

Las etapas de selección descritas anteriormente pueden realizarse particularmente a continuación de una primera secuenciación de una mezcla de sustancias para analizar. Para ello, pueden retirarse ya en una primera etapa de secuenciación las TNS o sustancias acopladas con TNS identificadas de la mezcla de sustancias para analizar. Además, puede conseguirse de este modo también un aislamiento de las TNS o sustancias acopladas con TNS. Con ello, pueden identificarse TNS adicionales mediante secuenciación de nuevo de la mezcla de sustancias restante, que se presentan en una menor cantidad. Pueden disponerse por tanto una o varias etapas de selección para la selección de al menos una TNS o sustancia acoplada con TNS para la provisión de la mezcla de sustancias para analizar.

Respecto al procedimiento de secuenciación para usar, para la secuenciación de una mezcla de sustancias para analizar, no se predeterminan limitaciones. Preferiblemente, la secuenciación puede llevarse a cabo según Sanger con el uso de didesoxinucleótidos marcados fluorescentemente, al menos una polimerasa y al menos un cebador que sea complementario de un segmento de secuencia de al menos una TNS.

A continuación, se ilustra detalladamente el procedimiento según la invención mediante ejemplos de realización y aplicación en relación con las Figuras.

A este respecto, se muestra:

Figura 1: un ejemplo de un algoritmo para la generación *in silico* de secuencias nucleotídicas diana (TNS) diferentes.

Figura 2: un ejemplo de una etapa de selección para la selección de TNS.

Figura 3: un ejemplo de una deconvolución de un espectro de secuencia según el algoritmo predeterminado.

Figura 4a/b: un diagrama esquemático de un ejemplo de aplicación del procedimiento según la invención.

En la Figura 1, se representa la generación de diferentes secuencias nucleotídicas diana (TNS) en un ejemplo. La generación de las TNS se realiza mediante el algoritmo X *in silico*. En primer lugar, se establece la longitud de las TNS para formar mediante el requisito de posiciones de secuencia. En el presente caso, la longitud de secuencia asciende a 14 posiciones N_0 - N_{13} , en el que para cada posición de secuencia N_0 - N_{13} se predetermina una operación z, a, d o e. En la posición de secuencia N_0 , la operación z no predetermina ninguna limitación para un nucleótido A, C, G, T. Para cada posición de secuencia adicional N_1 - N_{13} , se predetermina respectivamente según las operaciones a, d o e una limitación referida a la posición de secuencia precedente de un nucleótido A, C, G, T respectivamente para dos nucleótidos A, C, G o T. El algoritmo X presenta la forma z-a-a-d-d-e-a-a-d-d-e-a-d-e. La elección de la TNS así formada puede extraerse de la tabla de códigos adjunta en el anexo (véanse las págs. 15 a 18 de la descripción) Como puede deducirse de la Tabla con la referencia 10 en la Figura 1, un par cualquiera de TNS seleccionadas (generadas) presenta al menos seis diferencias de secuencia.

No representado, existe la posibilidad de una ampliación/realización adicional de las TNS así formadas con posiciones de secuencia adicionales mediante las que pueden depositarse (codificarse) informaciones adicionales. Así, pueden disponerse posiciones de secuencia o segmentos de secuencia adicionales que codifican

respectivamente propiedades de las sustancias para acoplar, como por ejemplo un grupo de sustancias, tamaño de sustancia, coordenadas geográficas de una variedad de exposición o una fecha. Además, pueden disponerse también posiciones de secuencia complementarias mediante las cuales pueden identificarse las TNS como tales. A este respecto, las TNS de igual género deberían presentar iguales longitudes de secuencia.

5 A continuación de ello, se lleva a cabo una síntesis *in vitro* de las TNS generadas *in silico* mediante el procedimiento de triéster fosfito.

10 El acoplamiento químico de las TNS sintetizadas con la sustancia que representan puede realizarse mediante la formación de un enlace amida con la ayuda de reactivos de acoplamiento peptídico.

15 La Fig. 2 muestra un ejemplo de una etapa de selección en el que en una primera secuenciación se pueden identificar las TNS como enriquecidas y retirarse de una mezcla. De este modo, se posibilita poder identificar TNS adicionales mediante secuenciación que, a causa de su menor número en la mezcla de sustancias (y por tanto menor intensidad de señal asociada) en la primera secuenciación, se ocultan por la intensidad de señal de las TNS de aparición frecuente. De este modo, pueden identificarse no solo las TNS más enriquecidas sino también otras enriquecidas en menor medida.

20 En el presente ejemplo, se representa en el lado izquierdo una mezcla de TNS con 11 posiciones, en el que cada posición representa una TNS combinada. Las mayúsculas representan a este respecto respectivamente una TNS individual. Se caracteriza con la referencia 20 una fase estacionaria con la que se inmovilizan las secuencias nucleotídicas A1', A3', A5', B3', B4' y B9' que son complementarias frente a las TNS A1, A3, A5, B3, B4 y B9 y por tanto posibilitan la conexión de las TNS citadas. Si se asocia la mezcla de TNS que contiene las posiciones 1 a 11 como fase móvil con la fase estacionaria 20, se unen las posiciones 1 a 9 con la fase estacionaria 20 y se retiran de la fase móvil. La mezcla de TNS procedente de este experimento de selección (lado derecho) contiene entonces solo las TNS de posiciones 10 y 11.

30 La Figura 3 muestra un esquema de flujo de un ejemplo de deconvolución de un espectro de secuencia 31, que se extrae de una secuenciación de una mezcla de sustancias que contienen ácidos nucleicos. En el presente ejemplo, se llevó a cabo la secuenciación según Sanger con el uso de didesoxinucleótidos marcados fluorescentemente. A causa de la intensidad de señal de los didesoxinucleótidos, podía determinarse en las posiciones de secuencia respectivas N_0-N_{13} en el espectro de secuencia 31, la frecuencia relativa de los nucleótidos A, G, C, T, que se representa en forma de barras. A causa de la distribución de frecuencia de los nucleótidos respectivos en las posiciones de secuencia N_0-N_{13} , pueden deducirse a partir del espectro de secuencia 31 distintos candidatos a TNS que se tienen en cuenta para una identificación. En el presente ejemplo, se establecieron los candidatos a TNS 321 y 322. Estos se criban mediante el algoritmo X predeterminado con la serie de operaciones z-a-a-d-d-e-a-a-d-d-e-a-d-e y se deconvolucionan. A causa de la deconvolución, pueden identificarse dos TNS 331 y 332, en los que la TNS 331 representa la sustancia i y la TNS 332 la sustancia ii.

40 La Fig. 4a muestra un diagrama esquemático de sustancias acopladas a una TNS de ejemplo de aplicación. Según el ejemplo de realización, debería entenderse la propagación de partículas acopladas con TNS o marcadas con TNS en el ambiente. Para ello, se exponen las partículas acopladas con TNS-A1B3 en la posición X, las partículas con acoplamiento con TNS A3B4 en la posición Y y las partículas con acoplamiento con TNS A8B7 en la posición Z. En las partículas, se trata por ejemplo de partículas nocivas con un tamaño en el intervalo de 10 nm a 100 μ m. Las respectivas posiciones X, Y o Z en las que se exponen las partículas están codificadas en un segmento de secuencia separado de la TNS respectiva. Además, las TNS pueden presentar también un segmento de secuencia separado que codifica la fecha de exposición (en la posición X, Y, Z respectiva) de las partículas. Puede realizarse una extracción de muestras por ejemplo en una posición L. Esto puede concretarse por ejemplo con un correspondiente filtro de aire. Las partículas recogidas con el filtro de aire se dispersan entonces en un líquido. La mezcla de sustancias que contienen ácidos nucleicos así proporcionada puede someterse a continuación a una secuenciación. A este respecto, puede ser ventajoso someter la mezcla de sustancias que contienen ácidos nucleicos para analizar antes de la secuenciación a una amplificación mediante PCR, en la que se utilizan los correspondientes cebadores de TNS A1B3, TNS A3B4 y TNS A8B7. De esta manera, puede establecerse la presencia de partículas marcadas con TNS A1B3, TNS A3B4 y TNS A8B7 en la posición L, con lo que puede entenderse una migración de las partículas de las posiciones X, Y y Z.

50 La Figura 4b muestra un ejemplo de un esquema según el cual pueden exponerse y atraparse las partículas marcadas con TNS. Para ello, se representa una superficie A formada por 16 cuadrados pequeños 1.1-1.16. La longitud del borde de los cuadrados pequeños asciende, p.ej., respectivamente a 40 km, de modo que una superficie A presenta en este caso una extensión de 1600 km². Los puntos representados con forma redonda en las esquinas de los cuadrados pequeños 1.1-1.16 caracterizan respectivamente una ubicación en la que se exponen partículas marcadas con TNS, en la que preferiblemente se dispone respectivamente centrada en los cuadrados pequeños una superficie de al menos 16 m² para atrapar o detectar las partículas marcadas con TNS.

65 Con las partículas marcadas con TNS recogidas, puede conseguirse una asignación local de la posición respectiva en la que se localizan y analizan una o varias partículas marcadas con TNS. Esto es particularmente ventajoso

cuando debe recogerse una distribución local determinada de partículas marcadas con TNS que se desplazan por influencias externas de una posición a otras posiciones.

El ejemplo de aplicación descrito es también adaptable a sistemas acuáticos.

5 Se encuentra un campo de aplicación adicional en la investigación química, biológica y médica. Así, pueden utilizarse por ejemplo TNS para identificar estructuras moleculares con afinidades de unión específica por proteínas. Además, pueden utilizarse también colecciones de moléculas químicas codificadas por TNS o ADN como herramienta eficaz para la localización de ligandos para proteínas farmacéuticamente relevantes. Así, pueden enriquecerse moléculas codificadas por TNS o ADN por ejemplo mediante una selección basada en afinidad y a 10 continuación decodificarse a causa de su codificación por TNS o ADN definida. Las mezclas de sustancias codificadas por TNS o ADN obtenidas en tales experimentos de selección pueden deconvolucionarse entonces según el procedimiento según la invención de manera sencilla, sin requerir un aislamiento/purificación o 15 amplificación de la mezcla.

Anexo

<u>Tabla de códigos</u>		
Número (=SEQ ID NO)	Código de TNS	Nº de ronda
1	GCGATGAGACATGT	0
2	ATCATATACGTATA	1
3	TAGACATCATAGAG	2
4	TATGTGCTCGCGAG	3
5	AGATGCTATGTCAC	4
6	TCTCGTAGTCTCGT	5
7	TCGATGATCACTCT	6
8	GCTCAGCTGTGCAG	7
9	CTCGAGATCGCTGC	8
10	CGCACTAGATGCGT	9
11	GCTCGCGCGAGCAC	10
12	AGCGTGAGTCTCAG	11
13	CTCGACTATCAGAC	12
14	CGATGCTCATAGTA	14
15	ATCGACGCATGCAG	15
16	GAGTGCGATCAGAG	17
17	CTACACTCACTACA	18
18	TATCGTGGATGATA	20
19	TCTGTATATCTCAC	21
20	GATCACTCGTATCA	23
21	AGCATAGCGACGTA	24
22	TCTGAGCGATAGTA	25
23	ATATCTCTGACGTG	27
24	GAGTCTAGACTCAG	29
25	GAGTGTAGTGTACA	30
26	GCTGTGCTGAGATA	31
27	CTACGTATGTATCT	32
28	TATCGCGACGTATA	34
29	GAGACTCGTGCGTG	36
30	TATCGCTCACAGAC	37

ES 2 643 712 T3

Número (=SEQ ID NO)	Código de TNS	Nº de ronda
31	TATCGCTACAGCGT	40
32	CTCACTCTCAGCAG	41
33	GATCAGCTCACTGT	47
34	AGCGTGCTGTATGT	50
35	CGACGCGCGACGAG	52
36	TCGATAGACAGATG	53
37	GCGATGCTGTATCA	54
38	CTACGCGATGCTGC	55
39	CGCACAGCACAGTG	56
40	GCGTCAGATGCTCA	57
41	TAGTCTCGATGCGC	59
42	ATCACAGCATGACA	60
43	CTATGCTACACGAC	68
44	CGCACAGATGTCGT	72
45	AGACAGATGAGACT	74
46	ATATCATCGTATGT	79
47	GATGTAGCACTACT	84
48	TATGTATCGACTCT	88
49	CGACAGAGACAGTG	92
50	CTCGTAGATCATGT	99
51	CGACGTCTCGTCGT	100
52	TAGTGCGACGCTCT	105
53	GAGACAGACACTGT	108
54	GCGACTCGATGACA	109
55	AGACACGCGTGATA	114
56	ATCATATCACTCAG	118
57	GATGTATCATATGC	122
58	GCTCGTCGTCAGTA	124
59	CGATGTATCACGTA	139
60	CTATGCTCGTGA	140
61	TATGTGAGACTATA	147
62	CTCGTGAGTCAGTA	148
63	TCTGACGCGAGACT	155
64	TAGTCTATGACTGC	161
65	CGCGACTATGCTGT	163
66	GATCGTATGTGCGC	167
67	AGATCATAACAGACT	168
68	TATGAGCGTGCTGC	170
69	TCGATATATGCTGC	177
70	AGCACTCTCGTATG	185
71	CTATCATATGCGTA	187

ES 2 643 712 T3

Número (=SEQ ID NO)	Código de TNS	Nº de ronda
72	TCGTGTCGTGCTGT	197
73	GCTGTAGCGTAGTG	198
74	TCTGTGCGTGTACT	202
75	GAGATATCGTGATA	209
76	TCGACTAGACAGTA	235
77	GCGACATATCAGTG	239
78	CGCACTATGTAGTA	245
79	AGCGACGACAGCGT	261
80	CTCATATCGTGCGC	263
81	GAGACTATGAGATG	270
82	GCTCGCTACGTCAG	295
83	CGCACTCGTGCTCA	299
84	GCGTGCTACAGACA	311
85	AGCATGCTCACGAC	322
86	TCGTCAGACACGAC	325
87	AGCGTAGCATAGAC	330
88	TATCAGAGATGCGT	342
89	TCGTCTATCGTCGT	344
90	CTCGACGACACGTA	360
91	AGACACGACGCTCA	362
92	CTATCTCGTCTCAC	368
93	TCTCACTATCATGC	380
94	GCTGAGATCACGAC	381
95	ATATGCGATCATCT	386
96	GCGTGCTCATATCT	401
97	TCTCGTATCGCGTG	405
98	AGATCTCTGTGACA	414
99	GAGATATACGCGAG	424
100	ATCGTGAGATGATG	437
101	TCTCACGCACTATG	440
102	GCGATGCGTGTCAC	445
103	TCGTGTCGATAGAG	458
104	GATCACTATGCGTA	484
105	CTACGTAGTCTATA	535
106	GAGTGCGCACTCGT	541
107	TCGTGCTATGTATG	542
108	AGATGTAGACTACT	581
109	ATACAGAGATATGC	594
110	CTACAGCTCGTATA	596
111	CGACGTCTGAGATG	608
112	CTCGAGCGATGCGC	625

ES 2 643 712 T3

Número (=SEQ ID NO)	Código de TNS	Nº de ronda
113	TAGATAGATCATCA	664
114	AGCATGCGACTACA	669
115	CGCGAGCTGACTCA	680
116	GATGTAGACGTTCGC	756
117	CGCATGATCAGATG	779
118	GCGTCATCGAGCGC	790
119	CTACAGCGTCATCT	817
120	ATACGCGCACTCAC	893
121	CTATGTATGAGCAC	899
122	CTCGTGATGACGAG	951
123	CGCGACGCACATGC	997
124	CTATCAGACGTCAG	1049
125	ATACAGAGTGCGAG	1112
126	ATACGTATCACTGC	1251
127	CGACAGCTCAGCAC	1311
128	CTATCAGCGAGATG	1355
129	ATCACTCGATAGTG	1376
130	CGCATATCATATCT	1399
131	CGCGAGAGTGACA	1525
132	GAGATGCTGAGCGC	1568
133	AGCATGAGTGCTGT	1589
134	CGCGTAGCGAGCAG	1778
135	AGATCATCATGCAC	1890
136	TCTGTATACACGTA	1909
137	TCTGTATCATGACA	2169
138	TCGTCTCTCAGATA	2196
139	ATACACGATGTATG	2609
140	ATACGCTCGACTCA	2833
141	TCTGACGATGCGAG	2857
142	GATGAGCGTCAGAG	2910
143	GAGTCTCGTCATCT	3395
144	ATACGTAGATGCAG	3415
145	GATCACGATCTCAC	3428
146	ATCACTAGTCATGC	3651
147	TAGTGCGCGTGATG	3680
148	GCTCAGATCGTACT	4170
149	GCTGACTACACTCT	4243
150	AGCGAGAGACATCT	4391
151	TCTCGTCTGACTCT	4568
152	GATGACGCACAGTA	5440
153	ATCACAGATGCGAC	5554

ES 2 643 712 T3

Número (=SEQ ID NO)	Código de TNS	Nº de ronda
154	GCGACTCTCGCTGC	5938
155	CGACAGCGATGACA	7003
156	AGATCAGCACATCA	7229
157	GCTGTGAGATGCAC	7364
158	GCGATAGCATGCAG	8343
159	GATCGTAGTGCGAC	8520
160	CTCACATACGCTCT	8522
161	CGATGTCGATATGT	8744
162	ATCACTATGACTCT	8809
163	TCGATAGCGTATGT	8963
164	ATACACTCGTAGAG	9308
165	GCGTGTATGTGACT	9847
166	TATCAGATGACGTA	10090
167	GAGACTCTGTAGAC	10629
168	GCGTGCGCGACGTA	10690
169	TAGTCAGATCTCGC	11570
170	AGACAGCGACTCGT	12132
171	CGATCAGATCAGAC	12906
172	GATGACGACAGATG	13442
173	AGCGAGATGTGCAC	13445
174	GCGTCAGCACTATA	13898
175	AGCGTATACACTGC	14072
176	CTCATGCTGTGACT	15131
177	GAGATGATCGTATA	15693
178	TATGACTATCTACT	16035
179	TAGATGATGTAGTG	17311
180	TCGACATCACTCGT	19265
181	GCTGACTCACTCGC	19811
182	GAGTGTCTCGTCAC	19988
183	CGACACTCGAGCGT	21276
184	TATCACGCGACTGC	21882
185	AGATGTATGTAGAG	21955
186	AGACGTCGACAGAC	22214
187	ATCGAGCTCGTCAG	22328
188	TCTGAGATGTATGC	25668
189	AGCGACTCGACGAC	25671
190	GAGACATATGTA CT	31084
191	GCTGTGAGTGCGTG	34906
192	CTCATGCGTCTCGT	34923
193	TATGACTCGTGACAC	42593
194	AGACGCTCACATGT	43050

ES 2 643 712 T3

Número (=SEQ ID NO)	Código de TNS	Nº de ronda
195	CGCGTAGATGCGTG	48624
196	CTACACGACAGACT	50109
197	CGATGTCGTGCGAG	66603
198	TAGATGCGACAGAC	93912

Listado de secuencias

- 5 <110> Universidad Técnica de Dresde
- <120> Procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos
- <130> 1
- 10 <150> DE 10 2014 200 446.2
- <151> 13-01-2014
- <160> 198
- 15 <170> BiSSAP 1.3
- <210> 1
- <211> 14
- <212> ADN
- 20 <213> -
- <400> 1
- gcgatgagac atgt 14
- 25
- <210> 2
- <211> 14
- <212> ADN
- <213> -
- 30
- <400> 2
- atcatatacg tata 14
- 35
- <210> 3
- <211> 14
- <212> ADN
- <213> -
- 40
- <400> 3
- tagacatcat agag 14
- 45
- <210> 4
- <211> 14
- <212> ADN
- <213> -
- 50
- <400> 4
- tatgtgctcg cgag 14
- 55
- <210> 5
- <211> 14
- <212> ADN
- <213> -
- <400> 5

ES 2 643 712 T3

agatgctatg tcac 14

5 <210> 6
<211> 14
<212> ADN
<213> -

10 <400> 6
tctcgtatgc tcgt 14

15 <210> 7
<211> 14
<212> ADN
<213> -

20 <400> 7
tcgatgatca ctct 14

25 <210> 8
<211> 14
<212> ADN
<213> -

30 <400> 8
gctcagctgt gcag 14

35 <210> 9
<211> 14
<212> ADN
<213> -

40 <400> 9
ctcgagatcg ctgc 14

45 <210> 10
<211> 14
<212> ADN
<213> -

50 <400> 10
cgcactagat gcgt 14

55 <210> 11
<211> 14
<212> ADN
<213> -

60 <400> 11
gctcgcgcga gcac 14

65 <210> 12
<211> 14
<212> ADN
<213> -

70 <400> 12
agcgtgagtc tcag 14

ES 2 643 712 T3

<210> 13
<211> 14
<212> ADN
<213> -
5
<400> 13
ctcgactatc agac 14

10 <210> 14
<211> 14
<212> ADN
<213> -

15 <400> 14
cgatgctcat agta 14

20 <210> 15
<211> 14
<212> ADN
<213> -

25 <400> 15
atcgacgcat gcag 14

30 <210> 16
<211> 14
<212> ADN
<213> -

35 <400> 16
gagtgcgatc agag 14

40 <210> 17
<211> 14
<212> ADN
<213> -

<400> 17
ctacactcac taca 14

45

50 <210> 18
<211> 14
<212> ADN
<213> -

<400> 18
tatcgtcgat gata 14

55 <210> 19
<211> 14
<212> ADN
<213> -

60 <400> 19
tctgtatatc tcac 14

65 <210> 20
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 20
gatcactcgt atca 14
5
<210> 21
<211> 14
<212> ADN
10 <213> -
<400> 21
agcatagcga cgta 14
15
<210> 22
<211> 14
<212> ADN
20 <213> -
<400> 22
tctgagcgat agta 14
25 <210> 23
<211> 14
<212> ADN
<213> -
30 <400> 23
atatctctga cgtg 14
35 <210> 24
<211> 14
<212> ADN
<213> -
40 <400> 24
gagtctagac tcag 14
45 <210> 25
<211> 14
<212> ADN
<213> -
50 <400> 25
gagtgtagtg taca 14
55 <210> 26
<211> 14
<212> ADN
<213> -
60 <400> 26
gctgtgctga gata 14
65 <210> 27
<211> 14
<212> ADN
<213> -
<400> 27

ES 2 643 712 T3

ctacgtatgt atct 14

5 <210> 28
<211> 14
<212> ADN
<213> -

10 <400> 28
tatcgcgacg tata 14

15 <210> 29
<211> 14
<212> ADN
<213> -

20 <400> 29
gagactcgtg cgtg 14

25 <210> 30
<211> 14
<212> ADN
<213> -

30 <400> 30
tatcgctcac agac 14

35 <210> 31
<211> 14
<212> ADN
<213> -

<400> 31
tatcgctaca gcgt 14

40 <210> 32
<211> 14
<212> ADN
<213> -

45 <400> 32
ctcactctca gcag 14

50 <210> 33
<211> 14
<212> ADN
<213> -

55 <400> 33
gatcagctca ctgt 14

60 <210> 34
<211> 14
<212> ADN
<213> -

65 <400> 34
agcgtgctgt atgt 14

ES 2 643 712 T3

<210> 35
<211> 14
<212> ADN
<213> -
5
<400> 35
cgacgcgcga cgag 14

10 <210> 36
<211> 14
<212> ADN
<213> -

15 <400> 36
tcgatagaca gatg 14

20 <210> 37
<211> 14
<212> ADN
<213> -

25 <400> 37
gcgatgctgt atca 14

30 <210> 38
<211> 14
<212> ADN
<213> -

35 <400> 38
ctacgcgatg ctgc 14

40 <210> 39
<211> 14
<212> ADN
<213> -

45 <400> 39
cgcacagcac agtg 14

50 <210> 40
<211> 14
<212> ADN
<213> -

55 <400> 40
gcgtcagatg ctca 14

60 <210> 41
<211> 14
<212> ADN
<213> -

65 <400> 41
tagtctcgat gcgc 14

70 <210> 42
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 42
atcacagcat gaca 14
5
<210> 43
<211> 14
<212> ADN
10 <213> -
<400> 43
ctatgctaca cgac 14
15
<210> 44
<211> 14
<212> ADN
20 <213> -
<400> 44
cgcacagatg tcgt 14
25 <210> 45
<211> 14
<212> ADN
<213> -
30 <400> 45
agacagatga gact 14
35 <210> 46
<211> 14
<212> ADN
<213> -
40 <400> 46
atatcatcgt atgt 14
45 <210> 47
<211> 14
<212> ADN
<213> -
50 <400> 47
gatgtagcac tact 14
55 <210> 48
<211> 14
<212> ADN
<213> -
60 <400> 48
tatgtatcga ctct 14
65 <210> 49
<211> 14
<212> ADN
<213> -
<400> 49

ES 2 643 712 T3

cgacagagac agtg 14

5 <210> 50
<211> 14
<212> ADN
<213> -

10 <400> 50
ctcgtagatc atgt 14

15 <210> 51
<211> 14
<212> ADN
<213> -

20 <400> 51
cgacgtctcg tcgt 14

25 <210> 52
<211> 14
<212> ADN
<213> -

30 <400> 52
tagtgcgacg ctct 14

35 <210> 53
<211> 14
<212> ADN
<213> -

40 <400> 53
gagacagaca ctgt 14

45 <210> 54
<211> 14
<212> ADN
<213> -

50 <400> 54
gcgactgat gaca 14

55 <210> 55
<211> 14
<212> ADN
<213> -

60 <400> 55
agacacgcgt gata 14

65 <210> 56
<211> 14
<212> ADN
<213> -

70 <400> 56
atcatatcac tcag 14

ES 2 643 712 T3

<210> 57
<211> 14
<212> ADN
<213> -
5
<400> 57
gatgtatcat atgc 14

10 <210> 58
<211> 14
<212> ADN
<213> -

15 <400> 58
gctcgtcgtc agta 14

20 <210> 59
<211> 14
<212> ADN
<213> -

25 <400> 59
cgatgtatca cgta 14

30 <210> 60
<211> 14
<212> ADN
<213> -

35 <400> 60
ctatgctcgt gact 14

40 <210> 61
<211> 14
<212> ADN
<213> -

45 <400> 61
tatgtgagac tata 14

50 <210> 62
<211> 14
<212> ADN
<213> -

55 <400> 62
ctcgtgagtc agta 14

60 <210> 63
<211> 14
<212> ADN
<213> -

65 <400> 63
tctgacgcga gact 14

70 <210> 64
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 64
tagtctatga ctgc 14
5
<210> 65
<211> 14
<212> ADN
10 <213> -
<400> 65
cgcgactatg ctgt 14
15
<210> 66
<211> 14
<212> ADN
20 <213> -
<400> 66
gatcgtatgt ggcg 14
25 <210> 67
<211> 14
<212> ADN
<213> -
30 <400> 67
agatcataca gact 14
35 <210> 68
<211> 14
<212> ADN
<213> -
40 <400> 68
tatgagcgtg ctgc 14
45 <210> 69
<211> 14
<212> ADN
<213> -
50 <400> 69
tcgatatatg ctgc 14
55 <210> 70
<211> 14
<212> ADN
<213> -
60 <400> 70
agcactctcg tatg 14
65 <210> 71
<211> 14
<212> ADN
<213> -
<400> 71

ES 2 643 712 T3

ctatcatatg cgta 14

5 <210> 72
<211> 14
<212> ADN
<213> -

10 <400> 72
tcgtgctgtg ctgt 14

15 <210> 73
<211> 14
<212> ADN
<213> -

20 <400> 73
gctgtagcgt agtg 14

25 <210> 74
<211> 14
<212> ADN
<213> -

30 <400> 74
tctgtgctgt tact 14

35 <210> 75
<211> 14
<212> ADN
<213> -

40 <400> 75
gagatatcgt gata 14

45 <210> 76
<211> 14
<212> ADN
<213> -

50 <400> 76
tcgactagac agta 14

55 <210> 77
<211> 14
<212> ADN
<213> -

60 <400> 77
gcgacatc agtg 14

65 <210> 78
<211> 14
<212> ADN
<213> -

cgcactatgt agta 14

ES 2 643 712 T3

<210> 79
<211> 14
<212> ADN
<213> -
5
<400> 79
agcgacgaca gcgt 14

10 <210> 80
<211> 14
<212> ADN
<213> -

15 <400> 80
ctcatatcgt gcgc 14

20 <210> 81
<211> 14
<212> ADN
<213> -

25 <400> 81
gagactatga gatg 14

30 <210> 82
<211> 14
<212> ADN
<213> -

35 <400> 82
gctcgctacg tcag 14

40 <210> 83
<211> 14
<212> ADN
<213> -

45 <400> 83
cgcactcgtg ctca 14

50 <210> 84
<211> 14
<212> ADN
<213> -

55 <400> 84
gcgtgctaca gaca 14

60 <210> 85
<211> 14
<212> ADN
<213> -

65 <400> 85
agcatgctca cgac 14

70 <210> 86
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 86
tcgtcagaca cgac 14
5
<210> 87
<211> 14
<212> ADN
10 <213> -
<400> 87
agcgtagcat agac 14
15
<210> 88
<211> 14
<212> ADN
20 <213> -
<400> 88
tatcagagat gcgt 14
25 <210> 89
<211> 14
<212> ADN
<213> -
30 <400> 89
tcgtctatcg tcgt 14
35 <210> 90
<211> 14
<212> ADN
<213> -
40 <400> 90
ctcgacgaca cgta 14
45 <210> 91
<211> 14
<212> ADN
<213> -
50 <400> 91
agacacgacg ctca 14
55 <210> 92
<211> 14
<212> ADN
<213> -
60 <400> 92
ctatctcgtc tcac 14
65 <210> 93
<211> 14
<212> ADN
<213> -
<400> 93

ES 2 643 712 T3

tctcactatc atgc 14

5 <210> 94
<211> 14
<212> ADN
<213> -

10 <400> 94
gctgagatca cgac 14

15 <210> 95
<211> 14
<212> ADN
<213> -

20 <400> 95
atatgcatc atct 14

25 <210> 96
<211> 14
<212> ADN
<213> -

30 <400> 96
gcgtgctcat atct 14

35 <210> 97
<211> 14
<212> ADN
<213> -

40 <400> 97
tctcgatcg cgtg 14

45 <210> 98
<211> 14
<212> ADN
<213> -

50 <400> 98
agatctctgt gaca 14

55 <210> 99
<211> 14
<212> ADN
<213> -

60 <400> 99
gagatatcg cgag 14

65 <210> 100
<211> 14
<212> ADN
<213> -

70 <400> 100
atcgtgagat gatg 14

ES 2 643 712 T3

<210> 101
<211> 14
<212> ADN
<213> -
5 <400> 101
tctcagcac tatg 14

10 <210> 102
<211> 14
<212> ADN
<213> -

15 <400> 102
gcgatgcgtg tcac 14

20 <210> 103
<211> 14
<212> ADN
<213> -

25 <400> 103
tcgtgtcgat agag 14

30 <210> 104
<211> 14
<212> ADN
<213> -

35 <400> 104
gatcactatg cgta 14

40 <210> 105
<211> 14
<212> ADN
<213> -

45 <400> 105
ctacgtagtc tata 14

50 <210> 106
<211> 14
<212> ADN
<213> -

55 <400> 106
gagtgcgcac tcgt 14

60 <210> 107
<211> 14
<212> ADN
<213> -

65 <400> 107
tcgtgctatg tatg 14

<210> 108
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 108
agatgtagac tact 14
5
<210> 109
<211> 14
<212> ADN
10 <213> -
<400> 109
atacagagat atgc 14
15
<210> 110
<211> 14
<212> ADN
20 <213> -
<400> 110
ctacagctcg tata 14
25 <210> 111
<211> 14
<212> ADN
<213> -
30 <400> 111
cgacgtctga gatg 14
35 <210> 112
<211> 14
<212> ADN
<213> -
40 <400> 112
ctcgagcgat ggcg 14
45 <210> 113
<211> 14
<212> ADN
<213> -
50 <400> 113
tagatagatc atca 14
55 <210> 114
<211> 14
<212> ADN
<213> -
60 <400> 114
agcatgcgac taca 14
65 <210> 115
<211> 14
<212> ADN
<213> -
<400> 115

ES 2 643 712 T3

cgcgagctga ctca 14

5 <210> 116
<211> 14
<212> ADN
<213> -

10 <400> 116
gatgtagacg tcgc 14

15 <210> 117
<211> 14
<212> ADN
<213> -

20 <400> 117
cgcatgatca gatg 14

25 <210> 118
<211> 14
<212> ADN
<213> -

30 <400> 118
gcgtcatcga gcgc 14

35 <210> 119
<211> 14
<212> ADN
<213> -

<400> 119
ctacagcgtc atct 14

40 <210> 120
<211> 14
<212> ADN
<213> -

45 <400> 120
atacgcgcac tcac 14

50 <210> 121
<211> 14
<212> ADN
<213> -

55 <400> 121
ctatgtatga gcac 14

60 <210> 122
<211> 14
<212> ADN
<213> -

65 <400> 122
ctcgtgatga cgag 14

ES 2 643 712 T3

<210> 123
<211> 14
<212> ADN
<213> -
5
<400> 123
cgcgacgcac atgc 14

10 <210> 124
<211> 14
<212> ADN
<213> -

15 <400> 124
ctatcagacg tcag 14

20 <210> 125
<211> 14
<212> ADN
<213> -

25 <400> 125
atacagagtg cgag 14

30 <210> 126
<211> 14
<212> ADN
<213> -

35 <400> 126
atacgtatca ctgc 14

40 <210> 127
<211> 14
<212> ADN
<213> -

45 <400> 127
cgacagctca gcac 14

50 <210> 128
<211> 14
<212> ADN
<213> -

55 <400> 128
ctatcagcga gatg 14

60 <210> 129
<211> 14
<212> ADN
<213> -

65 <400> 129
atcactcgat agtg 14

70 <210> 130
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 130
cgcatatcat atct 14
5
<210> 131
<211> 14
<212> ADN
10 <213> -
<400> 131
cgcgagagtg taca 14
15
<210> 132
<211> 14
<212> ADN
20 <213> -
<400> 132
gagatgctga gcgc 14
25 <210> 133
<211> 14
<212> ADN
<213> -
30 <400> 133
agcatgagtg ctgt 14
35 <210> 134
<211> 14
<212> ADN
<213> -
40 <400> 134
cgcgtagcga gcag 14
45 <210> 135
<211> 14
<212> ADN
<213> -
50 <400> 135
agatcatcat gcac 14
55 <210> 136
<211> 14
<212> ADN
<213> -
60 <400> 136
tctgtataca cgta 14
65 <210> 137
<211> 14
<212> ADN
<213> -
<400> 137

ES 2 643 712 T3

tctgtatcat gaca 14

5 <210> 138
<211> 14
<212> ADN
<213> -

10 <400> 138
tcgtctctca gata 14

15 <210> 139
<211> 14
<212> ADN
<213> -

20 <400> 139
atacacgatg tatg 14

25 <210> 140
<211> 14
<212> ADN
<213> -

30 <400> 140
atacgctcga ctca 14

35 <210> 141
<211> 14
<212> ADN
<213> -

40 <400> 141
tctgacgatg cgag 14

45 <210> 142
<211> 14
<212> ADN
<213> -

50 <400> 142
gatgagcgtc agag 14

55 <210> 143
<211> 14
<212> ADN
<213> -

60 <400> 143
gagtctcgtc atct 14

65 <210> 144
<211> 14
<212> ADN
<213> -

70 <400> 144
atacgtagat gcag 14

ES 2 643 712 T3

<210> 145
<211> 14
<212> ADN
<213> -
5
<400> 145
gatcacgatc tcac 14

10 <210> 146
<211> 14
<212> ADN
<213> -

15 <400> 146
atcactagtc atgc 14

20 <210> 147
<211> 14
<212> ADN
<213> -

25 <400> 147
tagtgcgctg gatg 14

30 <210> 148
<211> 14
<212> ADN
<213> -

35 <400> 148
gctcagatcg tact 14

40 <210> 149
<211> 14
<212> ADN
<213> -

45 <400> 149
gctgactaca ctct 14

50 <210> 150
<211> 14
<212> ADN
<213> -

55 <400> 150
agcgagagac atct 14

60 <210> 151
<211> 14
<212> ADN
<213> -

65 <400> 151
tctcgtctga ctct 14

70 <210> 152
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 152
gatgacgcac agta 14
5
<210> 153
<211> 14
<212> ADN
10 <213> -
<400> 153
atcacagatg cgac 14
15
<210> 154
<211> 14
<212> ADN
20 <213> -
<400> 154
gcgactctcg ctgc 14
25 <210> 155
<211> 14
<212> ADN
<213> -
30 <400> 155
cgacagcgat gaca 14
35 <210> 156
<211> 14
<212> ADN
<213> -
40 <400> 156
agatcagcac atca 14
45 <210> 157
<211> 14
<212> ADN
<213> -
50 <400> 157
gctgtgagat gcac 14
55 <210> 158
<211> 14
<212> ADN
<213> -
60 <400> 158
gcgatagcat gcag 14
65 <210> 159
<211> 14
<212> ADN
<213> -
<400> 159

ES 2 643 712 T3

gatcgtagtg cgac 14

5 <210> 160
<211> 14
<212> ADN
<213> -

10 <400> 160
ctcacatacg ctct 14

15 <210> 161
<211> 14
<212> ADN
<213> -

20 <400> 161
cgatgctgat atgt 14

25 <210> 162
<211> 14
<212> ADN
<213> -

30 <400> 162
atcactatga ctct 14

35 <210> 163
<211> 14
<212> ADN
<213> -

40 <400> 163
tcgatagcgt atgt 14

45 <210> 164
<211> 14
<212> ADN
<213> -

50 <400> 164
atacactcgt agag 14

55 <210> 165
<211> 14
<212> ADN
<213> -

60 <400> 165
gcgtgtatgt gact 14

65 <210> 166
<211> 14
<212> ADN
<213> -

70 <400> 166
tatcagatga cgta 14

ES 2 643 712 T3

<210> 167
<211> 14
<212> ADN
<213> -
5
<400> 167
gagactctgt agac 14

10 <210> 168
<211> 14
<212> ADN
<213> -

15 <400> 168
gcgtgcgcga cgta 14

20 <210> 169
<211> 14
<212> ADN
<213> -

25 <400> 169
tagtcagatc tcgc 14

30 <210> 170
<211> 14
<212> ADN
<213> -

35 <400> 170
agacagcgac tcgt 14

40 <210> 171
<211> 14
<212> ADN
<213> -

45 <400> 171
cgatcagatc agac 14

50 <210> 172
<211> 14
<212> ADN
<213> -

55 <400> 172
gatgacgaca gatg 14

60 <210> 173
<211> 14
<212> ADN
<213> -

65 <400> 173
agcgagatgt gcac 14

70 <210> 174
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -
<400> 174
gcgtcagcac tata 14
5
<210> 175
<211> 14
<212> ADN
10 <213> -
<400> 175
agcgataca ctgc 14
15
<210> 176
<211> 14
<212> ADN
20 <213> -
<400> 176
ctcatgctgt gact 14
25 <210> 177
<211> 14
<212> ADN
<213> -
30 <400> 177
gagatgatcg tata 14
35 <210> 178
<211> 14
<212> ADN
<213> -
40 <400> 178
tatgactatc tact 14
45 <210> 179
<211> 14
<212> ADN
<213> -
50 <400> 179
tagatgatgt agtg 14
55 <210> 180
<211> 14
<212> ADN
<213> -
60 <400> 180
tcgacatcac tcgt 14
65 <210> 181
<211> 14
<212> ADN
<213> -
<400> 181

ES 2 643 712 T3

gctgactcac tcgc 14

5 <210> 182
<211> 14
<212> ADN
<213> -

10 <400> 182
gagtgtctcg tcac 14

15 <210> 183
<211> 14
<212> ADN
<213> -

20 <400> 183
cgacactcga gcgt 14

25 <210> 184
<211> 14
<212> ADN
<213> -

30 <400> 184
tatcagcga ctgc 14

35 <210> 185
<211> 14
<212> ADN
<213> -

40 <400> 185
agatgtatgt agag 14

45 <210> 186
<211> 14
<212> ADN
<213> -

50 <400> 186
agacgtcgac agac 14

55 <210> 187
<211> 14
<212> ADN
<213> -

60 <400> 187
atcgagctcg tcag 14

65 <210> 188
<211> 14
<212> ADN
<213> -

<400> 188
tctgagatgt atgc 14

ES 2 643 712 T3

<210> 189
<211> 14
<212> ADN
<213> -
5
<400> 189
agcgactcga cgac 14

10 <210> 190
<211> 14
<212> ADN
<213> -

15 <400> 190
gagacatatg tact 14

20 <210> 191
<211> 14
<212> ADN
<213> -

25 <400> 191
gctgtgagtg cgtg 14

30 <210> 192
<211> 14
<212> ADN
<213> -

35 <400> 192
ctcatgcgtc tcgt 14

40 <210> 193
<211> 14
<212> ADN
<213> -

45 <400> 193
tatgactcgt gcac 14

50 <210> 194
<211> 14
<212> ADN
<213> -

55 <400> 194
agacgctcac atgt 14

60 <210> 195
<211> 14
<212> ADN
<213> -

65 <400> 195
cgcgtagatg cgtg 14

70 <210> 196
<211> 14
<212> ADN

ES 2 643 712 T3

<213> -

<400> 196
ctacacgaca gact 14

5

<210> 197
<211> 14
<212> ADN
<213> -

10

<400> 197
cgatgtcgtg cgag 14

15

<210> 198
<211> 14
<212> ADN
<213> -

20

<400> 198
tagatgcgac agac 14

REIVINDICACIONES

1. Procedimiento para la deconvolución de mezclas de sustancias que contienen ácidos nucleicos, en el que:
- 5 a) se generan a partir de varios nucleótidos (A, C, G, T/U) según un algoritmo predeterminado varias secuencias nucleotídicas diana (TNS) diferentes entre sí ($A_1-A_n, B_1-B_n, \dots, Z_n$) con posiciones de secuencia N_0-N_n , de las que
- 10 b) respectivamente al menos una TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) se asigna respectivamente a al menos una sustancia o combinación de sustancias y se acopla químicamente con esta, y
- c) se proporciona al menos una mezcla de sustancias para analizar con al menos dos TNS distintas contenidas ($A_1-A_n, B_1-B_n, \dots, Z_n$) o sustancias acopladas con TNS, que
- 15 d) se secuencian según un procedimiento de secuenciación, en el que simultáneamente se recogen todas las TNS contenidas en la mezcla de sustancias ($A_1-A_n, B_1-B_n, \dots, Z_n$) en un espectro de secuencia común, en el que
- e) las secuencias superpuestas en el espectro de secuencia se deconvolucionan mediante cribado de las posiciones de secuencia N_0-N_n según el algoritmo predeterminado y se identifican según su asignación,
- 20 caracterizado porque se predetermina una ocupación nucleotídica para los nucleótidos (A, C, G, T/U) en las posiciones de secuencia N_0-N_n de una TNS para formar ($A_1-A_n, B_1-B_n, \dots, Z_n$) mediante una condición establecida por el algoritmo predeterminado que está ligada a una ocupación nucleotídica de al menos una posición de secuencia adicional.
- 25 2. Procedimiento según la reivindicación 1, caracterizado porque se criban según el algoritmo predeterminado las posiciones de secuencia N_0-N_n que presentan una intensidad de señal significativamente elevada de nucleótidos individuales (A, C, G, T/U).
- 30 3. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se sustraen por etapas las intensidades de señal significativamente elevadas de nucleótidos individuales (A, C, G, T/U) en el espectro de secuencia hasta alcanzar en cada posición de secuencia N_0-N_n como máximo un nucleótido (A, C, G, T/U) de intensidad de señal detectable mínima, en el que los espectros de sustracción obtenidos a este respecto, que presentan respectivamente al menos segmentos de secuencia o fragmentos de secuencia, se criban en las
- 35 posiciones de secuencia N_0-N_n según el algoritmo predeterminado.
4. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se predetermina según el algoritmo predeterminado para cada posición de secuencia $N_{0+1}-N_n$ de una TNS para formar ($A_1-A_n, B_1-B_n, \dots, Z_n$), una limitación referida a un nucleótido (A, C, G, T/U) de una posición de secuencia precedente respectivamente para
- 40 al menos un nucleótido (A, C, G, T/U).
5. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se diferencian entre sí las TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) en al menos dos posiciones de secuencia y/o por al menos una serie de nucleótidos consistente en al menos 5 posiciones de secuencia.
- 45 6. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se forman TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) con una longitud de al menos 5 posiciones de secuencia.
7. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se asignan TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) que presentan una diferencia de secuencia de al menos un 50 %, preferiblemente un 75 %, con especial
- 50 preferencia un 90 %, respectivamente a sustancias que presentan las mayores diferencias estructurales y/o funcionales entre sí.
8. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se forman TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) que presentan al menos un segmento de secuencia que codifica un grupo de sustancias, tamaño de
- 55 sustancia, coordenadas geográficas de una variedad de exposición o una fecha.
9. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque las TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) presentan al menos un segmento de secuencia mediante el cual son identificables como TNS.
- 60 10. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se trata en las TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) de moléculas de ARN o ADN monocatenarias o bicatenarias.
11. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se combinan entre sí las
- 65 TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$) y/o segmentos de secuencia de TNS ($A_1-A_n, B_1-B_n, \dots, Z_n$).

12. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se unen covalentemente las TNS (A_1 - A_n , B_1 - B_n , ..., Z_n) con su sustancia o sustancias asignadas.
- 5 13. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se lleva a cabo al menos una etapa de procedimiento para la selección de al menos una de las TNS (A_1 - A_n , B_1 - B_n , ..., Z_n) y/o al menos una de las sustancias acopladas a TNS, en el que se proporciona una mezcla de sustancias para analizar a partir de una fase móvil líquida y/o una elución líquida de una fase estacionaria.
- 10 14. Procedimiento según una de las reivindicaciones precedentes, caracterizado porque se secuencian la mezcla de sustancias para analizar mediante el uso de dideoxinucleótidos marcados fluorescentemente, al menos una polimerasa y al menos un cebador que es complementario de un segmento de secuencia de al menos una TNS (A_1 - A_n , B_1 - B_n , ..., Z_n).

Fig. 1

TNS según el siguiente algoritmo

N_0
z-a-a-d-d-e-a-a-d-d-e-a-d-e N_{13} ← X

Operación-z: A/C/G/T

Operación-a: A->G/T; C->G/T; G->A/C; T->A/C

Operación-d: A->C/T; C->A/G; G->A/T; T->C/G

Operación-e: A->C/G; C->A/T; G->C/T; T->A/G

Número de diferentes TNS por número de diferencias de secuencia

Diferencia de 0 ----0
 Diferencia de 1 ----0
 Diferencia de 2 ----0
 Diferencia de 3 ----0
 Diferencia de 4 ----0
 Diferencia de 5 ----0
 Diferencia de 6 ----820
 Diferencia de 7 ----1141
 Diferencia de 8 ----1569
 Diferencia de 9 ----2269
 Diferencia de 10 ----3027
 Diferencia de 11 ----3544
 Diferencia de 12 ----3382
 Diferencia de 13 ----2535
 Diferencia de 14 ----1216

← 10

Fig. 2

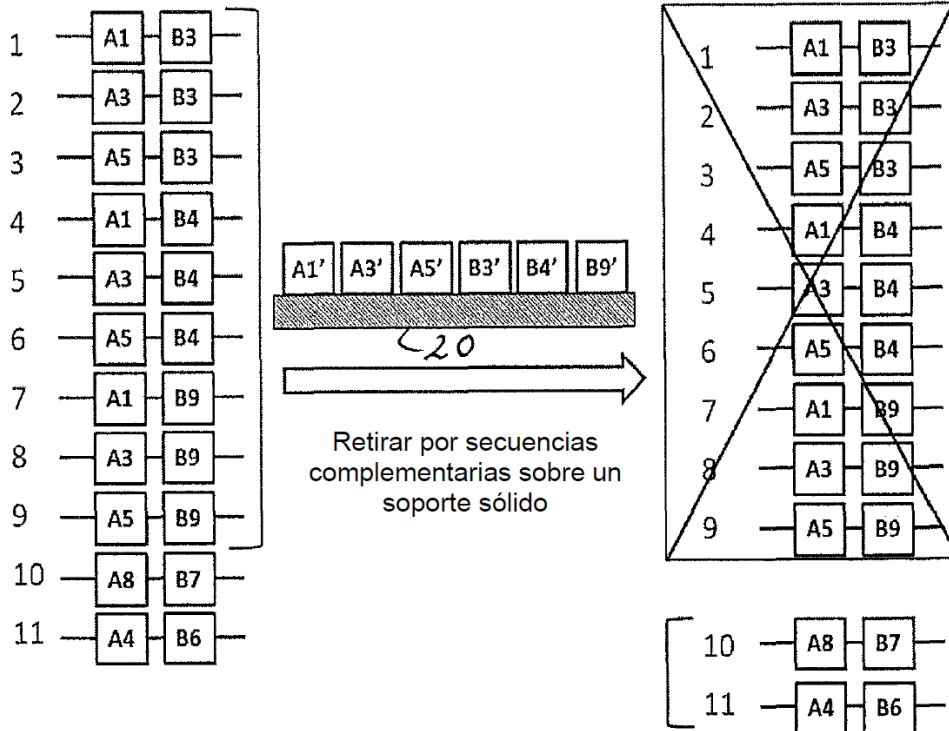


Fig. 3

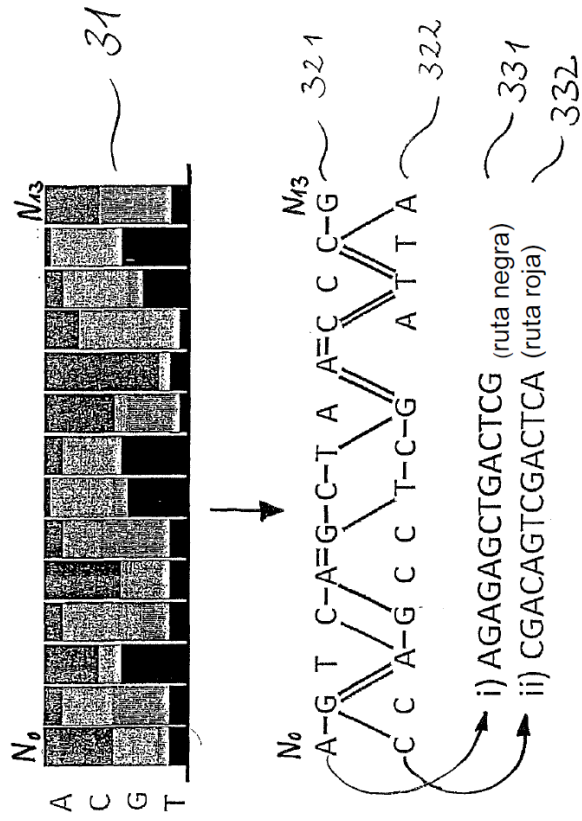


Fig. 4a

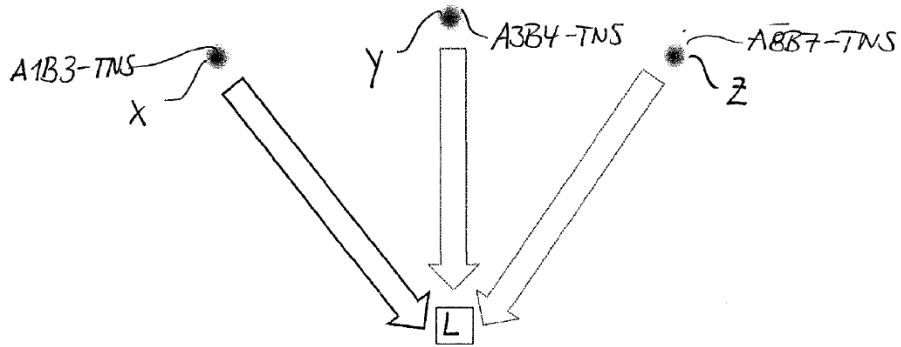


Fig. 4b

