(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 551 577**

(51) Int. CI.:

G06F 19/00 (2011.01)
C12N 15/10 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **19.06.2007 E 12167464 (2)**

(97) Fecha y número de publicación de la concesión europea: **29.07.2015 EP 2487616**

(54) Título: **Alargamiento Iterado Programable: Un método para elaborar genes sintéticos y ADN combinatorio y librerías de proteína**

(30) Prioridad:

19.06.2006 US 814570 P
27.04.2007 US 924042 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
20.11.2015

(73) Titular/es:

YEDA RESEARCH AND DEVELOPMENT COMPANY LIMITED (100.0%)
The Weizmann Institute of Science P.O. Box 95
76100 Rehovot, IL

(72) Inventor/es:

SHAPIRO, EHUD Y.;
LINSHIZ, GREGORY;
BEN-YEHEZKEL, TUVAL;
KAPLAN, SHAI;
ADAR, RIVKA;
GRONAU, ILAN y
TUVI, SIVAN

(74) Agente/Representante:

LAZCANO GAINZA, Jesús

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

Alargamiento Iterado Programable: Un método para elaborar genes sintéticos y ADN combinatorio y librerías de proteína.

Campo de la invención

La presente divulgación es un sistema y método para construir objetos físicos, incluyendo pero no estando limitado a polímeros tales como polinucleótidos y polipéptidos, y librerías de los mismos, preferiblemente de manera automática pero opcionalmente al menos semiautomáticamente.

Antecedentes de la invención

Los avances en la tecnología de la síntesis de oligodesoxiribunocleótido (ODN) y la purificación han conducido a oportunidades para la síntesis de genes de novo. Existen dos aproximaciones principales a la construcción de un gen sintético. La primera aproximación involucra las síntesis de los ODN que comprende la secuencia completa. Los ODN son hibridados en una forma fragmentaria seguida por la unión con la ligasa del ADN. Los fragmentos de ADN son entonces clonados en un vector plásmido directamente o después de la amplificación por PCR. La segunda aproximación es un método a base de PCR en el cual múltiples ODN que pertenecen a las dos hebras de la secuencia del gen de interés son hibridados mediante traslapos cortos y luego extendidos mediante una polimerasa estable al calor utilizando regiones traslapadas como cebadores. Las secuencias de genes sintéticos se pueden optimizar para la expresión máxima al eliminar codones raros y utilizar codones óptimos para unas especies particulares o para múltiples sistemas de hospedero, optimización de secuencia de la transcripción e iniciación de la traducción/región de terminación, modificación de la estabilidad del mensajero, minimizar la estructura secundaria y ajustar el contenido GC. Los sitios de restricción existentes se pueden retirar y/o se pueden agregar nuevos sitios de restricción para preparar el gen para futuras estrategias de clonación.

Otro ejemplo de un método para el montaje de un gen involucra le tecnología de fase sólida, como se describió por Stahl et al., en “Solid Phase Gene Assembly of Constructs Derived from the Plasmodium falciparum Malaria Blood-Stage Antigen Ag332”[iii]. El oligonucleótido se sintetiza mientras se une a un soporte sólido. El procedimiento sintético utiliza ligasas en lugar de polimerasa, con el fin de enlazar los segmentos de ADN preformados. Se describe un método similar por Hostomsky et al., “Solid Phase Assembly of Cow Colostrum Trypsin Inhibitor Gene”[iv].

Los autores hacen notar que los rendimientos de los oligonucleótidos de longitud completa disminuyen sustancialmente en la medida en que la longitud de los oligonucleótidos (gen sintetizado) se incrementa. Este resultado se esperaría para cualquier proceso de ensamblaje de oligonucleótido/gen que utiliza ligasa, en razón de que no existe una función de corrección. Adicionalmente, no existe posibilidad de incrementar el rendimiento al permitir que se efectúen ciclos adicionales de ligado, de tal manera que los rendimientos de los oligonucleótidos correctamente formados de longitud completa se esperarían que cayeran fuertemente en la medida en que la longitud del oligonucleótido deseado se incrementa.

La Patente US No. 6, 386, 861 también describe un método para crear una librería de oligonucleótidos de ADN, al recombinar o “barajar” familias de oligonucleótidos. Sin embargo, en lugar de crear una librería de acuerdo a un conjunto de principios matemáticamente definidos, la librería se crea solamente de acuerdo a homología con la secuencia o grupo de secuencias previamente definidas, limitando de esta manera el alcance de las librerías que se puedan crear al utilizar este método.

Resumen de la Invención

La técnica anterior no enseña ni sugiere un método o sistema para construir de manera recursiva un objeto físico de una pluralidad de subcomponentes. La técnica anterior tampoco enseña ni sugiere tal método o sistema a para construir una librería que contenga una pluralidad de objetos físicos.

La técnica anterior tampoco enseña ni sugiere un método o sistema para construir tal objeto físico que esté libre de error o al menos reduzca los errores, de una pluralidad de subcomponentes de los cuales al menos uno tiene al menos un error. La técnica anterior tan poco enseña ni sugiere tal método o sistema para construir una librería que contenga una pluralidad de objetos físicos.

La técnica anterior tampoco enseña ni sugiere un método o sistema para elaborar un ADN sintético de manera directa, fácilmente automatizada mayor de 100 nucleótidos (nts). La técnica anterior tampoco enseña ni sugiere un método para síntesis sistemática de ADN combinatorio o la librería de proteína de una pluralidad de especificaciones o características definidas.

La presente divulgación soluciona estas deficiencias de la técnica anterior al suministrar un sistema y método para construir un objeto final de una pluralidad de subcomponentes, de una manera eficiente que sea opcional y preferiblemente capaz de ser al menos parcialmente automatizado y más preferiblemente completamente automatizado. La presente divulgación también caracteriza, en algunas realizaciones, un sistema y método para construir una librería para tales objetos finales.

De acuerdo con unas realizaciones, se suministra un proceso de construcción jerárquico, recursivo, preferiblemente un proceso de construcción recursivo modificado, para construir un objeto final de una pluralidad de subcomponentes. Estos subcomponentes se utilizan para producir el objeto final. El proceso es preferiblemente "de recurrencia modificada" porque el tipo de objeto final puede ser opcionalmente diferente del tipo de los subcomponentes de entrada. Opcionalmente y preferiblemente, la entrada y salida de cada subproceso dentro del proceso de construcción puede ser de un tipo diferente. La construcción recursiva, los pares de elementos más pequeños (u opcionalmente los grupos de al menos tres elementos más pequeños) están compuestos sistemáticamente en elementos siempre mayores hasta que se construya el elemento objetivo. Un prerrequisito de una construcción recursiva matemáticamente estricta es que la salida de la operación de composición sea del mismo tipo que sus dos o más entradas. Sin embargo, tal como se anota en la presente divulgación puede opcionalmente caracterizar la construcción recursiva o recursiva modificada, de tal manera que el objeto final pueda ser opcionalmente un tipo diferente de uno o más subcomponentes y/o la salida de cada subproceso pueda ser opcionalmente de un tipo diferente de una o más entradas al subproceso.

De acuerdo con las realizaciones preferidas, al menos un subcomponente puede opcionalmente ser erróneo o al menos potencialmente erróneo, de tal manera que se utilice al menos un subcomponente erróneo o potencialmente erróneo en el proceso de construcción que es luego preferiblemente ajustado para solucionar tal error o error potencial. En general, un objeto compuesto construido de bloques de construcción propensos al error se espera que tengan un mayor número de errores que cada uno de sus bloques de construcción. Sin embargo, si los errores se distribuyen de manera aleatoria entre los bloques de construcción y ocurren de manera aleatoria durante la construcción, y si suficientes copias de un objeto son construidas, se espera que algunas de las copias puedan contener componentes libres de error. Si tales componentes se pudieran identificar y extraer de los objetos fallidos, ellos podrían ser reutilizados como bloques de construcción para otra construcción recursiva del objeto. Ya que esta construcción inicia desde los bloques de construcción típicamente mayores que están libres de error, el número de errores en el objeto resultante se espera que disminuya, posiblemente bajando a cero. Aún si los objetos producidos de esta manera tuvieran errores, se esperaría que ellos tuvieran menores errores que sus predecesores y de esta manera tuvieran aún componentes mayores libres de error, lo cual se podría utilizar como bloques de construcción para una construcción recursiva subsecuentemente correctiva, hasta que un objeto libre de error (o al menos con error reducido) se forme.

De acuerdo con algunas realizaciones de la técnica anterior, el sistema y método son para elaborar genes sintéticos y ADN combinatorio y librerías de proteína, o cualquier otro polímero biológico al menos semiautomáticamente, pero preferiblemente automáticamente. El método se puede utilizar en de una manera sistemática o automatizada para sintetizar una molecula de ADN de cualquier largo, u otro polímero biológico, y, más generalmente, cualquier librería molecular combinatoria que tenga una pluralidad de moléculas de polímero biológicas que opcionalmente tengan la propiedad matemática de ser un conjunto regular de cadenas.

Gran parte de la investigación en biología molecular requiere moléculas de ADN, u otros biopolímeros, preparados de acuerdo con la especificación. Aunque existe un método automatizado estándar para preparar oligonucleótidos de ADN de alrededor de 100 nts[v], moléculas de ADN mayores y librerías de molécula se preparan típicamente a través de una multitud de métodos ad hoc, intensivos en trabajo como se describió previamente en la sección de antecedentes anterior. No existe un lenguaje común para especificar librerías de ADN y/o moléculas de ADN mayores, ni un método automatizado de propósito general, para preparar estas librerías y/o moléculas más largas.

Se debe notar que estos procesos se efectúan preferiblemente al menos semiautomáticamente, y más preferiblemente de manera completamente automática. El proceso híbrido es susceptible de automatización de extremo a extremo, teniendo como entrada una especificación matemática y produciendo como salida las moléculas especificadas, y como tal tengan el potencial de transformar la naturaleza y marcha de la investigación en biología molecular.

De acuerdo con realizaciones preferidas, se suministra un método para la construcción recursiva de un oligonucleótido. El método es preferiblemente automatizado. El método opera preferiblemente de manera jerárquica, al calcular primero los bloques de construcción jerárquicos requeridos para construir el polinucleótido y luego construir el polinucleótido de los bloques de construcción en un orden determinado de acuerdo con la jerarquía. Opcionalmente, se puede efectuar una pluralidad de procesos de construcción en paralelo, cuyos bloques resultantes pueden opcional y preferiblemente ser utilizados para construir la molécula final (en cuyo caso más preferiblemente se utilizan dos de tales procesos paralelos). El conjunto completo de reacciones se expresa preferiblemente como un árbol jerárquico y el conjunto completo de reacciones para la corrección del error se determina más preferiblemente de acuerdo con un algoritmo de corte mínimo.

El método también utiliza preferiblemente bloques de construcción perfectos, aunque los bloques de construcción imperfectos pueden ser opcionalmente utilizados como una etapa intermedia. Se pueden utilizar bloques de construcción imperfectos inicialmente, pero los bloques de construcción perfectos se obtienen preferiblemente de los bloques de construcción imperfectos y luego se utilizan para construir directamente el polinucleótido.

De acuerdo con realizaciones preferidas, se suministra un método para reducir sino eliminar los errores que ocurran durante la síntesis del polinucleótido, preferiblemente a través del proceso de eliminación de error iterativo. Tal proceso se efectúa preferiblemente como parte del proceso recursivo de la construcción del oligonucleótido. La construcción recursiva con eliminación de error iterativo mejora sobre las aproximaciones previas a la síntesis de ADN al posibilitar la construcción rápida, completamente automatizada y libre de error de moléculas de ADN largas. Tal método también puede ser opcionalmente utilizado para combinar segmentos de ADN sintéticos y naturales. Adicionalmente, tal método posibilita opcionalmente el diseño eficiente, la síntesis y la corrección del error de las librerías de ADN combinatorias con componentes compartidos y variantes. Opcionalmente, tal método se podría utilizar para sintetizar nuevos genomas artificiales sintéticos.

Opcionalmente y preferiblemente, la librería (o cualquier otra librería de moléculas de polinucleótido que tengan al menos una región variable y al menos una región compartida) se construye al primero analizar la pluralidad de secuencias de polinucleótido para crear un conjunto de bloques de construcción de acuerdo a las regiones compartidas y variables, y luego construir en paralelo cada molécula en la librería de los bloques de construcción.

De acuerdo con otras realizaciones preferidas, el método de la presente invención se puede opcionalmente utilizar para preparar una librería de variantes de una secuencia de ADN que utilice cualquiera de los ácidos nucleicos existentes de manera natural (por ejemplo plásmido, cADN etc..) relacionado con esa secuencia como entrada. Por ejemplo, con el fin de generar diferentes variantes de un gen en la forma de una librería de variantes para el gen, el método de la presente invención puede opcionalmente utilizar el ADN del gen en sí mismo como el material de partida para la construcción de la librería.

Como se utiliza aquí, el término “librería molecular combinatoria” incluye preferiblemente una pluralidad de polímeros biológicos, y más preferiblemente incluye un número grande de tales polímeros biológicos que representan muchas combinaciones diferentes de los bloques de construcción de los polímeros biológicos.

En lo sucesivo, el término “oligonucleótido largo” incluye preferiblemente un oligonucleótido de al menos aproximadamente 100 bases de longitud, más preferiblemente incluye un oligonucleótido de al menos 200 bases de longitud y más preferiblemente incluye un oligonucleótido de al menos aproximadamente 400 bases de longitud.

Breve descripción de los dibujos

La invención se describe aquí, a manera de ejemplo solamente, con referencia a los dibujos que la acompañan, en donde:

La FIG. 1 muestra la síntesis recursiva para uso en el método de la presente invención;

La FIG. 2 muestra un perfil esquemático de un protocolo de síntesis de D&C de ejemplo de acuerdo con la presente invención;

La FIG. 3 muestra la producción de ssADN con exonucleasa lambda de acuerdo con el método de D&C;

La Fig. 4A muestra los procesos recursivos básicos para la síntesis de ADN D&C, mientras que la Fig. 4B muestra un proceso para determinar el o los protocolos de síntesis óptima.

La FIG. 5 muestra la medición de la capacidad de unión del ssADN marcado con biotina ((A) unión específica; (B) Unión no específica ((A);

La FIG. 6 muestra una etapa de alargamiento, transferencia, separación y liberación con los siguientes carriles: (1) Marcador; (2) Control (ADN no alargados); (3) alargamiento mediante T7 Pol; (4) Alargamiento Mediante Fragmento Klenow; (5) Sobrenadante de Lavado 1; (6) sobrenadante no unido a perlas; (7) Sobrenadante de Lavado 2; (8) Separación de NaON; (9) Sobrenadante de Lavado 3; (10) perlas después de hervir; (11) cadena biotinilada liberada, marcada con P32;

La FIG. 7 muestra oligo alargamiento con cantidades diferentes de Fragmento Klenow (gel nativo). C, alargamiento antes de control. 10u, 20u, 30u, cantidad total de polimerasa utilizada para el alargamiento de 50 pmol oligos. Los alargamientos se analizaron después de 0,60, 120 y 180 min;

La FIG. 8 muestra el alargamiento de dos pares de oligos mediante la Versión Sequenase 2.0;

La FIG. 9 muestra (A) alargamiento de 30 pmol de ADN mediante 1.65u de la Version Sequenase 2.0 en diferentes puntos de tiempo. (a) ssADN antes de alargamiento (b) Heterodímero después de hibridación (c) puntos de tiempo (B) Alargamiento de 30 pmol de ADN mediante 1, 4, 8, 16 u de la Versión Sequenase 2.0 después de 30 min.

La FIG. 10 muestra una prueba de actividad de exonucleasa de la Exonucleasa Lambda. El ADN con sentido fue marcado en 3' mediante Fluorescina y el ADN contra sentido se visualizó mediante tenido con bromuro de Etidio. (1) Digestión de ADN contra sentido después de la Fosforilación Química. (II) Digestión de ADN con sentido después de la fosforilación mediante PNK. (III) Digestión de ADN contra sentido después de la fosforilación mediante PNK.

La FIG. 11 muestra un árbol D&C de ejemplo que muestra el concepto de acuerdo con del protocolo de diseño. Los círculos rosados muestran las cadenas de ADN que son fosforiladas.

La FIG. 12 muestra el progreso de dos niveles de síntesis. S29_ 112 oligo marcada mediante Cy5 en 5' (rojo). El teñido con Bromuro de Etidio es verde. (1), (2) y (3) son etapas de reacción de acuerdo con el protocolo. En cada reacción: Hibridación (Ann), Alargamiento (Elon), selección de cadena (Exo).

La FIG. 13 muestra electroforesis capilar. El resumen de las dos etapas de síntesis 84nt-> 145nt y 145nt->239 nt; la muestra naranja fue marcada mediante Cy5. El marcador es Marcado con Naranja mediante Liz.

La FIG. 14 muestra errores en la secuencia. (A) sustitución del "C" en lugar del "G". (B) supresión del "C".

La FIG. 15 muestra la digestión del exceso de ssADN mediante Exonucleasa I; el carril 1 muestra la hibridación de dos cadenas; el carril 2 muestra el alargamiento mediante ADN polimerasa; el carril 3 la digestión del exceso de ssADN mediante exonucleasa 1.

La FIG. 16 muestra la purificación de la nucleasa Mung Bean a 80ºC 30 min. El carril A muestra hibridación; el carril E muestra alargamiento; el carril MB muestra limpieza de la nucleasa Mung Bean;

La FIG. 17 muestra una prueba de concepto con secuencias naturales que describen el protocolo de síntesis de D&C para un polinucleótido de ADN de 737 bp de longitud.

La FIG. 18 muestra el alineamiento de secuencia del alargamiento del producto 3 de alargamiento obtenido con la secuencia esperada.

La FIG. 19 muestra la secuencia del producto PCR de un oligonucleótido;

La FIG. 20 muestra el alineamiento de la secuencia de los productos PCR obtenidos después de clonación;

La FIG. 21 muestra una etapa de alargamiento proveniente de 2 productos PCR de oligonucleótido antes de clonación.

La FIG. 22 muestra un alineamiento de las secuencias clonadas de una etapa de alargamiento de 2 productos PCR de oligonucleótido;

La FIG. 23 muestra la selección y la purificación del oligonucleótido;

La FIG. 24 muestra los resultados experimentales de una división no coincidente enzimática mediante endonucleasa T7; solo luego de la no coincidencia, existe digestión del dúplex de ADN mediante la endonucleasa. El color naranja representa el marcador, el color verde representa el sustrato de ADN;

La FIG. 25 muestra un método de ejemplo con corrección de error de acuerdo con la presente invención;

La FIG. 26 muestra un método de ejemplo para la síntesis del gen GFP;

La FIG. 27A-E muestra el análisis de fermento C.E después de las reacciones PCR durante la construcción del GFP;

La FIG. 28A-B muestran el análisis de fragmento C.E después de las reacciones PCR durante la construcción del GFP proveniente de dos fragmentos;

La FIG. 29 muestra el error versus la longitud objetivo para los clones;

La FIG. 30A muestra seis variantes predefinidas del gen p53 (componentes compartido (gris) y únicos (coloreados) de la librería de ADN combinatoria p53 se muestran);

Las FIGS. 30B-D muestran una gráfica que describe el proceso de construcción de la librería que corresponde al D&C basado en división in sílico; las secuencias de librería objetivo fueron recursivamente divididas en componentes compartidos y únicos y luego divididas adicionalmente en oligonucleótidos básicos utilizando D&C. Estas oligosecuencias fueron hechas (con errores) mediante síntesis oligo química convencional. Utilizando la construcción recursiva in vitro con estos oligos como entrada, fueron compuestas las aproximaciones propensas a error de todas las seis moléculas de la librería p53 objetivo. Estas variantes p53 fueron clonadas, secuenciadas y se identificaron los errores (marcados con rojo). Un corte mínimo libre de error (por debajo de la línea negra, no sombreada) del plan de construcción de la librería original se computó de solamente cuatro clones propensos a error (variantes 1, 3, 5, 6) para todas las seis variantes de librería. Los segmentos libres de error por fuera de estos cuatro clones (delimitados) se utilizaron como entrada para la reconstrucción recursiva libre de error de todos los seis miembros de librería objetivo a través de los clones y a través de las variantes de acuerdo con la gráfica de corte mínima computarizada;

La FIG. 31A-F muestra el análisis de fragmento C.E después de las reacciones PCR durante el método de construcción de librería p53 tal como se efectuó anteriormente:

La FIG. 32A-C muestra el análisis de gel después de las reacciones PCR;

La FIG. 33 muestra el análisis de fragmento C.E. de un ejemplo de una reacción de alargamiento;

La FIG. 34 muestra un ejemplo de una reacción de exonucleasa lambda, que genera los ssADN provenientes del dsADN mediante exonucleasa lambda;

La FIG. 35 se relaciona con una vigilancia en tiempo real del proceso de amplificación PCR que muestra que se utiliza la población única para el smPCR;

La FIG. 36 se relaciona con el secuenciamiento de una reacción PCR que inicia desde una molécula de ADN plantilla con una secuencia;

La FIG. 37 muestra la electroforesis de gel del ADN proveniente del PSR de molécula única con un gen sintético recursivamente construido que codifica para GFP como plantilla. La FIG. 37A muestra un control positivo que contiene más de una molécula por reacción. La FIG. 37B muestra el PCR del gen GFP con una concentración de 0,5 moléculas por reacción de PCR.

La FIG. 38 muestra un sistema automatizado de ejemplo de acuerdo con la presente invención;

La FIG. 39 muestra una primera realización ilustrativa de un aparato de acuerdo con la presente invención, que muestra un aparato 3900;

La FIG. 40 muestra otra realización opcional, que se relaciona con la configuración de un aparato con un componente sintetizador OLIGOS;

La FIG. 41 se relaciona con la configuración de un aparato con un soporte sólido; y

La FIG. 42 se relaciona con una configuración de aparato con chips microfluidos y de ADN.

Descripción de las realizaciones preferidas

De acuerdo con algunas realizaciones, se suministra un proceso de construcción jerárquico, recursivo, preferiblemente un proceso de construcción recursivo modificado, para la construcción de un objeto final de una pluralidad de subcomponentes. Estos subcomponentes se utilizan para producir el objeto final. El proceso es preferiblemente “recurrencia modificada” porque el tipo de objeto final puede opcionalmente ser diferente del tipo de los subcomponentes de entrada. Opcionalmente y preferiblemente, la entrada y salida de cada uno de los subprocesos dentro del proceso de construcción puede ser de tipo diferente. En una construcción recursiva, los pares de elementos más pequeños (u opcionalmente los grupos de al menos tres elementos más pequeños) se componen sistemáticamente en elementos siempre mayores hasta que se construye el elemento objetivo. Un prerrequisito de la construcción recursiva matemáticamente estricta es que la salida de la operación de composición sea del mismo tipo que sus dos o más entradas. Sin embargo, como se anota en la presente divulgación puede opcionalmente caracterizar la construcción recursiva o recursiva modificada, de tal manera que el objeto final pueda opcionalmente ser de un tipo diferente de uno o más de los subcomponentes y/o la salida de cada subproceso pueda opcionalmente ser de un tipo diferente de una o más entradas al subproceso.

Los subcomponentes pueden opcionalmente ser de un tipo o alternativamente de una pluralidad de tipos. Por “tipo” se significa una categoría de subcomponentes en lugar de una instancia particular del mismo. Un ejemplo no limitante de un tipo es un tipo de un polímero o un tipo de una unidad de un polímero. Ejemplos no limitantes de tales tipos incluyen aminoácidos, nucleótidos, polinucleótidos, oligonucleótidos, péptidos, polipéptidos, ADN o ARN de cadena doble o cadena sencilla y/o cualquier tipo de grupo funcional y/o molécula, o combinaciones de los mismos. De acuerdo con algunas realizaciones de la presente invención, el tipo es preferiblemente ADN o ARN de doble cadena o de cadena sencilla, o combinaciones de los mismos. Otros ejemplos no limitantes de un tipo son un componente eléctrico, que incluye (sin limitación) cualquier componente de un dispositivo electrónico, o cualquier librería molecular combinatoria que tenga la propiedad matemática de ser un conjunto regular de cadenas.

De acuerdo con las realizaciones preferidas, el proceso para construir de manera recursiva el objeto final se determina preferiblemente al menos de manera parcial automáticamente, al determinar una senda eficiente para combinar una pluralidad de subcomponentes. Por “eficiente” se significa que al menos uno de una mayoría de subprocesos para combinar una pluralidad de subcomponentes se prefiere de acuerdo con uno o más criterios de tiempo, entrada de energía, subproductos y gastos monetarios, y/o que al menos el número de subcomponentes y/o subprocesos se requieran para preparar el objeto final.

Preferiblemente, la senda eficiente se determina de acuerdo con el método de “divide y conquistarás” como se describe aquí o una variación del mismo. El método de divide y conquistarás inicia con el objeto final y luego lo descompone en una pluralidad de subcomponentes. El método busca un protocolo óptimo para la construcción del objeto blanco bajo un conjunto de restricciones y de conjunto de parámetros de coste. El método caracteriza preferiblemente una descomposición óptima o al menos eficiente del objeto final suministrado por el usuario a una pluralidad de subcomponentes, que pueden ser entonces preferiblemente ensamblados y/o combinados con el objeto final de acuerdo a uno o más subprocesos, cada uno de los cuales tiene un protocolo. Este proceso de descomposición se efectúa preferiblemente iterativamente una pluralidad de veces, de tal manera que uno o más subcomponentes de los subprocesos que se efectúan posteriormente en el proceso de construcción se ensamblan y/o construyen preferiblemente a través de subprocesos anteriores.

De acuerdo con las realizaciones preferidas, el proceso de descomposición preferiblemente incluye la búsqueda de un punto de división óptimo y válido del objeto final. Tal punto preferiblemente cumple con un conjunto de restricciones tal como la existencia de subcomponentes específicos y la especificidad de traslapo para el punto de división, además de la existencia de dos protocolos válidos que pueden construir las dos partes divididas. Una vez que se encuentra el punto de división válido el algoritmo busca de manera recursiva un protocolo para construir sus dos partes. La recurrencia preferiblemente se detiene cuando el objeto blanco ha llenado uno o más requisitos. Una función de costos se computa para cada subprotocolo basado en el número de subcomponentes y los costos de los mismos, el número de reacciones y el número de tres niveles de protocolos requeridos para construir su blanco. El protocolo de coste más pequeño se selecciona como el protocolo óptimo. En razón a que el espacio del protocolo es muy grande, el algoritmo de programación dinámico se utiliza preferiblemente para mantener los subprotocolos computados previamente en una caché la cual es entonces preferiblemente reutilizada cuando se requiere en una diferente senda de búsqueda. Además, una rama y un algoritmo de unión se utilizan preferiblemente para cortar el espacio de búsqueda cuando el coste intermedio busca que no se pueda mejorar el mejor coste corriente para un protocolo.

Ejemplos de restricciones y conjuntos de parámetros de coste para síntesis de polinucleótidos incluyen la preferencia por los subcomponentes a ser sintetizables como cebadores oligo, el tamaño de los oligonucleótidos, el número de los oligonucleótidos, el número de las reacciones (como un ejemplo del número de subproceso), el número de niveles en la jerarquía de construcción, la especificidad de unión, la afinidad de unión sea que los subcomponentes, tales como los oligonucleótidos, estén listo para usarse o sea que ellos deban se preparados, y el uso de ADN o ARN sintético o natural. Otros parámetros que pueden ser opcionalmente generalizados para cualquier tipo de proceso de construcción incluyen entrada de energía y cualquier coste o dificultad en cada uno de los subprocesos. Algunos de los parámetros específicos anteriores también pueden opcionalmente ser generalizados para otros tipos de objetos finales, que incluyen pero no están limitados a, el número de subprocesos, el número de niveles de jerarquía de construcción, el tamaño y/o otros costos involucrados en crear los subcomponentes, el número de subcomponentes y si los subcomponentes están listos para uso o alternativamente se deben preparar.

De acuerdo con una realización, la construcción recursiva (u opcionalmente construcción recursiva modificada) puede ser opcionalmente utilizada para la construcción de una librería que comprende una pluralidad de objetos finales. El algoritmo recibe una descripción de librería con regiones variables y/o aspectos separados al compartir las regiones y/o aspectos de los objetos finales. Cada región variable puede tener dos o más variantes con una o más diferentes propiedades. Opcionalmente y preferiblemente, una senda eficiente para combinar una pluralidad de subcomponentes para formar la librería de una pluralidad de objetos finales, más preferiblemente de acuerdo a una o más propiedades de los aspectos compartidos y/o variables y/o las regiones de los objetos finales.

De acuerdo con realizaciones preferidas de la misma, la aproximación de divide y conquistarás se utiliza para determinar la senda eficiente para la construcción combinatoria de la librería. Utilizando la aproximación de divide y conquistarás, el algoritmo encuentra el protocolo de librería óptimo para construir la librería desde sus regiones compartidas y variables con números mínimos de reacciones o subprocesos. El algoritmo entonces encuentra un traslapo válido específico dentro de las regiones compartidas que es adecuado para sintetizar dos regiones adyacentes con todas sus variantes. El traslapo define los bloques de construcción (subcomponente) de la librería, preferiblemente para tanto las regiones compartidas como variantes y/o los aspectos de los objetos finales. Más generalmente, la optimización combinatoria se efectúa preferiblemente con el fin de minimizar el número total de reacciones requeridas de los subcomponentes. Cada objeto final es entonces planeado utilizando el algoritmo de divide y conquistarás para el objeto final (descrito anteriormente). Los protocolos para preparar los objetos finales se mezclan preferiblemente y se agregan a reacciones adicionales de acuerdo al protocolo de construcción óptima como se calculó previamente.

De acuerdo con las realizaciones preferidas, al menos un subcomponente puede opcionalmente ser erróneo o al menos potencialmente erróneo, de tal maneras que al menos un subcomponente erróneo o potencialmente erróneo se utiliza en el proceso de construcción que es luego preferiblemente ajustado para solucionar tal error o error potencial. En general, un objeto compuesto construido de bloques de construcción propensos a error se espera que tenga un número mayor de errores que cada uno de sus bloques de construcción. Sin embargo, si los errores son distribuidos aleatoriamente entre los bloques de construcción y ocurren aleatoriamente durante la construcción, y si se construyen suficientemente muchas copias de un objeto, se espera que algunas de las copias puedan contener componentes libres de error. Si tales componentes se pudieran identificar y extraer de los objetos fallidos, ellos podrían ser reutilizados como bloques de construcción para otra construcción recursiva del objeto. En razón a que esta construcción inicia desde los bloques de construcción típicamente mayores que están libres de error, el número de errores en el objeto resultante se espera que disminuya, posiblemente bajando a cero.

Aún si los objetos producidos de esta manera tuvieran errores, se esperaría que tuvieran menores errores que sus predecesores y de esta manera tuvieran aún componentes libres de error mayores, lo cual se podría utilizar como bloques de construcción para una construcción recursiva correctiva posterior, hasta que se forme preferiblemente un objeto libre de errores (o al menos con errores reducidos).

De acuerdo con realizaciones preferidas, los subcomponentes máximos libres de errores se utilizan como bloques de construcción para otro intento de construcción. Específicamente, los subcomponentes se analizan preferiblemente para encontrar el corte mínimo en el árbol de construcción recursivo (descrito con mayor detalle adelante). El corte mínimo se utiliza preferiblemente para determinar que partes de cuales objetos de salida (sub componentes de salida) están libres de error y para utilizar estas partes libres de error con el fin de construir un objeto que tenga errores reducidos (sino libres de error). Más preferiblemente se utiliza una estructura de árbol para este método. Un nodo en el árbol está cubierto por un conjunto de subcomponentes de salida si la parte del objeto final representada por el nodo está libre de error en al menos un subcomponente de salida. El corte mínimo define preferiblemente todas las partes del objeto final de acuerdo a la mayoría de los nodos superficiales (aquellos más cercanos a la raíz). El corte mínimo existe y es único si todas las hojas del árbol están cubiertas por las partes de los subcomponentes.

Si alguna hoja no está cubierta significa que su parte relacionada es errónea en todos los subcomponentes de salida, en cuyo caso los subcomponentes adicionales se deben analizar y el corte mínimo recomputar, o el subcomponente se puede reconstruir y el proceso efectuar de nuevo si el subcomponente es erróneo. Las partes correctas (libres de error) de los subcomponentes pueden ser opcionalmente extraídas y utilizadas para el proceso de construcción recursiva de acuerdo al corte mínimo.

De acuerdo con realizaciones opcionales, una librería de una pluralidad de objetos finales puede opcionalmente experimentar corrección de error separadamente tal como se describió anteriormente. De manera alternativa y preferiblemente, sin embargo, cada subcomponente de la librería, compartida y/o variante, es preferiblemente corregida para los errores tal como se describió anteriormente, después de lo cual la librería completa se construye de estos subcomponentes libres de error.

No existen ejemplos limitantes de otros bloques de construcción (subcomponentes) que incluyan opcionalmente oligonucleótidos sintéticos, construidos preferiblemente de una manera altamente paralela sobre un chip de ADN en el cual cada bloque de construcción de ADN sobre el chip de ADN sea más preferiblemente accesible de manera individual para uso en la construcción utilizando este método; la síntesis de estado sólido; los fragmentos de ADN natural o los fragmentos de ADN producidos de acuerdo a cualquiera de los otros métodos sintéticos (se debe notar que) donde se hace referencia aquí a ADN que el ARN y/o cualquier tipo de polinucleótido u oligonucleótido no natural está incluido). Otros ejemplos no limitantes opcionalmente incluyen péptidos sintetizados de acuerdo con cualquier método sintético; péptidos o proteínas producidas de manera natural; o cualquier otro tipo de grupo funcional y/o molécula, que incluye preferiblemente unidades poliméricas para construir un polímero.

De acuerdo con realizaciones preferidas, el método de divide y conquistarás determina un “corte mínimo” de acuerdo con lo cual una pluralidad de subcomponentes erróneos potenciales o posibles se utilizan para construir al menos un objeto final libre de error de acuerdo a una senda eficiente. Una prueba y descripción del corte mínimo se suministra adelante:

Un corte en un árbol es un conjunto de nodos que incluye un nodo único sobre cualquier senda de la raíz a una hoja.

Sea T un árbol de protocolo de construcción recursivo y S un conjunto de subcomponentes. Se establece que S cubre T si existe un conjunto de subcomponentes C tales que cada subcomponente en C sea un subcomponente en algún subcomponente de S y C es un corte C de T. En tal caso se debe establecer que S cubre T con C.

Reivindicación: si S cubre T, entonces existe un conjunto mínimo único C tal que S cubre T con C.

Prueba de algoritmo de reconstrucción libre de error: dado un protocolo RC T y un conjunto de subcomponentes S, encuentre un C mínimo tal que S cubra T con C.

Entonces C se crea y se efectúa la construcción recursiva partiendo con C.

Se suministran diversos ejemplos específicos adelante para computar el corte mínimo.

Sección 1- Polinucleótidos y otros biopolímeros

Sin querer estar limitados de ninguna manera, esta sección se relaciona con realizaciones de ejemplo, ilustrativas de la presente invención (método y sistema) con relación a la construcción de polinucleótido (y también la construcción de otros biopolímeros, ya que los métodos de adelante se pueden utilizar opcionalmente para la construcción de los polipéptidos). Se debe notar que tales polinucleótidos pueden opcionalmente comprender ADN y/o ARN y/o cualquiera de los aminoácidos no naturales y/o modificados, y puede opcionalmente ser de cadena doble y/o cadena sencilla (incluidas dentro de la misma molécula). También, cualquier referencia hecha a “ADN” aquí, es con el propósito de explicación solamente ya que el polinucleótido puede comprender opcionalmente cualquiera de los tipos descritos anteriormente o combinación de los mismos.

Las realizaciones descritas anteriormente para la construcción general de los objetos finales se utilizan preferiblemente para la construcción de las secuencias de polinucleótido. Una descripción más detallada de las realizaciones de ejemplo para los métodos y sistemas para el uso con secuencias de polinucleótido se suministra adelante.

De acuerdo con las realizaciones preferidas, se suministra un proceso de construcción jerárquico, recursivo, preferiblemente un proceso de construcción recursivo modificado, para construir una secuencia de polinucleótido de una pluralidad de secuencias de nucleótido más cortas. Estas secuencias de nucleótido más cortas pueden opcionalmente ser, pero no estar limitadas a, oligonucleótidos, fragmentos de ADN o ARN natural, o pueden ser opcionalmente derivadas al utilizar técnicas varias, incluyendo pero no estando limitados a, PCR, clonación y así sucesivamente.

De acuerdo con otras realizaciones preferidas, el método de la presente divulgación utiliza el método D&C para dividir la secuencia de ADN blanco in sílico en fragmentos suficientemente cortos para hacer sintetizados mediante síntesis de oligonucleótido convencional, aunque con errores; estas moléculas propensas a error son recursivamente combinadas in vitro, formando moléculas de ADN blanco propensas a error. Las partes libres de error de estas moléculas se identifican, extraen y utilizan como entradas nuevas, típicamente mayores y más precisas a otra iteración del procedimiento de construcción recursivo y el proceso completo se repite hasta que se forma una molécula blanco libre de error.

En la construcción recursiva, los pares de elementos más pequeños están compuestos sistemáticamente en elementos siempre mayores hasta que se construya el elemento blanco. De acuerdo con algunas realizaciones, la salida de la operación de composición es del mismo tipo que sus dos entradas; sin embargo se debe notar que otros tipos de recurrencias o procesos similares a recurrencia se pueden implementar opcionalmente para la presente invención. Para estas realizaciones, se suministra una reacción de composición novedosa que toma dos fragmentos de ssADN traslapantes como entrada y producen una molécula de ssADN alargada como salida.

En general, un objeto compuesto construido de los bloques de construcción propensos a error se espera que tenga un número mayor de errores que cada uno de sus bloques de construcción. Sin embargo, si los errores son distribuidos aleatoriamente entre los bloques de construcción y ocurren aleatoriamente durante la construcción, y si se construyen suficientes copias de un objeto, se espera que alguna de las copias pueda contener componentes libres de error. Si tales componentes se pudieran identificar y extraer de los objetos fallidos, ellos podrían ser reutilizados en bloques de construcción para otra construcción recursiva del objeto. Ya que esta construcción inicia desde los bloques de construcción típicamente mayores que están libres de error, el número de errores en el objeto resultante se espera que

disminuya, posiblemente bajando a cero. Aún si los objetos producidos de esta manera tuvieran errores, ellos esperarían que tuvieran menos errores que sus predecesores y de esta manera tuvieran siempre mayores componentes libres de error, lo cual se podría utilizar como bloques de construcción para una construcción recursiva correctiva posterior hasta que se forme un objeto libre de error. Así, aunque el método utiliza bloques de construcción que son oligonucleótidos que pueden caracterizar errores, el producto final es un poli nucleótido que está libre de error.

De acuerdo con realizaciones preferidas, las secuencias de ADN máximas libre de error se utilizan como bloques de construcción para otros intentos de construcción. Específicamente, las secuencias de ADN se analizan preferiblemente para encontrar un corte mínimo en el árbol de construcción recursivo, definido como sigue: un nodo en el árbol se dice que está cubierto por un conjunto de secuencias de ADN si sus secuencias ocurren libres de error en al menos una de las secuencias de ADN. Un conjunto de secuencias de ADN de moléculas construidas de acuerdo con el árbol de construcción recursivo induce a un corte mínimo sobre el árbol, definido por ser el conjunto de nodos más superficiales (más cercanos a la raíz) que estén cubiertos por las secuencias de ADN. El corte mínimo existe y es único si todas las hojas del árbol están cubiertas por las secuencias de ADN. Si alguna hoja no está cubierta significa que su secuencia es errónea en todas las secuencias de ADN, en cuyo caso las secuencias de ADN adicionales deben ser analizadas y recomputado el corte mínimo, y el oligo se puede resintetizar y el proceso de efectuar de nuevo si el oligo sintético es erróneo. Ya que los fragmentos de ADN que constituyen el corte mínimo ocurren en el árbol de construcción recursivo, sus límites coinciden con los límites de los fragmentos del árbol de construcción recursivo inicial. Por lo tanto, ellos se pueden extraer de sus respectivas secuencias de ADN y ser utilizados como bloques de construcción para la construcción recursiva correctiva. Esto se ejecuta al utilizar sus respectivas secuencias de ADN como entradas a la etapa de composición en el árbol de construcción recursivo que corresponde a ese fragmento específico, tal como se describe con mayor detalle adelante.

Un nodo en el árbol se dice que está cubierto por un conjunto de secuencias de ADN si la secuencia del nodo ocurre libre de error en al menos una de las secuencias de ADN. Un conjunto de secuencias de ADN de las moléculas construidas de acuerdo a un árbol de construcción recursivo induce un corte mínimo en el árbol, definido por ser el conjunto de los nodos más superficiales (más cercanos de la raíz) en el árbol que este cubierto por las secuencias de ADN. El corte mínimo existe y es único si todas las hojas del árbol están cubiertas por las secuencias de ADN. Si alguna hoja no está cubierta significa que su secuencia es errónea en todas las secuencias de ADN, en cuyo caso las secuencias de ADN adicionales se deben analizar y recomputar el corte mínimo, o el oligo se puede resintetizar y el proceso efectuar de nuevo si el oligo sintético es erróneo. Ya que los fragmentos de ADN que constituyen el corte mínimo ocurren en el árbol de construcción recursivo, sus límites coinciden con los límites de los fragmentos del árbol de construcción recursivo inicial. Por lo tanto, ellos se pueden extraer de sus respectivas secuencias de ADN y ser utilizadas como bloques de construcción para la construcción recursiva correctiva. Esto se ejecuta al utilizar sus respectivas secuencias de ADN como entrada para la etapa de composición en el árbol de construcción recursivo que corresponda a aquel fragmento específico, tal como se describió con mayor detalle adelante.

Computar el corte mínimo:

Una aproximación recursiva se utiliza para computar el corte mínimo de un árbol de protocolo. Cada nodo en el árbol representa un proceso bioquímico con un producto y dos precursores. El algoritmo inicia con la raíz del árbol (molécula objetivo) y para cada nodo revisa si su secuencia de producto existe sin errores en una de las secuencias de ADN o se están utilizando otras secuencias. Se debe notar que para los propósitos de explicación solamente y sin desear estar limitado de ninguna manera, esta descripción se centra alrededor del uso de las secuencias de ADN como bloques de construcción; sin embargo, opcionalmente y alternativamente, el PCR de molécula única o cualquier otro método de amplificación molecular (particularmente cualquier otro método de amplificación de molécula única) se podría utilizar para amplificar las moléculas de ADN únicas. También, cualquier tipo de bloque de construcción adecuado se podría utilizar opcionalmente.

Si tal secuencia de ADN existe este producto se marca como nuevo bloque de construcción básico para la reconstrucción de la molécula objetivo y su par cebador y secuencia de ADN relevante (como plantilla) se registran como su reacción PCR generante. Si no hay una secuencia de ADN que contenga una secuencia libre de error del producto nodo la reacción se registra como reacción existente en el nuevo protocolo y el algoritmo es ejecutado recursivamente en los dos precursores del producto. La salida de tal protocolo es un árbol de reacciones que comprende un corte mínimo del árbol original, Este contiene hojas para las cuales los productos libres de error existen y que todos sus nodos internos no tienen secuencias de ADN libre de error que los contengan. Un programa automatizado que utiliza estos nuevos bloques de construcción libres de error para la construcción recursiva de la molécula objetivo se genera para el robot.

Computar el número requerido de secuencias de ADN (bloques de construcción)

Para un fragmento del tamaño L bajo una tasa de mutación R la probabilidad de tener un fragmento libre de error en una secuencia de ADN simple se toma de la distribución de Poisson con LAMBDA = L* R (la probabilidad de tener 0 errores

cuando los errores esperados son L * R). Para encontrar el número más pequeño de secuencias de ADN requeridas para encontrar un fragmento libre de error con una probabilidad mayor del 95% se utilizó una distribución binomial y la probabilidad de tener al menos un fragmento libre de error en N secuencias de ADN se computa.

En la aproximación divide y conquistarás la longitud del fragmento puro se puede reducir al tamaño de un oligo (~ 80bp) y el coste de tener que efectuar más etapas durante la reconstrucción. Así con el fin de garantizar que el árbol tiene un cubrimiento total libre de errores de la molécula de la secuencia blanco, la probabilidad de tener un fragmento puro del tamaño L en N secuencias de ADN- P Exitosas (L, N) es multiplicada por sí misma, el número de fragmento de tamaño L que se requiere para construir la molécula objetivo (la primera parte está libre de error y la segunda parte está libre de error etc…). Este número se calcula después de considerar el traslapo que reduce la contribución de cada oligo para ser más pequeño que su tamaño actual (~55bp). Luego, se encuentra el número más pequeño de secuencias de ADN (bloques de construcción) que satisface el requisito de que la probabilidad total de tener un corte mínimo excederá el 95%.

Una breve descripción de una realización particular de un método de ejemplo de acuerdo con la presente invención se discute adelante:

Algoritmo

INPUT:

TARGET_SEQUENCE y CONFIGURATION con valores de parámetro.

OUTPUT:

Un protocolo para construir la molécula objetico que incluye:

● SEQUENCE LIST: una lista de todos los oligos y los productos objetivo intermedios

● REACTION LIST: una lista de reacciones, cada reacción tiene dos secuencias de la lista de secuencias y su entrada y una secuencia como su producto describiendo así el árbol de protocolo.

● Lista de OLIGOS y PRIMER: una lista de oligos que es necesaria para sintetizar los bloques de construcción para los protocolos.

1. Pre – procesamiento

1.1 Encontrar el mejor traslapo para cada punto y computar su rango de especificidad

1.2 Encontrar el mejor cebador para cada punto y computar su rango de especificidad.

2. Dividir & Conquistar (TRAGET_ SEQUENCE)

2.1 Si TRAGET_ SEQUENCE está en la Cache ENTONCES regresar el protocolo de la cache

2.2 Si TRAGET_SEQUENCE es más corto que MAX_ OLIGO: SIZE de regresar el OLIGO al protocolo corriente.

2.3 Establecer CURRENT_ PROTOCOL a NO_ PROTOCOL y CURRENT_ BEST_ COST a Inf.

2.4 Para cada Punto de División revisar lo siguiente:

i. Existe traslapo válido y cumplir con las restricciones de traslapo.

ii. Existen cebadores para división y cumplir con las restricciones de los cebadores

iii. Revisar dímeros de Cebadores para cada subobjetivo.

iv. IF cualquiera de las revisiones 2.3.1 – 2.3.3 falló THEN continuar al siguiente punto.

v. Computar la unión inferior sobre el coste del protocolo IF LOWER_ BOUND_ COST > CURRENT _BEST_COST continuar al siguiente punto.

vi. Divide &Conquistarás (LEFT_SUBTRAGET)

Vii. Divide & Conquistarás (RIGHT_SUBTRAGET)

Viii. Mezclar protocolos y computar el CURRENT_COST

ix. IF CURRENT_COST < BEST_CURRENT_COST establecer CURRENT_BEST_PROTOCOL = CURRENT_ PROTOCOL (autorizar cache).

2.5 Regresar CURRENT_ BEST_ PROTOCOL

Parámetros

max_oligo_len: una longitud de 80 oligo máximo

min_oligo_len: longitud 30 oligo mínimo

max_primer_Tm: 70 máximo de temperatura de fusión del cebador

min_primer_Tm: 60 mínimo de temperatura de fusión de traslapo

min_primer_len: 14 longitud de cebador mínima.

max_primer_len: longitud cebador 30 máximo

min_overlap_Tm: 60 mínimo temperatura de fusión de traslapo

min_overlap: 15 mínimo longitud de traslapo

max_overlap: 70 máximo longitud de traslapo

levelcost: 50 coste de nivel adicional en el protocolo

reactioncost: 10 coste de reacción adicional en el protocolo

oligo_constant_cost: 2 Costo constante para un oligo simple

oligo_nuc_cost: 0,2500 de longitud dependiente del coste para un oligo.

primer_ident_nucs_allowed: 15

primer_ sequential_ident_nucs_allowed:8

primer_3p_match_allowed: 5

prim3_len: 6

overlap_ident_nucs_allowed: 10

overlap_sequential_ident_nucs_allowed: 7

overlap_3p_match_allowed: 6

overlap_prim3_len: 7

min_external_fragment: 200

primer_dimer_any: 8

primer_hetrodimer_any: 8

skip_oligo_oligo_primers: 1

Algoritmos Divide& Conquistarás para síntesis de librería combinatoria de ADN.

Las búsquedas de algoritmo que utilizan la aproximación de divide y conquistarás para un protocolo para construir una librería combinatoria descrita por el usuario con una utilización eficiente de las secuencias compartidas de librería.

Descripción general:

El algoritmo recibe una descripción de librería con regiones variables separadas por regiones compartidas. Cada región variable pude tener 2 o más variantes de diferente secuencia y tamaño.

Utilizando la aproximación de divide y conquistarás, el algoritmo encuentra el protocolo de librería óptimo para construir la librería de subregiones compartidas y variables con número mínimo de reacciones considerando que el producto intermedio puede ser una multiplicación de los tamaños de dos regiones variables. El algoritmo tiende a encontrar un traslapo válido específico dentro de las regiones compartidas que es adecuado para sintetizar las dos regiones adyacentes con todas sus variantes. El traslapo define los bloques de construcción de la librería, tanto los compartidos como los fragmentos variantes. Cada bloque de construcción es entonces planeado utilizando el algoritmo Divide & conquistarás para una secuencia única (descrita anteriormente).

Los protocolos de los bloques de construcción se mezclan y se agregan reacciones adicionales de acuerdo al protocolo de construcción óptimo de las librerías previamente calculadas.

Ejemplos no limitantes de otros bloques de construcción incluyen oligonulceótidos sintéticos, construidos preferiblemente de una manera altamente paralela sobre el chip de ADN en el cual cada bloque de construcción de ADN sobre el chip de ADN es más preferiblemente accesible de manera individual para uso en la construcción utilizando este método; síntesis de estado sólido; fragmentos de ADN natural y cualquier otro método sintético.

La presente invención también puede opcionalmente ser utilizada para construir genes largos y librerías de ADN. Las realizaciones opcionales de la presente invención con relación a las aplicaciones en biología molecular e ingeniería incluyen pero no están limitadas a, descubrir nuevos aspectos en las interacciones proteína – proteína, creaciones de proteínas con nueva función catalítica; creación de nuevas proteínas provenientes de plegamientos de proteínas conocidas; investigación de módulos de transcripción y muchos más.

La presente invención puede ser opcionalmente utilizada para un número de diferentes aplicaciones, ejemplos no limitantes de los cuales incluyen librerías de ADN de construcción que consisten de todas las secuencias específicas en tanto que la longitud de los genes completos, y la creación de las dependencias dentro de una secuencia y entre secuencias en la librería (mutagénesis condicional) . Adicionalmente ejemplos no limitantes incluyen nuevas posibilidades de investigación en biología molecular e ingeniería, que incluyen:

1. Descubrir los nuevos aspectos de las interacciones proteína- proteína. La mutación de punto sobre la superficie que está encargada de las interacciones proteína-proteína afecta su estructura de proteína. Sin embargo la mutagénesis condicional puede crear unas nuevas mutaciones relacionadas sobre ambas superficies de proteínas interactuantes, permitiendo la preservación de la estructura de la proteína.

2. Creación de las proteínas con nueva función catalítica: Crear librerías de proteína que consisten de mutaciones condicionales en lados activos de proteínas conocidas permite la creación de nuevas proteínas con una alta probabilidad de tener nuevas funciones catalíticas. Esto también ayuda a encontrar las conexiones entre los patrones de secuencia y las estructuras tridimensionales creadas por aquellos patrones.

3. La creación de nuevas proteínas provenientes de los plegamientos de las proteínas conocidas: el PIE soporta la creación de nuevas proteínas construyéndolas desde los plegamientos de proteínas conocidas, y prediciendo sus funciones y estructuras con alta probabilidad.

4. Investigar los módulos de transcripción: utilizar mutagénesis condicional, se pueden crear nuevos módulos de transcripción, permitiendo la investigación de las interacciones entre diferentes factores de transcripción.

5. Explorar la evolución del ARN: Crear librerías de ARN que consisten de secuencias con dependencias condicionales que permiten explorar la evolución de la molécula y las interacciones mutuas en la sopa de ARN.

6. Explorar el uso del codón en diferentes sistemas de expresión: el PIE permite construir sistemas que prueban la influencia del uso del codón sobre la traducción y el plegamiento de proteínas en diferentes sistemas de expresión.

7. Evolución paralela del gen: Efectuar la evolución de varios genes en paralelo y ensamblarlos en operones o virus largos.

8. Creación de proteínas con función o funciones nuevas y/o de unión modificada.

9. Creación de anticuerpos

10. Generar genomas artificiales/nuevos/ modificados.

Ácidos nucleicos no naturales.

La presente invención puede ser ejecutada opcionalmente con uno o más ácidos nucleicos no naturales. Ejemplos no limitantes de tales ácidos nucleicos no naturales incluyen índoles (5- sustituidos (que incluyen pero no están limitados a derivados de 5- trifluorometilo, derivados de 5-trifluorometoxi y derivados de 5-nitroindol), 3- nitropirroles, bases de carbazol (que incluyen pero no están limitadas a 3-nitrocarbazol y 3,6- dinitrocarbazol) y bases de fenilo 4-sustituidas, así como también ácidos nucleicos peptídicos (PNA), análogos de ADN con estructuras neutras, seudopéptidicas capaces de hibridar con una cadena de ADN complementaria, y/o hipoxantina.

Aminoácidos no naturales.

La siguiente tabla se relaciona con aminoácidos no convencionales o modificados que se pueden utilizar con la presente invención.

Aminoácido no convencional	Código	Aminoácido no convencional	Código
Ácido α-aminobutírico	Abu	L-N-metilalanina	Nmala
α -amino- α -metilbutirato	Mgabu	L-N-metilarginina	Nmarg
aminociclopropano-	Cpro	L-N-metilasparagina	Nmasn
Carboxilato		Ácido L-N-metilaspártico	Nmasp
Ácido aminoisobutírico	Aib	L-N-metilcisteína	Nmcys
Aminonorbornilo-	Norb	L-N-metilglutamina	Nmgin
Carboxilato		Ácido L-N-metilglutámico	Nmglu
Ciclohexilalanina	Chexa	L-N-metilhistidina	Nmhis
Ciclopentilalanina	Cpen	L-N-metilisoleucina	Nmile
D-alanina	Dal	L-N-metileucina	Nmleu
D-arginina	Darg	L-N-metilisina	Nmlys
Ácido D-aspártico	Dasp	L-N-metilmetionina	Nmrnet
D-cisteína	Dcys	L-N-metilnorleucina	Nmnle
D-glutamina	Dgln	L-N-metilnorvalina	Nmnva
Ácido D-glutámico	Dglu	L-N-metilomitina	Nmorn
D-histidina	Dhis	L-N-metilfenilalanina	Nmphe
D-isoleucina	Dile	L-N-metilprolina	Nmpro
D-leucina	Dleu	L-N-metilserina	Nmser
D-lisina	Dlys	L-N-metiltreonina	Nmthr
D-metionina	Dmet	L-N-metiltriptofano	Nmtrp
D-omitina	Dom	L-N-metiltirosina	Nmtyr
D-fenilalanina	Dphe	L-N-metilvalina	Nmval
D-prolina	Dpro	L-N-metiletilglicina	Nmetg
D-serina	Dser	L-N-metil-t-butilglicina	Nmtbug
D-treonina	Dthr	L-norleucina	Nle
D-triptofano	Dtrp	L-norvalina	Nva
D-tirosina	Dtyr	α-metil-aminoisobutirato	Maib
D-valina	Dval	α-metil-γ-aminobutirato	Mgabu
D-α-metilalanina	Dmala	α-metilciclohexilalanina	Mchexa
D-α-metilarginina	Dmarg	α-metilciclopentilalanina	Mcpen
D-α-metilasparagina	Dmasn	α-metil- α-naftilalanina	. Manap
D-α-metilaspartato	Dmasp	α-metilpenicilamina	Mpen
D-α-metilcisteína	Dmcys	N-(4-aminobutil)glicina	Nglu
D-α-metilglutamina	Dmgln	N-(2-arninoetil)glicina	Naeg
D-α-metilhistidina	Drnhis	N-(3-arninopropil)glicina	Nom
D-α-metilisoleucina	Dmile	N-amino-α-metilbutirato	Nmaabu
D-α-metileucina	Dmleu	α-naftilalanina	Anap
D-α-metilisina	Dmlys	N-bencilglicina	Nphe
D-α-metilmetionina	Drnmet·	N-(2-carbamiletil)glicina	Ngln
D-α-metilomitina	Dmorn	N-(carbamilmetil)glicina	Nasn

Aminoácido no convencional	Código	Aminoácido no convencional	Código
D-α-metilfenilalanina	Dmphe	N-(2-carboxietil)glicina	Nglu
D-α-metilprolina	Dmpro	N-(carboximetil)glicina	Nasp
D-α-metilserina	Dmser	N-ciclobutilglicina	Ncbut
D-α-metiltreonina	Dmthr	N-cicloheptilglicina	Nchep
D-α-metiltriptofano	Dmtrp	N-Ciclohexilglicina	Nchex
D-α-metiltirosina	Dmty	N-Ciclodecilglicina	Ncdec
D-α-metilvalina	Omval	N-Ciclododeclglicina	Ncdod
D-α-metilalnina	Dnmala	N-Ciclooctilglicina	Ncoct
D-α-metilarginina	Dnmarg	N-Ciclopropilglicina	Ncpro
D-α-metilasparagina	Onmasn	N-Cicloundecilglicina	Ncund
D-α-metilaspartato	Onmasp	N-(2,2-difeniletil)glicina	Nbhm
D-α-metilcisteína	Dnmcys	N-(3,3-difenilpropil)glicina	Nbhe
D-N-metileucina	Dnmleu	N-(3-indoliletil) glicina	Nhtrp
D-N-metilisina	Dnmlys	N-metil-γ-aminobutirato	Nmgabu
N-metilciclohexilalanina	Nmchexa	D-N-metilmetionina	Dnmmet
D-N-metilomitina	Dnmorn	N-metilciclopentilalanina	Nmcpen
N-metilglicina	Nala	D-N-metilfenilalanina	Dnmphe
N-metilaminoisobutirato	Nmaib	D-N-metilprolina	Dnmpro
N-(1-metilpropil)glicina	Nile	D-N-metilserina	Dnmser
N-(2-metilpropil)glicina	Nile	D-N-metilserina	Dnmser
N-(2-metilpropil)glicina	Nleu	D-N-rnetiltreonina	DNmthr
D-N-metiltriptofano	Dnmtrp	N-(1-metiletil)glicina	Nva
D-N-metiltirosina	DNmtyr	N-metila-naftilalanina	Nmanap
D-N-metilvalina	Dnmval	N-metilpenici llamine	Nmpen
Ácido γ-aminobutírico	Gabu	N-(p-hidroxifenil)glicina	Nhtyr
L-t-butilglicina	Tbug	N-(tiometil)glicina	Ncys
L-etilglicina	Etg	Penicilamina	Pen
L-homofenilalanina	Hphe	L-α-metilalanina	Mala
L-α-metilarginina	Marg	L-α-metilasparagina	Masn
L-α-metilaspartato	Masp	L-α-metil-t-butilglicina	Mtbug
L-α-metilcisteína	Mcys	L-metiletilglicina	Metg
L-α-metilglutamina	Mgln	L-α-metilglutamato	Mglu
L-α-metilhistidina	Mhis	L-α-metilhomo fenilalanina	Mhphe
L-α-metilisoleucina	Mile	N-(2-metiltioetil)glicina	Nmet
D-N-metilglutamina	Dnmgln	N-(3-guanidinopropil)glicina	Narg
D-N-metilglutamato	Dnmglu	N-(1-hidroxietil)glicina	Nthr
D-N-metilhistidina	Dnmhis	N-(hidroxietil)glicina	Nser
D-N-metilisoleucina	Dnmile	N-(imidazoliletil)glicina	Nhis
D-N-metileucina	Dnmleu	N-(3-indoliletil)glicina	Nhtrp
D-N-metilisina	Dnmlys	N-metil-γ-aminobutirato	Nmgabu
N-metilciclohexilalanina	Nmchexa	D-N-metilmetionina	Dnmmet
D-N-metilornitina	Dnmom	N-metilciclopentilalanina	Nmcpen
N-metilglicina	Nala	D-N-metilfenilalanina	Dnmphe
N-metilaminoisobutirato	Nmaib	D-N-metilprolina	Dnmpro
N-(1-metilpropil)glicina	Nile	D-N-metilserina	Dnmser
N-(2-metilpropil)glicina	Nleu	D-N-metiltreonina	DNmthr
D-N-metiltriptofano	Dnmtrp	N-(1-metiletil)glicina	Nval
D-N-metiltirosina	Dnmtyr	N-metila-naftilalanina	Nmanap
D-N-metilvalina	Dnmval	N-metilpenicilamina	Nmpen
Ácido γ-aminobutírico	Gabu	N-(p-hidroxifenil)glicina	Nhtyr
L-t-butilglicina	Tbug	N-(tiometil)glicina	Ncys
L-etilglicina	Etg	Penicilamina	Pen
L-homofenilalanina	Hphe	L-α-metilalanina	Mala
L-α-metilarginina	Marg	L-α-metilasparagina	Masn
L-α-metilaspartato	Masp	L-α-metil-t-butilglicina	Mtbug
L-α-metilcisteína	Mcys	L-metiletilglicina	Metg
L-α-metilglutamina	Mgln	L-α-metilglutamato	Mglu
L-α-metilhistidina	Mhis	L-α-metilhomofenilalanina	Mhphe
L-α-metilisoleucina	Mile	N-(2-metiltioetil)glicina	Nmet

Aminoácido no convencional	Código	Aminoácido no convencional	Código
L-α-metilleucina	Mleu	L-α-metilisina	Mlys
L-α-metilmetionina	Mmet	L-α-metilnorleucina	Mnle
L-α-metilnorvalina	Mnva	L-α-metilomitina	Morn
L-α-metilfenilalanina	Mphe	L-α-metilprolina	Mpro
L-α-metilserina	Mser	L-α-metiltreonina	Mthr
L-α-metilvalina	Mtrp	L-α-metiltirosina	Mtyr
L-α-metileucina	Mval Nnbhm	L-N-metilhomofenilalanina	Nmhphe
N-(N-(2,2-difeniletil)		N-(N-(3,3-difenilpropil)	
carbamilmetil-glicina	Nnbhm	Carbamilmetil(1) glicina	Nnbhe
1-carboxi-1-(2,2-difenil etilamino)ciclopropano	Nmbc		

Se debe notar que aunque se hace referencia arriba a un biopolímero, de hecho la presente invención también puede ser opcionalmente utilizada con otros tipos de polímeros así como también descritos con mayor detalle adelante en la sección 2.

EJEMPLO 1

Metodo D&C

El método sintético soporta la síntesis de moléculas de ADN largas o librerías de ADN combinatorias de una manera eficiente y opcionalmente automatizada de manera completa.

Descripción del Proceso de Síntesis de ADN-D&C

El método de síntesis de ADN- D&C permite la síntesis in vitro de moléculas de ADN largas o librerías de ADN combinatorias, utilizando la aproximación Divide y Conquistarás (D&C).

El D&C Es un paradigma de diseño de algoritmo que trabaja al descomponer recursivamente un problema en dos o más subproblemas del mismo tipo. Los subproblemas se resuelven opcionalmente y preferiblemente de manera independiente y sus soluciones son entonces combinadas para dar una solución al problema original. Esta técnica se ha aplicado en muchos problemas bien conocidos en la ciencia de cómputo. La aproximación D&C suministra al menos dos beneficios potenciales. Primero, se puede suministrar una aproximación simple para resolver los problemas de dificultad conceptual al reducir el tamaño del problema recursivamente hasta que se alcanza un caso de base trivial. Segundo, en algunos casos se puede reducir sustancialmente el coste de cómputo.

El D&C es opcional y preferiblemente aplicado a síntesis de ADN largo (o síntesis de las secuencias largas) de la siguiente manera. La división de las secuencias de ADN largas se hace in sílico. El programa de ordenador procesa la secuencia y diseña el mejor protocolo de síntesis. El ensamblaje de la secuencia se hace in vitro. Las secuencias cortas (hasta de 85 nt) se sintetizan directamente, mientras que las secuencias más largas se sintetizan utilizando dos secuencias más cortas, denominadas progenitores, que se producen recursivamente (Figura 1). Los dos progenitores son complementarios en los extremos de tal manera que la secuencia larga se puede sintetizar de ellos por vía de polimerización, opcionalmente en una reacción de alargamiento (aunque esta reacción puede opcionalmente no ser una reacción de alargamiento).

El diseño de un protocolo de síntesis para una secuencia larga consiste de un árbol binario, en el cual cada nodo representa una secuencia intermedia. Los nodos internos se crean en las reacciones de alargamiento de sus nodos hijos y las hojas se sintetizan directamente. Después de cada alargamiento solo la cadena de ADN pasa al siguiente nivel en el árbol (Figura 2). De acuerdo con algunas realizaciones, la selección de cadena se efectuó de acuerdo con la fosforilación de una de las cadenas de ADN, aunque por supuesto a otros tipos de selección de cadena se pueden efectuar de manera opcional. Preferiblemente, antes de cada etapa de alargamiento, una de las cadenas de ADN es fosforilada mediante Polinucleótido de Quinasa (PNK). El producto alargado es tratado con Exonucleasa Lambda, que digiere solamente la cadena fosforilada y deja intacta la cadena no fosforilada (Figura 3).

La Figura 4 muestra un proceso recursivo de ejemplo básico para la síntesis de ADN - D&C

1. Fosforilación de una de las dos cadenas de entrada

2. Hibridación de las dos cadenas

3. Alargamiento

4. Producción de la salida de ssADN

La Figura describe el “corazón” del proceso de síntesis de ADN - D&C y representa la solución bioquímica para el procedimiento recursivo. La entrada al procedimiento se compone de dos ADN de cadena simple que producirán una salida de ADN de cadena simple larga.

Se debe notar que la selección de cadena se puede efectuar de numerosas maneras; Para los propósitos de discusión solamente y sin desear estar limitados, dos métodos ilustrativos de ejemplo se suministran como ejemplos no limitantes. Adicionalmente, el protocolo de construcción recursivo/jerárquico también puede potencialmente efectuarse sin la selección de cadena.

Preferiblemente los cálculos anteriores se efectúan mediante un programa de ordenador que determina el o los protocolos de síntesis óptima de acuerdo al proceso esbozado en la Figura 4B. El programa se basa en un modelo matemático, en el cual cada protocolo recibe una calificación que corresponde a su coste de producción. El programa encuentra el protocolo de coste mínimo para generar la cantidad q de la molécula M; en resumen: min(C(M,q)).

Resultados Experimentales de la Síntesis de ADN- D&C

Algunas realizaciones preferidas de la síntesis de ADN- D&C se describieron anteriormente. Se consideraron dos opciones para efectuar la síntesis de ADN- D&C: primero, elaborar el alargamiento y la separación de la cadena sobre soporte sólido, segundo alargamiento y selección de la cadena enzimática en solución.

En cada etapa de alargamiento una de las cadenas de ADN es biotinilada. La marcación con biotina se efectúa durante la síntesis oligo. Después del alargamiento todos los dsADN son pescados por perlas magnéticas recubiertas con estreptavidina. Las cadenas se separan mediante 0.1 M de NaOH y las cadenas biotiniladas se calientan y se desconectan de las perlas. Una de las cadenas de ADN pasa al siguiente nivel en el árbol de acuerdo con el árbol de síntesis.

Se examinó la eficiencia del ADN biotinilado que une las perlas magnéticas con estreptavidina y la separación de cadena. La eficiencia de unión está influenciada por las impurezas del oligonucleótido. No todos los oligos marcados con biotina después de la síntesis llevan biotina. La purificación del oligonucleótido (oligo) ayuda a solucionar este problema.

Con el fin de ensayar la eficiencia de unión de los oligos biotinilados a las perlas utilizadas así como también los oligos no biotinilados de la misma longitud (25nt), los oligos fueron marcados con P^{32}. La unión específica y no específica se ensayó en dos experimentos. En uno 25 pmol de oligos biotinilados se agregaron en cada momento a un tubo que contiene 10µg de perlas (Dynabeads MyOne Streptavidin) para revisar la unión específica de las perlas (Fig. 5A). El otro experimento se hizo como anteriormente excepto que se agregó el oligo agregado se desbiotiniló lo que probó la unión no específica de las perlas (Fig. 5B). La capacidad de unión se midió al contar la radiación radioactiva de los oligos en solución y la desunión y acoplado a las perlas. Los oligos de 25 pmol fueron incubados en un amortiguador IX B&W (5mM de Tris-HCl) (pH 7.5); 0.5 mM de EDTA; 1,0 M de NaCl) con Dynabeads MyOne™ Streptavidin (Dynal) 10 µg 15 min. Las perlas se separaron por magnetos, luego las perlas fueron lavadas con un amortiguador IX B&W. El nivel de radiación de muestras de 1 µl se midió en un sobrenadante no unido y las soluciones de lavado. Luego las perlas fueron incubadas con la siguiente porción de oligos de 25 pmol.

Del trabajo experimental se notó que el ADN biotinilado más largo había reducido su capacidad de unión.

De acuerdo con el protocolo los dsADN alargados se pescaron utilizando perlas magnéticas con estreptavidina. La cadena no biotinilada fue liberada utilizando 0.1 M NaOH neutralizada mediante 0.1 M de HCl. La cadena biotinilada se liberó de las perlas mediante ebullición. La cadena biotinilada fue marcada radioactivamente después de liberarse para visualización (FIG. 6).

El alargamiento de los oligonucleótidos (cada uno 50 pmol) mediante 10, 20 y 30 unidades de fragmento Klenow se hizo en 30 µl de amortiguador 1X (10 mM de Tris-HCl, pH 7.5, 5 mM de $MgCl_2$, 7.5 mM de ditiotreitol; 200 mM de dNTP) durante 0, 60, 120, y 180 min a 37ºC. La Figura 7 muestra que el alargamiento ocurre en dos fases; en la medida en que cada cadena se alarga separadamente y el alargamiento consume tiempo, cada dúplex hibridado se alarga en dos direcciones. Sin embargo, esto no ocurre simultáneamente y la velocidad de alargamiento es dependiente de la eficiencia de alargamiento de la ADN polimeraza. Se observó que el fragmento Klenow no es el mejor candidato para el alargamiento porque este tiene eficiencia de alargamiento baja.

Como una alternativa al Fragmento Klenow, se consideró la Versión Sequenasa 2.0. Esta enzima tiene una eficiencia de elongación de alargamiento alta y es muy precisa. Esta se desarrolló sobre la base de ADN Polimeraza T7. El alargamiento de los nucleótidos (50 pmol cada uno) mediante 20 unidades de la Versión Sequenasa 2.0 se hizo en 30 µl de amortiguador 1X (40 mM de Tris-HCl, pH 7.5, 20 mM de $MgCl_2$, 50 mM de NaCl, 5 mM de ditiotreitol; 200 mM de dNTP) durante 30 min a 37ºC. La Figura 8 muestra el alargamiento de dos pares de oligos mediante la versión Sequenase 2.0. Después de un afinamiento cuidadoso de las condiciones de reacción se encontró que para el alargamiento de 30 pmol se pueden utilizar 1.65 unidades de enzima, y la reacción ocurre en 15 min. (Fig. 9).

Durante el trabajo sobre la mejora de la separación de la cadena mecánica y la selección sobre el soporte sólido, se consideró un método simple alternativo, basado en la marcación de cadena selectiva enzimática y la selección enzimática eficiente en solución. La principal idea fue que antes de cada etapa de alargamiento una de las cadenas de ADN se fosforila mediante Polinucleótido de Quinasa (PNK). El producto alargado se trató con Exonucleasa Lambda. Esta Exonulceasa digiere solamente la cadena fosforilada y deja intacta la cadena no fosforilada.

La siguiente figura demuestra la actividad de Exonucleasa de la Exonucleasa Lambda. La Exonucleasa Lambda es muy específica y tiene una eficiencia de exonucleasa muy alta (Fig. 10). La fosforilación de ADN enzimático: de 2000 pmol de ADN fueron hechos en 120 µl de regulador PNK 1X (Tris-HCl 70 mM pH 7.6 en 25° C; $MgCl_2$ 10 mM ditiotreitol 5 mM) que contiene ATP 1 mM y 80u de polinucleótido Quinasa T4 (PNK) (NEB) durante 60 min a 37° C. El ADN que hibrida se hace en 20 µl de Tris-HCl 40 mM pH 7.5; MgCl 20 mM, NaCl 50 mM al enfriarse lentamente desde 96° C hasta 25° C en un índice de 0.04° C/sec. Se realiza la digestión de Lambda Exonucleasa de 60 pmol de dsADN en la que una hebra se fosforila en 50 µl de glicina-KOH 67 mM (pH 9.4; $MgCl_2$ 5 mM; 0.01% Triton X-100 utilizando 10 u de Lambda Exonucleasa durante 45 min a 37° C. (Epicentre).

Se determinan las condiciones adecuadas para todas las reacciones durante una etapa de recursividad en la síntesis de DC-ADN. Bioquímicamente, el proceso recursivo se puede describir opcionalmente como sigue:

1. Fosforilación de uno o dos ssADN que entran en la recursividad.

2. Hibridación de estos dos ssADN.

3. Alargamiento

4. Selección de soporte

El diseño del protocolo para el concepto anterior está de acuerdo con el software desarrollado por los inventores. El programa elige el mejor y más barato protocolo (Figura 11) de acuerdo con la función de puntuación teniendo en cuenta muchos parámetros tales como cantidades de síntesis de oligos y su coste, cantidades de productos intermedios, rendimiento de hibridación del heterodímero deseado en cada etapa en el protocolo, tasa de éxito en cada etapa y coste de reacción.

Para la realización de este protocolo, se sintetizan 5 oligos. Los nombres de oligos y productos intermedios son: S, sentido; A, antisentido; los números muestran la ubicación de secuencia en el producto final.

S0_64 (65 nt)
ACCGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTA

S29_112 (84 nt)
TCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAAT

A90_173 (84 nt)
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTT

S153_231 (79 nt)
TCGTTCATCCATAGTTGCCTGACTGCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCT

A217_267 (51 nt)
GAAACGATTCCGGTCGAGACCTTGCGGTATCATTGCAGCACTGGGGCCAGA

La siguiente tabla muestra los parámetros De todas las reacciones de hibridación de acuerdo con el algoritmo vecino más cercano[vi].

Heterodidímero	Superposición (bp)	T_M (° C)	$\Delta G°_{37.0}$(kcal/mol)	ΔH°(kcal/mol)	ΔS (eu)
S29 112 A90 173	23	68.7	-24.92	-172.4	-475.51
S153 213 A217_267	15	70.0	-19.98	-114.8	-305.72
S29_173 A 153 267	21	67.7	-23.93	-116.6	-460.00
SO 64 29 267	36	67.2	-32.92	-270.6	-766.34

La descripción adelante se relaciona con un protocolo de iteración.

Protocolo de una etapa

1. Fosforilación de hebra de ssADN(-P) de acuerdo con el protocolo

Nombre	Concentración inicial		Dilución X	Volumen [ul]	Concentración final	
Regulador A	10	X	10	5	1	X
ATP	100	mM	100	0.5	1	mM
DTT	100	mM	20	2.5	5	mM
Enzima A-PNK	10	u/ul	50	1	0.2	u/ul
H2O	-		-	5	-	
ssADN(-P)	3.0	Pmol/ul	1.39	36	2	pmol/ul
		Volumen total		50		

2. Incubación durante 30 min a 37° C

3. Inactivación durante 10 min a 75° C

4. Purificación de ssADN por filtración en gel

5. Hibridación de hebras fosforiladas y no fosforiladas

Hibridación Volumen de reacción constnte (110 ul), vol de ADN 36.50 ul

Nombre	Concentración inicial	Dilución X		Volumen [ul]	Concentración final	
Regulador B	5	X	5	22	1	X
H2O	-		-	2	-	
ssADN	3.0	pmol/ul	3	36	1.0	pmol/ul
ssADN (+P)	2	pmol/ul	2.2	50	1.0	pmol/ul
		Volumen total		110		

6. Hibridación por enfriamiento lento 95° C → 25° C 0.04° C/seg

7. Alargamiento de hebras hibridadas

Alargamiento Volumen de reacción constante (125 ul)

Nombre	Concentración inicial		Dilución X	Volumen [ul]	Concentración final	
Regulador B	5	X	5.0	3.8	1	X
dNTP	25	mM cada uno	85.91	1.5	0.3	mM
DTT	100	mM	20.0	6	5	mM
Enzima B-Seq II	1.63	u/ul	76.9	2.52	0.02	u/ul
H2O	-		-	4.97		
ADN híbrido	0.98	pmol/ul	1.18	106		pmol/ul
		Volumen total		125		

8. Incubación durante 30 min a 37° C

9. Inactivación durante 10 minutos a 75° C y enfriamiento lento 80° C → 25° C 0.1° C/seg

10. Purificación de dsADN mediante el equipo de purificación de PCR. Elución en 30 µl

11. Degradación de la hebra fosforilada

Exonucleación cantidad de enzima constante (20 u), volumen de ADN (30 ul)

Nombre	Concentración inicial		Dilución X	Volumen [ul]	Concentración final	
Regulador C	10	X	10	3.8	1	X
DTT	100	mM	20	1.9	5	mM
Enzima C Exo. Lam.	10	u/ul	19.0	2	0.53	u/ul
H2O	-		-	0.3		
ADN alargado	3.4	pmol/ul	1.3	30	2.65	pmol/ul
		Volumen total		38		

12. Incubación durante 30 min a 37° C

13. Inactivación durante 10 min a 75° C

14. Purificación de ssADN mediante filtración en gel

Las Muestras de ADN están listas para la siguiente etapa de iteración.

La Figura 12 muestra el progreso de los dos niveles de síntesis. En este experimento el oligo S29_112 se marca por Cy5 en 5'. Todos los otros oligos no se marcaron. Se puede observar que al final de el alargamiento, el producto dsADN tiene la longitud esperada.

Los productos se probaron después de cada etapa de síntesis mediante electroforesis capilar. Se utilizaron dos marcadores de etiquetas fluorescentes VIC y Cy5. Las moléculas de ADN se alargan de etapa a etapa. Las diferencias en el tamaño después de cada etapa entre los dos tintes son el resultado del cambio provocado por los tintes (Figura 13).

Se secuenciaron los productos en ambas direcciones después de cada etapa de alargamiento. Se encontraron algunos errores. Una sustitución y una supresión se muestran en la Figura 14. Después del análisis de la posibilidad de fuentes de estos errores, se encontró que parte de los errores resulta durante la síntesis de los oligonucleótidos y parte resulta durante el alargamiento. Estos problemas se abordan adelante.

En el proceso de síntesis de dos puntos que pueden provocar acumulación de error son: hibridación y alargamiento de hebra. Durante la hibridación no todas las moléculas encuentran sus parejas debido a algunos errores de medición de concentraciones de ADN y pipeteo. En este caso, algunas moléculas de ssADN crean autobucles. Esta estructura secundaria tiene alargamiento potencial. A veces durante el alargamiento no todas las dos hebras hibridadas se alargan, en algunos casos sólo se alarga una hebra. Para eliminar la acumulación de errores, se consideraron los métodos para selección de ADN específica de acuerdo a sus propiedades. En la Figura 15 se muestra la limpieza postalargamiento del producto alargado dsADN a partir del exceso de ssADN mediante exonucleasa I (30 min a 37° C).

La exonucleasa I tiene actividad exonucleasa específica de ssADN 3'→5'. Esta enzima tiene una actividad de exonucleasa muy alta y específica y puede trabajar directamente con el regulador de alargamiento. Sin embargo a partir de los experimentos se determinó que la enzima no podría progresar en estructuras secundarias fuertes. Las muestras se calentaron para abrir la estructura secundaria, pero se inactivó la exonucleasa I a 65° C. La otra solución para la limpieza de ADN específica es utilizar nucleasa Mung Bean. Esta tiene actividad nucleasa de ssADN 3'→5', 5'→3' y actividad endonucleasa de ssADN. Esta enzima tiene propiedades termoestables. La Figura 16 muestra su actividad después de 30 minutos en 80° C. La nucleasa Mung Bean limpia el exceso de ssADN y un ADN de hebra alargada. Las condiciones de reacción son acetato de sodio 30 mM (pH 4.6), cloruro de sodio 50 mM, acetato de zinc 1 mM y 0.01% de Triton X-100, 10 unidades nucleasa Mung Bean (Epicentre) y 10-70 pmol de ADN después de reacción alargamiento. Esta es una herramienta muy importante para la librería de "ADN basura", ya que es capaz de eliminar las moléculas que no deben estar presentes después del final de cada ciclo de síntesis.

Prueba de concepto con secuencias naturales.

La prueba de concepto se lleva a cabo al utilizar una secuencia conocida tomada del plásmido pBK. Se utiliza una secuencia de 737 pb para crear un protocolo de síntesis D&C-ADN para con la longitud mínima de aproximadamente 200 pb. El protocolo se presenta en la figura 17.

Las secuencias de las hojas del árbol de protocolo D&C se obtuvieron de cuatro reacciones de PCR individuales (en 50 µl de Tris-HCl 10 mM, pH 8.3, KCl 50 mM, $MgCl_2$ 1.1 mM, 0,01% de Gelatina, dNTPs 200 mM, 0.5 µM de cada cebador y plásmido pGEM 0.04 nM; 25 ciclos). En cada reacción de PCR uno de los cebadores se fosforiló y antes del proceso de síntesis de alargamiento la hebra fosforilada fue digerida por exonucleasa lambda I. Para obtener el producto final (alargamiento 3), se aplicó un procedimiento de síntesis recursiva de dos niveles. En el primer nivel de la síntesis, se obtuvieron los alargamientos 1 y 2. El alargamiento 1 se obtuvo a partir PCR 1 y 2 y el alargamiento 2 se obtuvo a partir de PCR 3 y 4. En el segundo nivel de síntesis, se obtuvo un producto que corresponde al alargamiento 3 a partir de los alargamientos 1 y 2. Las secuencias de los progenitores y productos obtenidos en cada etapa se confirman mediante secuenciación. Se han visto coincidencias perfectas con las secuencias esperadas.

La Figura 18 muestra la secuencia completa de la estructura de ADN 737 pb que fue producida por los presentes inventores, denominado Amp_PCR1-PCR4; la Elon3F-Prm1F representa la secuencia del producto final desde su inicio; la E3-F representa la secuencia del producto final a partir de su centro.

Con el fin de resolver el problema de las tasas de error durante el alargamiento mientras se trabaja con oligonucleótidos sintéticos, se realizó PCR en estos oligonucleótidos utilizando dos cebadores para cada oligonucleótido (uno de ellos fosforilado). Los resultados obtenidos mostraron que es posible secuenciar el producto de PCR de los oligonucleótidos (en contraste con el oligonucleótido en sí mismo que no se puede secuenciar). Con el fin de evaluar la tasa de errores se clonaron y secuenciaron los productos de PCR de oligonucleótidos. La Figura 19 muestra la secuencia del producto de PCR de un oligonucleótido. Los PCR de oligonucleótidos se clonaron en vector fácil pGEM-T. Se secuenciaron pocos clones positivos. Las secuencias obtenidas fueron todas idénticas y no revelaron emparejamientos erróneos (Figura 20).

Se secuenciaron los productos alargados obtenidos a partir de los PCR de oligonucleótidos y se encontró que eran correctos (Figura 21). Adicionalmente, los productos alargados se clonaron y secuenciaron. Las secuencias clonadas todas fueron correctas e idénticas a las esperadas (Figura 22).

Todos los datos recibidos hasta el momento han abierto la ruta para la realización de este método en la plataforma completamente automática y empezar a trabajar en la síntesis de librerías combinatorias definidas específicamente. Actualmente se está utilizando un robot de manipulación de líquidos de laboratorio (Tecan Freedom 2000). Se desarrolló un programa que traduce el protocolo de síntesis D&C-DNA para el lenguaje de escritura del robot. El robot realiza todas las operaciones de manejo de líquidos, incubación de reacción, y recolección de muestras para control de calidad.

Purificación de Oligo nucleótidos enzimática

La presente descripción proporciona un método y un sistema para opcional y preferiblemente utilizar oligonucleótidos sintéticos (también denominados aquí como “oligos”) para la construcción de moléculas mucho más largas. Dichas moléculas tienen propiedades y características que se deben considerar para la implementación de la presente invención.

Por ejemplo, durante la síntesis de oligo, cada etapa de acoplamiento tiene una eficiencia del 98 al 99%. El 1-2% de oligos que no hacen pareja en cada etapa se inactivan y permanecen en el producto de síntesis final; estos son llamados productos n-menos. En oligos largos, la cantidad de material n puede ser significativa. Como un ejemplo, consideremos un 75-mer sintetizado con 99% de eficiencia en cada etapa. La cantidad de producto de longitud completa será (.99)75 = 47%. El 53% restante del producto final sintetizado consiste de productos n. Debido a que la síntesis procede de 3’ a 5’, los productos n consisten de supresiones del extremo 5’. Adicionalmente los métodos para generar oligonucleótidos tienen algún nivel de errores. Existen algunos métodos básicos de purificación de oligonucleótidos, que incluyen pero no se limitan a purificación de cartucho de fase inversa, purificación por HPLC, purificación PAGE.

El principio de purificación de cartucho de fase inversa se basa en la selección de producto de longitud completa con el grupo DMT. El n-1 oligo sin el grupo DMT no se seleccionará cuando el oligo se pasa a través de la columna.

La purificación por HPLC elimina las secuencias de oligo más truncados. La HPLC de intercambio iónico se utiliza para separar las secuencias de oligo sobre la base de carga.

La PAGE separa los oligos sobre la base de carga y peso molecular. La purificación PAGE se lleva a cabo en un gel de poliacrilamida desnaturalizante que contiene urea. El oligo se detecta a través de sombreado UV. El producto de longitud completa se escinde y se eluye a partir del gel.

Estas técnicas purifican los oligos de productos n pero no resuelven el problema de los errores, no son automatizadas.

Resultados experimentales - Corrección de error

Como se señaló anteriormente, un problema de cualquier tipo de síntesis de oligonucleótidos es la ocurrencia de errores. Todos los métodos sintéticos producen oligonucleótidos con errores en la secuencia de varios tipos, que incluyen nucleótidos con emparejamiento erróneo, nucleótidos incorrectos, nucleótidos agregados, truncamiento prematuro y similares. De acuerdo con las realizaciones preferidas, se proporciona un método para reducir, si no eliminar dichos errores, preferiblemente a través de un proceso de eliminación de error iterativo. Dicho proceso se realiza preferiblemente como parte de un proceso recursivo de la construcción de oligonucleótidos. La construcción recursiva con eliminación de errores iterativa mejora en los métodos previos a la síntesis de ADN, al permitir la construcción rápida, completamente automatizada y sin errores de moléculas de ADN largas. Dicho método también puede opcionalmente ser utilizado para combinar segmentos de ADN sintéticos y naturales. Adicionalmente, dicho método permite opcionalmente diseño, síntesis y corrección de errores eficiente de las librerías combinatorias de ADN con componentes compartidos y variantes y, opcionalmente, permite la síntesis de los componentes básicos de los genomas. Por lo tanto, la presente descripción proporciona una nueva base para el estudio y diseño de moléculas y organismos biológicos sintéticos.

De acuerdo con algunas realizaciones, la construcción recursiva se realiza al combinar pares de elementos más pequeños en elementos cada vez más grandes hasta que se construye un elemento suficientemente grande, como se muestra en la Figura 25A. Como se muestra, la superposición de fragmentos de ADN de cualquier fuente se puede utilizar como entrada para la construcción recursiva (superior). Una mezcla de fragmentos naturales y sintéticos se muestra con el propósito de descripción y sin ninguna intención de ser limitante, ya que opcionalmente sólo se pueden utilizar fragmentos sintéticos o naturales. Los fragmentos de ssADN se combinan de forma recursiva para formar la molécula objetivo (centro), que se clona y secuencia. Si existe un clon sin errores, entonces éste clon forma la salida del proceso y termina el proceso (parte inferior). De lo contrario, un conjunto de fragmentos sin errores de los clones que constituyen un corte mínimo del árbol de construcción recursiva se utilizan como entradas para la siguiente iteración del proceso de construcción, que normalmente es más corto y produce una salida sin o por lo menos, pocos errores que la iteración previa.

Un ejemplo de realización pero preferido de las etapas internas del proceso de construcción recursiva se muestra en la Figura 25B. Para habilitar la recursividad, la operación de composición se diseña de tal manera que su salida es del mismo tipo que sus dos entradas. Por ejemplo, una reacción composición que toma dos fragmentos de ssADN superpuestos como entrada y produce la molécula de ssADN alargada correspondiente como salida no se puede utilizar de forma recursiva, debido a que la molécula de ssADN de salida no es una entrada adecuada para la siguiente reacción de composición. Por la misma razón, una reacción que toma dos moléculas de dsADN superpuestas como entrada y produce como salida una mezcla de moléculas de entrada y algunas moléculas de dsADN alargadas no se puede utilizar de forma recursiva, ya que, por definición, dicha mezcla no es una entrada legal a la reacción de composición. Para abordar este requisito básico de construcción recursiva, se utiliza una reacción de composición novedosa que toma dos fragmentos de ssADN superpuestos como entrada y produce una molécula de ssADN alargada como salida, mostrada en la Figura 25B.

Como se muestra, se hibridan dos fragmentos de ssADN superpuestos de entrada y se ceban entre sí por una reacción de alargamiento para formar un fragmento de dsADN, que se amplifica por PCR con uno de los dos cebadores fosforilados en su extremo 5'. La hebra de PCR marcada con fosfato luego se degrada con exonucleasa Lambda, produciendo un fragmento de ssADN alargado como salida. El proceso de construcción recursiva luego continúa como se describe.

La salida de un proceso de construcción recursiva puede contener fallas o errores, que opcionalmente pueden haber estado presentes ya en los elementos de entrada y/o introducidos por el proceso de construcción en sí mismo. Este problema se resuelve a través de la explotación de la naturaleza específica del ADN. La molécula construida (construida de acuerdo con el proceso de la Figura 25) se clona y se secuencia, como en otros protocolos de construcción de ADN conocidos en la técnica. Si uno de los clones es sin errores, es la salida del proceso. De lo contrario, los fragmentos sin errores de los clones preferiblemente se identifican y preferiblemente se utilizan para calcular un corte mínimo como se describió anteriormente a través del árbol de construcción recursiva. El corte mínimo es un conjunto de fragmentos sin errores de longitud máxima de los clones, con límites que coinciden con límites de fragmentos del árbol de construcción recursiva inicial, que puede servir como entradas mejores (más largo y más precisa) a la siguiente iteración del proceso de construcción. Este intercambio de límites asegura que los fragmentos de corte mínimo pueden ser "levantados" de clones con PCR utilizando los mismos cebadores que en la etapa de composición correspondiente (Figura 25B) del proceso de construcción recursiva.

A menos que se indique lo contrario, los ejemplos adelante se relacionan con métodos de construcción recursiva como se describe en el presente documento, por ejemplo con respecto a los protocolos D&C y otros protocolos como se describe aquí.

Ejemplo - Síntesis de GFP

El proceso de construcción recursiva de una molécula particular, el gen GFP 783nt-largo y su corrección de errores iterativa se muestran en la Figura 26. Los materiales y métodos más detallados se dan a continuación.

La secuencia GFP objetivo se dividió in silico de forma recursiva en secuencias oligo básicas utilizando D&C (Figura 26A). La secuencia objetivo se carga un algoritmo que genera protocolo de construcción que divide de forma recursiva la secuencia objetivo en oligos superpuestos básicos utilizando D&C, un proceso como se describe aquí. El algoritmo encuentra una solución óptima para la construcción de ADN de acuerdo con las limitaciones termodinámicas (afinidad, especificidad etc.), topología del árbol de construcción y una función de coste (número de oligos, longitud de superposición etc.). Cada nodo en el árbol de la secuencia GFP dividida representa dos secuencias de ADN superpuestas, con superposiciones delimitadas en el centro de cada nodo.

Los bloques de construcción de ADN básicos en la división in silico se elaboran (con errores) mediante síntesis de oligonucleótidos convencionales y se utilizan como entradas para la construcción recursiva in vitro (Figura 26B). Si cualquier fragmento de la secuencia objetivo está disponible como ADN natural (por ejemplo en un plásmido), entonces el algoritmo puede tener en cuenta esta información y utilizar este fragmento natural en la construcción en lugar de sintetizarlo a partir de oligonucleótidos básicos (véanse Figuras 25A, 26B).

Se realiza la construcción recursiva, utilizando estos oligonucleótidos, hasta que se forman moléculas GFP de longitud completa (esperando que sea errónea) (véase Figuras 27 a 29 para los resultados, como se describe a continuación). El algoritmo de la Figura 26B también genera un programa de control de robot de manipulación de líquidos, utilizando un lenguaje desarrollado por los inventores, que ejecuta el protocolo de construcción (véase Figura 25B). El programa ordena al robot construir automáticamente el Gen GFP utilizando reactivos estándar fuera de plataforma. El protocolo realmente utilizado en los experimentos incluye procedimientos que se realizaron manualmente debido a las limitaciones de los equipos de laboratorio, pero que podrían opcionalmente ser realizados automáticamente (véase materiales y métodos). El azul claro y oscuro corresponde las hebras sentido y antisentido de ADN combinado de forma recursiva. Las moléculas de GFP objetivo erróneas se clonan, secuencian y se identifican los errores (marcados con rojo) (Figura 26C). Un corte mínimo secuencias sin errores (no desvanecido, por debajo de la línea de corte mínima) del proceso de construcción original se calculó in silico a partir de un conjunto de tres clones propensos a errores. Los fragmentos que constituyen este corte mínimo (paréntesis negros en clones) se amplifican a partir de los clones y se utilizan como entradas para reconstrucción recursiva in vitro de acuerdo con el corte mínimo. Esto da lugar a un clon GFP objetivo sin errores (Figura 26D).

La construcción recursiva de un fragmento de ADN 3Kb se lleva a cabo utilizando esta construcción de GFP sintética y ADN a partir de una fuente natural (plásmido bacteriano) como entradas para el proceso de construcción recursiva. Esto produjo una molécula 3 Kb objetivo sin errores (Figura 26E).

Materiales y Métodos –

Una descripción más detallada de los Materiales y Métodos se proporciona en esta sección.

Fosforilación:

300 pmol de ADN de terminal 5’ en una reacción de 50 µl que contiene Tris-HCl 70 mM, $MgCl_2$ 10 mM, ditiotreitol 7 mM, pH 7.6 a 37° C, ATP 1 mM, 10 unidades de polinucleótido quinasa T4 (NEB). La incubación es a 37° C durante 30 min, inactivación a 65° C durante 20 min.

Alargamiento (ABgene):

1 -5 pmol de ADN de terminal 5’ de cada progenitor en una reacción que contiene TAPS 25 mM pH 9.3 a 25° C, $MgCl_2$ 2 mM, KCl 50 mM, β-mercaptoetanol 1 mM, 200 µM de cada uno de dNTP, 4 unidades de polimerasa de ADN Thermo-Start (ABgene). El programa Ciclador Térmico es: Activación de enzima a 95° C 15 min, hibridación lenta 0.1° C/seg a partir de 95° C a 62° C, alargamiento a 72° C durante 10 min.

Alargamiento (BioLINE):

1-5 pmol de ADN de terminal 5’ de cada progenitor en una reacción que contiene Tris-HCl 60 mM pH 8.3 a 60 25° C, $(NH_4)_2SO_4$ 6 mM, KCl 10 mM, $MgSO_4$ 2 mM, 200 µM de cada uno de dNTP, 5 unidades de polimerasa de ADN AccuSure (BioLINE). El programa Ciclador Térmico es: activación de enzima a 95° C 10 min, hibridación lenta 0.1° C/seg a partir de 95° C a 62° C, alargamiento a 72° C durante 10 min.

PCR (ABgene):

1-0.1 fmol de plantilla, 10 pmol de cada cebador en una reacción de 25 µl que contiene TAPS 25 mM pH 9.3 a 25° C, $MgCl_2$ 2 mM, KCl 50 mM, β-mercaptoetanol 1 mM, 200 µM de cada uno de dNTP, 0.9 unidades de polimerasa de ADN Thermo-Start (ABgene). El programa Ciclador Térmico es: Activación de enzima a 95° C 15 min, C desnaturalización 95°, hibridación a Tm de los cebadores, Extensión 72° C 1 min por kb para ser amplificada 20 ciclos.

PCR (Bioline):

1-0.1 fmol de plantilla, 10 pmol de cada cebador en una reacción de 25 µl que contiene TAPS 25 mM pH 9.3 a 25° C, $MgCl_2$ 2 mM, KCl 50 mM, β-mercaptoetanol 1 mM, 200 µM de cada uno de dNTP, 1.9 unidades polimerasa de ADN AccuSure (Bioline). El Programa ciclador térmico es: activación de enzima a 95° C 10 min, desnaturalización 95° C, hibridación a Tm de los cebadores, Extensión 72° C 1,5 min por kb para ser amplificado 20 ciclos.

Exonucleasa Lambda:

1 hasta 5 pmol de ADN de terminal 5' fosforilado en una reacción que contiene TAPS 25 mM pH 9.3 a 25° C, $MgCl_2$ 2 mM, KCl 50 mM, β-mercaptoetanol 1 mM 1,4-ditiotreitol 5 mM, 5 unidades de exonucleasa Lambda (Epicentre). El Programa ciclador térmico es 37° C 15 min, 42° C2 min, inactivación de enzima a 70° C 10 min.

Purificación de ADN Automatizada:

La purificación de ADN automatizada se realizó también con equipo de purificación PCR QIAquik 96 de QIAGEN utilizando el protocolo estándar que fue adaptado para trabajar con Tecan fredom 200 y un colector de vacío.

Automatización Protocolo

El Protocolo automatizado se realizó como se describe aquí por un robot Tecan Freedom 2000.

Purificación de ADN:

Se realiza purificación de ADN manual con el equipo de purificación de PCR MinElute de QIAGEN utilizando procedimientos estándar.

Clonación

Los fragmentos se clonaron en el pGEM T Easy Vector System 1 de Promega utilizando procedimientos estándar. Los vectores que contienen fragmentos clonados se transformaron en células competentes JM109 de PROMEGAl utilizando procedimientos estándar.

Resultados –

La Figura 27 muestra el análisis de fragmentos C.E. después de reacciones de PCR durante la construcción de GFP. La Figura 27A muestra el protocolo de construcción completa de la molécula de GFP dividida en niveles de construcción. La Figura 27B muestra que al nivel 1 de construcción de GFP, cada C.E. contiene dos fragmentos que serán alargados juntos en el siguiente nivel. Las Figuras 27C y 27D muestran los resultados de los niveles de construcción de GFP 2 y 3, respectivamente. La Figura 27E muestra los resultados del nivel de construcción de GFP 4 y el PCR de control en pGFP-N1.

La primera iteración del proceso de construcción recursiva resultó en clones objetivo con una tasa de error de 1\160 (Figura 29A), lo que refleja la tasa de error conocida de sus bloques de construcción de oligonucleótidos. Dado que los errores están distribuidos al azar con una tasa conocida, el número mínimo de clones necesarios para obtener un corte mínimo sin errores se puede calcular con alta probabilidad como se describió anteriormente (> 95%).

El corte mínimo para la secuencia de GFP se calculó utilizando tres clones (Figura 26). Los fragmentos sin errores que constituyen el corte mínimo se levantan por PCR y se utilizan como entrada para una segunda iteración de construcción recursiva, lo que da lugar a un clon sin errores. A partir de los clones producidos en la primera iteración habría sido posible elegir sólo un par de clones para la reconstrucción, uno para cada mitad de la molécula objetivo, pero fueron elegidos tres, uno que contribuye sobre un medio y dos que contribuyen aproximadamente un cuarto cada uno de la molécula objetivo, para propósitos ilustrativos. Los clones producidos en la segunda iteración muestran una tasa de error de <1/5000.

El protocolo de construcción recursiva se puede utilizar para construcciones multi-kilobases y también se puede integrar el ADN de cualquier fuente y en cualquier etapa en el proceso de construcción. Para ilustrar esto, una molécula larga de ~3Kb se construye de forma recursiva a partir de la molécula de GFP sintética construida previamente 768nt-larga y dos fragmentos surgidos de un plásmido natural, uno de 700nt-largo y el otro 1700nt-largo (Figura 26E). Dado que los fragmentos eran de entrada sin errores, los clones de 3kb tampoco tenían errores, de tal manera que no fue necesaria corrección de errores adicional. Los resultados se muestran con respecto a la Figura 27 (la Figura 27A muestra los resultados después del primer nivel de construcción mientras que la Figura 27B muestra los resultados después del segundo nivel de construcción).

Ejemplo - Construcción de librería combinatoria

Los avances en la construcción de ADN de novo mantiene una gran promesa para la ingeniería de biopolímeros. Las librerías de ADN son una fuente importante de las cuales se pueden seleccionar las moléculas que codifican secuencias genéticas novedosas. Estas a su vez pueden codificar nuevos polímeros para su uso en medicina, investigación e industria. Numerosos métodos para construir grandes librerías, en su mayoría se han desarrollado mediante recombinación aleatoria y mutagénesis de ADN para la evolución dirigida. Por otro lado, en la práctica de computación intensiva del diseño racional de sólo un pequeño número de construcciones pre-especificadas no aleatorias, normalmente generadas por mutagénesis dirigida al sitio se investigan experimentalmente. El método de construcción de ADN de ejemplo de la presente invención, puede opcional y potencialmente extenderse para producir librerías de ADN combinatorias racionalmente diseñadas a gran escala con los miembros pre-especificados.

Un ejemplo de construcción de una librería se da aquí con respecto a la Figura 30. La librería contiene seis variantes predefinidas del gen p53, especificadas en la Figura 30A (se muestran los componentes compartidos (gris) y únicos (colores) de la librería de ADN combinatoria p53). Las Figuras 30B-D muestran un gráfico que representa el proceso de construcción de librería correspondiente al D&C con base en la división in silico.

El proceso se realiza como sigue. En primer lugar, se analizan secuencias de ADN de librería objetivo in silico y se identifican los segmentos únicos compartidos entre los miembros de la librería. Estos segmentos se dividen adicionalmente en oligonucleótidos superpuestos (Figura 30B). El algoritmo de división recursiva busca un protocolo de construcción de librería óptima con base en limitaciones químicas y una función de coste, con el fin de minimizar el número de componentes y las reacciones necesarias para construir la librería completa (Figura 30B).

Todos los seis genes p53 diferentes se construyeron de forma recursiva a partir de oligos básicos (Figura 30B superior), y las moléculas resultantes se clonaron y secuenciaron (parte inferior Figura 30B). Un corte mínimo sin errores de toda la librería se calcula a partir de sólo cuatro clones, y una segunda iteración del proceso utilizando fragmentos que constituyen este corte mínimo produce clones sin errores de los seis miembros de la librería. Vale la pena señalar que la tasa de error de los clones en la primera iteración fue, como se esperaba, 1\160, y que los fragmentos sin errores que totalizan 1000nt eran suficientes para generar en la segunda iteración una librería sin errores completa con aproximadamente 5200nt (Figuras 30C y D). Los clones producidos en la segunda iteración mostraron una tasa de error mejor que 1/5800, calculada sobre 86,000nt de clones secuenciados.

La Figura 31 muestra el análisis del fragmento C.E. después de las reacciones de PCR durante el método de construcción de la librería P53 como se realizó anteriormente. La Figura 31A muestra los resultados con respecto al nivel 1 de construcción de la librería P53, mientras que las Figuras 31B-E muestran los resultados con respecto a los niveles 2-5, respectivamente. La Figura 31F muestra el nivel 6 de construcción de la librería P53 con todas las 6 variantes de la librería P53 completa.

Adicionalmente se puede economizar la corrección de errores de las librerías. Por ejemplo, en la construcción de una librería con 256 miembros (Figura 29A arriba) un subconjunto de los cuatro clones que contienen todos los componentes de la librería (Figura 29B arriba) se puede construir y corregir los errores. Luego, todos los 256 miembros de la librería se construyen a partir de estos cuatro clones corregidos sin errores.

Los materiales y métodos fueron como se describió anteriormente para la síntesis de GFP.

La Figura 32 muestra la composición de los fragmentos naturales y sintéticos. La Figura 32A muestra un producto de PCR de dos fragmentos de 700 pb. La Figura 32B muestra un producto de PCR de 1363bp compuesto de los fragmentos de la A y un fragmento 1700bp de un plásmido natural. La Figura 32C muestra un producto de PCR de un fragmento de 3000bp compuesta de los fragmentos de B.

La Figura 33 muestra el análisis de fragmentos de C.E. de un ejemplo de una reacción de alargamiento. Una hebra se marca por HEX (verde) y luego por FAM (azul). El producto alargado se marca de esta manera por dos fluoróforos (a la derecha). Un exceso de la hebra marcada por HEX (pico verde a la izquierda) no se alarga.

La Figura 34 muestra un ejemplo de una reacción de exonucleasa lambda, que genera ssADN a partir de dsADN mediante exonucleasa lambda. El carril 1 muestra el dsADN marcado con un fosfato 5'; mientras que el carril 2 muestra el mismo ADN después de tratamiento con exonucleasa lambda, con una banda de ssADN inferior.

Ejemplo - Molécula de PCR sencilla

En el proceso de corrección de errores de acuerdo con las realizaciones de ejemplo del método de la presente invención, el aislamiento y la amplificación de moléculas de ADN sencillas se puede realizar con la clonación en células. Adicionalmente, esto también se puede realizar con una molécula de PCR sencilla como se describe en este Ejemplo.

La molécula de PCR sencilla se realiza por dilución de la plantilla original, que tiene uno o más errores, a una concentración de menos de una molécula por reacción de PCR. Luego, se desarrolla una reacción de PCR de molécula sencilla en tiempo real como se describe en la sección de métodos del ejemplo anterior para la síntesis de GFP (para reacciones de PCR utilizando la máquina BioLine), excepto que los ciclos térmicos son más utilizados para la amplificación (50 ciclos). Luego se calcula el número de amplificaciones de molécula sencilla requerido para lograr un corte mínimo wa. Los fragmentos de PCR resultantes se secuenciaron y se corrigió el error de acuerdo con el método de corrección de errores de la presente invención.

La preparación de la plantilla de ADN para smPCR (PCR de molécula sencilla) se realiza como sigue. Es esencial evitar la hibridación de hebras de ADN con diferentes secuencias desde que el smPCR realizado en dichas poblaciones da como resultado dos poblaciones de ADN amplificado en lugar de una. Este es el resultado de hibridación de fase estacionaria. El PCR en tiempo real se utiliza con el fin de supervisar el proceso de amplificación y detener la plantilla que genera PCR a partir de alcanzar la fase estacionaria (los resultados se muestran a continuación). Esto proporciona la seguridad de que ambas hebras de una molécula de dsADN a partir de esta reacción tengan la misma secuencia. Para los experimentos en los que no se utilizó este método, los resultados de la secuenciación después de amplificación mostraron que el cromatograma se convierte desde esa sola población hasta esa más de una población en el mismo par de bases (datos no mostrados). Esto significa que en el sitio de inserción de las poblaciones ya no se sincronizaron, sino más bien se desplazaron en un par de bases. Este resultado se obtiene debido a que las dos hebras que comprenden la molécula de dsADN, que se amplifica y luego se secuencia, no eran idénticas en secuencia.

La Figura 35 se relaciona con la monitorización en tiempo real del proceso de amplificación por PCR que muestra que se utiliza una sola población para el smPCR.

La Figura 36 se relaciona con la secuenciación de una reacción de PCR que comienza a partir de una molécula de ADN plantilla con una secuencia. Las secuencias superior e inferior son las reacciones de secuenciación inversa y directa de la misma plantilla. Observe que aunque la secuencia contiene una mutación (inserción de A) en comparación con la secuencia original, ambas hebras continúan exhibiendo un cromatograma de secuenciación normal. Este resultado se obtiene debido a que la reacción de PCR comenzó con una molécula de ADN plantilla y por lo tanto todas las moléculas de ADN secuenciadas tienen la misma secuencia.

La Figura 37 muestra la electroforesis en gel de ADN de una sola molécula de PCR con el gen sintético construido de forma recursiva que codifica el GFP como plantilla. La Figura 37A muestra un control positivo que contiene más de una molécula por reacción. La Figura 37B muestra el PCR del gen de GFP con una concentración de 0.5 moléculas por reacción de PCR.

Ejemplo - Síntesis de ADN en un chip

El ADN sintetizado en los chips opcionalmente se puede utilizar como bloques de construcción en algunas realizaciones del protocolo de construcción de acuerdo con la presente invención. La liberación de bloques de construcción de ADN desde el chip en una forma controlada, de tal manera que todas y cada secuencia en el chip se pueda eliminar o copiar desde el chip es individualmente esencial. Esto se puede hacer al crear una "dirección" específica para cada secuencia en el chip. Esto se puede hacer de diferentes formas, que incluyen, pero no se limitan a, se pueden activar y dividir unidades estructurales foto-dividibles en el ADN mediante luz dirigida a una secuencia de ADN específica en el chip; la división química de unidades estructurales dividibles en el ADN con indicaciones de tecnología de impresión a una secuencia específica de ADN en el chip; y las secuencias de ADN específicas en el chip se pueden amplificar al utilizar una combinación de cebadores de PCR que amplifican específicamente solo esa secuencia a partir de todas las secuencias presentes en el chip. Por supuesto cualquiera de los métodos anteriores opcionalmente se puede realizar como se conoce en la técnica.

Sección 2 -polímeros

Sin desear estar limitado de ninguna manera, en esta Sección se relaciona con las realizaciones de ejemplo, ilustrativas de la presente invención (método y sistema) con respecto a la construcción de polímeros distintos de los biopolímeros.

Cabe señalar que las realizaciones de la presente descripción descrita anteriormente con respecto a los biopolímeros pueden, opcionalmente, y preferiblemente ser extendidos a los polímeros, con los únicos cambios que se requieren en el tipo de reacciones requeridas para la química sintética. Una descripción completa y definición de muchos tipos diferentes de reacciones y definiciones poliméricas de los polímeros se proporcionan en K. Horie et al, Pure Appl. Chem. 76 (4), 889-906, 2004. Brevemente, algunos tipos de reacciones químicas poliméricas incluyen, pero no se limitan a, amplificación y entrecruzamiento químico. Las unidades poliméricas químicas pueden por ejemplo reaccionar juntas con el fin de formar cadenas poliméricas largas bajo condiciones químicas definidas. El método de construcción recursiva de acuerdo con la presente invención opcionalmente se puede utilizar al comenzar con bloques de construcción poliméricos más cortos y construir de forma recursiva los polímeros más largos. Opcionalmente la modificación química del polímero y/o porciones de los mismos resultante también se puede llevar a cabo, como se conoce en la técnica.

Los aparatos conocidos para la síntesis química a gran escala también se pueden utilizar opcionalmente con la presente invención; por ejemplo, el documento US 5529756 describe un método para síntesis química a gran escala al utilizar una matriz. El documento US 6045755 describe un método para la síntesis química combinatoria, que opcionalmente se puede combinar con los métodos de la presente invención para sintetizar, por ejemplo una librería de polímero.

Sección 3 - Aparato para el sistema y método de la presente invención

Esta sección se relaciona con una pluralidad de diferentes realizaciones de ejemplo, ilustrativas de un aparato para implementar diversas realizaciones del sistema y método de la presente descripción. La Figura 38 es un diagrama de flujo detallado de la implementación en una arquitectura dada de acuerdo con una realización de ejemplo de la presente descripción, mientras que las Figuras 39-42 son ilustraciones arquitectónicas generalizadas de algunas posibles realizaciones, de ejemplo, ilustrativas, de máquinas de acuerdo con la presente descripción.

La Figura 38 muestra un sistema automatizado de ejemplo de acuerdo con la presente descripción. Como se muestra, un sistema 3500 preferiblemente cuenta con un dispositivo 3502 de cómputo que recibe las especificaciones de secuencia y/o librería 3503 para la secuencia y/o librería que se va a sintetizar. El dispositivo 3502 de cómputo luego determina el proceso para construcción, que es preferiblemente jerárquica y que se determina más preferiblemente de acuerdo con el algoritmo de corte mínimo descrito aquí. El dispositivo 3502 de cómputo entonces preferiblemente envía las instrucciones para la síntesis de oligonucleótidos a un sintetizador 3504 adecuado. Alternativa o adicionalmente, el dispositivo 3505 de cómputo también puede opcionalmente identificar bloques de construcción de otras fuentes (por ejemplo, ADN construido previamente y\o ADN de origen natural para uso en el proceso sintético).

El dispositivo 3502 de cómputo preferiblemente también envía el programa requerido a un robot de síntesis de 3508, que también recibe los bloques de construcción sintetizados a partir del sintetizador de 3504 y\u, opcional o alternativamente de otras fuentes como se describió anteriormente que opcionalmente se pueden almacenar en un almacenamiento 3507. El robot 3508 de síntesis luego sintetiza, la secuencia y/o librería necesaria, que preferiblemente se comprueba para exactitud por un secuenciador 3510. La secuencia y/o librería requerida preferiblemente luego se almacena en un almacenamiento 3511de salida. Los flujos de información muestran con flechas que tiene líneas de trazos, mientras que el flujo de reactivos y/o material físico se muestra con flechas que tienen líneas sólidas.

La Figura 39 muestra otra realización ilustrativa de un aparato de acuerdo con la presente descripción, que muestra un aparato 3900. El proceso de operación opcional y preferiblemente se inicia con los reactivos a un almacenamiento 3902 de reactivos que se comunica con un robot 3904 de manipulación de líquidos. Los reactivos se pueden obtener a partir de o colocar en el almacenamiento 3902 de reactivos por robot 3904 de manipulación de líquidos. Las operaciones del robot 3904 de manipulación de líquidos son controlados preferiblemente por un ordenador 3906 que preferiblemente se opera adicionalmente por un software de acuerdo con los métodos descritos anteriormente de la presente descripción para la síntesis de polinucleótidos.

El ordenador 3906 preferiblemente recibe información con respecto a la secuencia (s) a ser construida y luego realiza una o más métodos como se describe aquí para la determinación de la construcción recursiva de un polinucleótido, opcionalmente y más preferiblemente con corrección de errores. El robot 3904 de manipulación de líquidos se comunica preferiblemente con un dispositivo 3906 de secuenciación, un dispositivo 3908 de PCR, un dispositivo 3910 de purificación y un dispositivo 3912 de medición para dirigir el desempeño del proceso sintético. Por ejemplo, suponiendo que el proceso de síntesis comprende una pluralidad de subprocesos, opcional y preferiblemente cada subproceso incluye amplificación por PCR con dispositivo 3908de PCR, opcionalmente seguido mediante secuenciación con el dispositivo 3906 de secuenciación para buscar uno o más errores (alternativa y, opcionalmente, la secuenciación se realiza al final del proceso sintético, y/o después de haber llevado a cabo una pluralidad de subprocesos pero antes de que se complete el proceso de síntesis). El dispositivo 3910 de purificación se puede utilizar opcionalmente entre el desarrollo de subprocesos (por ejemplo si la secuenciación se lleva a cabo opcionalmente entre subprocesos) pero, alternativamente, se lleva a cabo opcionalmente después de que se han sintetizado una o más secuencias (y/o después de haber desarrollado una pluralidad de subprocesos pero antes de que se complete el proceso de síntesis).

El dispositivo 3912 de medición se utiliza preferiblemente durante el proceso sintético para monitorizar la cantidad de polinucleótido que se sintetiza.

La Figura 40 muestra otra realización opcional, que se relaciona con una configuración de aparato con un componente sintetizador de oligonucleótidos. Un aparato 4000 presenta muchos de los mismos y/o similares componentes como para la Figura 39, excepto que también está incluido un sintetizador 4002 de oligonucleótidos; el reactivo 4004 de almacenamiento es opcional y preferiblemente se configura para soportar el almacenamiento de dichos oligonucleótidos.

La Figura 41 se relaciona con una configuración de aparato con soporte sólido, que es altamente similar a aquella de la Figura 41, excepto que ahora el aparato de 4100 preferiblemente carece de un dispositivo de purificación separada, ya que toda la purificación preferiblemente se lleva a cabo sobre el soporte sólido al que se adhiere el ADN, como se conoce en la técnica.

La Figura 42 se relaciona con una configuración de aparato con chips de ADN y micro-fluídicos. Los métodos para la síntesis de chip de ADN descrito anteriormente también se pueden utilizar aquí opcionalmente. Adicionalmente, los métodos y dispositivos para microfluidos se describen en Whitesides, G.M. The origins and the future of microfluidics. Nature 442, 368-373 (2006). De nuevo este aparato es similar a la configuración de la Figura 40, excepto que en lugar de un robot de manipulación de líquidos, se proporciona preferiblemente un dispositivo 4202 micro-fluídico, que opera como se conoce en la técnica. También el aparato opcional y preferiblemente carece de un dispositivo de PCR separado.

Para las diferentes realizaciones del aparato mostrado en las Figuras 39-42, cada uno de dichos aparatos se puede opcionalmente y preferible implementar como un dispositivo pequeño, portátil o incluso en miniatura. También, para cada uno de dicho aparato anterior, el control se proporciona preferiblemente por el ordenador 3906. La comunicación entre el ordenador 3906 y, o el robot 3904 de manipulación de líquidos o el dispositivo 4202 micro-fluídico, es preferiblemente bidireccional. Dicha comunicación bidireccional preferiblemente proporciona retroalimentación para la operación adecuada del proceso de construcción y para el control de cantidad y calidad durante el proceso de construcción.

REFERENCIAS

Goeden-Wood NL, Conticello VP, Muller SJ, Keasling JD. Improved assembly of multimeric genes for the biosynthetic production of protein polymers. Biomacromolecules. 2002 Jul-Aug;3(4):874-9.

Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene. 1995 Oct 16; 164(1):49-53.

Stahl S, Hansson M, Ahlborg N, Nguyen TN, Liljeqvist S, Lundeberg J, Uhlen M. Solidphase gene assembly of constructs derived from the Plasmodium falciparurnmalaria bloodstage antigen Ag332.Biotechniques. 1993 Mar; 14(3):424-34.

Hostomsky Z, Smrt J, Amold L, Tocik Z, Paces V. Solid-phase assembly ofcow colostrum trypsin inhibitor gene.Nucleie Acids Res. 1987 Jun 25; 15(12):4849-56.

H Caruthers MH. Gene synthesis machines: DNA chernistry and its uses.Science. 1985 Oct 18;230(4723):281-5.

SantaLucia J Jr. A unified view ofpolyrner, dumbbell, and oligonucleotide DNA nearestneighbor thermodynamics. Proc Natl Acad Sci U S A. 1998 Feb 17;95(4):1460-5.

LISTA DE SECUENCIAS

<110> Yeda Research And Development Company Limited

<120> Alargamiento Iterado Programable: Un método para elaborar genes sintéticos y ADN combinatorio y librerías de proteína

<130> 783

<160> 22

<170> BiSSAP 1.0

<210> 1
<211> 748
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1..748

<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 1
atgagtattc aacatttccg tgtccgtgtc gcccttattc ccttttttgc ggcattttgc 60
cttcctgttt ttgctcaccc agaaacgctg gtgaaaagta aaagatgctg aagatcagtt 120
gggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc ttgagagttt 180
tcgccccgaa gaacgttttc caatgatgag catttaaagt tctgctatgt ggcgcggtat 240
tatcccgtat tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg 300
acttggttga gtgctgccat aaccatgagt gataacactg cggccaacct tacttctgac 360
aacgatctgc agtgctgcca taaccatgag tgattaacac tgcggccaac ttacttctga 420
caacgatcgg aggaccgaag agctaaccgc ttttttgcac aacatggggg atcatgaact 480
cgccttgatc gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc 540
acgatgcctg taggcgaact acttgcaatg gcaacaacgt tgcgcaaact attaactctc 600
tagcttcccg gccaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt 660
ctgcgctcgg cccttccggc tggctggtta ttgctgataa atctggagcc ggtgagcgtg 720
agcgtgggtc tcgcggtatc attgcagc 748

<210> 2
<211> 247
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1..247
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 2
gtcgccttat tccctttttt gcggcatttt gccttcctgt ttttgctcac ccagaaacgc 60
tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg 120
atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt ccaatgatga 180
gcattttaaa gttctgctat gtggcgcggt attatcccgt attgacgccg ggcaagagca 240
actcggt 247

<210> 3
<211> 587
<212> ADN

<213> Secuencia Artificial

<220>
<221> fuente
<222> 1. .587
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 3
ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc caatgatgag cattttaaag 60
ttctgctatg tggcgcggta ttatcccgta ttgacgccgg gcaagagcaa ctcggtcgcc 120
gcaatacact attctcagaa tgacttggtt gagtgctgcc ataaccatga gtgataacac 180
tgcggccaac ttacttctga ccaacgatct gcagtgctgc cataaccatg agtgataaca 240
ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc 300
acaacatggg ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca 360
taccaaacga cgagcgtgac accaacgatg cctgtagcaa tggcaacaac gttgcgcaaa 420
ctattaactg gcgaactact tactctagct tcccggcaac aattaataga ctggatggag 480
gcggataaag ttgcaggacc acttctgcgc tcggcccttc cggctggctg gttattgctg 540
ataaatctgg agccggtgag cgtgggtctc gcggtatcat tgcagca 587

<210> 4
<211> 307
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1. .307
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 4
cttgcccggc tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat 60
cattggaaaa cgttcttcgg ggcgacttgc ccggctcaat acgggataat accgcgccac 120
atacgagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga gggataatac 180
cqcqccacat aqcaqaactt taaaaqtqct catcattqqa aaaacqttct tcqqqqcqaq 240
qqataatacc qcqccacata qcaqaacttt aaaaqtgctc atcattggaa aacqttcttc 300
qqqqcqa 307

<210> 5
<211> 86
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 86
<223> /mol_type="ONA"
/note="sintético"
/orqanism="Secuencia Artificial"

<400> 5
ctqccccqaa qaacqttttc caatqatqaq cacttttaaa qttctqctat qtqqcqcqqt 60
attatcccqt attqacqccq qqcaaq 86

<210> 6
<211> 87
<212> ADN
<213> Secuencia Artificial

<220>

<221> fuente
<222> 1 .. 87
<223> /mol_type="ADN"
/note="sintético"
/orqanism="Secuencia Artificial"

<400> 6
tcqccccqaa qaaacqtttt ccaatqatqa qcacttttaa aqttctqcta tqtqqcqcqq 60
tattatcccq tattqacqcc qqqcaaq 87

<210> 7
<211> 86
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 86
<223> /mol_type="ONA"
/note="sintético"
/orqanism="Secuencia Artificial"

<400> 7
tcqccccqaa qaacqttttc caatqatqaq cacttttaaa qttctqctat qtqqcqcqqt 60
attatcccqt attgacqccq qgcaaq 86

<210> 8
<211> 86
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 86
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 8
tcgccccgaa gaacgttttc caatgatgag cacttttaaa gttctgctat gtggcgcggt 60
attatcccgt attgacgccg ggcaag 86

<210> 9
<211> 134
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 134
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 9
ttacatatcg aactggatct caacagcggt aagatcctga gagttttcgc cccgaagaac 60
gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtattg 120
acgccgggca agaa 134

<210> 10
<211> 107
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 107
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 10
ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc caatgatgag cacttttaaa 60
gttctgctat gggcgcggta ttatcccgta ttgacgccgg gcaagaa 107

<210> 11
<211> 107
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 107
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 11
ggtaagagcc ttgagaqttt tcgccccgaa gaacgttttc caatgtgagc acttttaaag 60
ttctgctatg tggcgcgqta ttatcccqta ttgacgccgg gcaagaa 107

<210> 12
<211> 137
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1. .137
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 12
tacatcgaac tggatctcaa cagcgqtaag atccttgaga qttttcgccc cgaagattat 60
qtggcggcgg tattatcccg tattgacggc cgggcaagcg ttttccaatg atgagcactt 120
ttaaaqttct gctatgc 137

<210> 13
<211> 130
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1. .130
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 13
tacatcgaac tggatctcaa cagcgqtaag atccttgaga qttttcgccc cgaagaacqt 60
ttccaatgat gagcactttt aaaqttctgc tatqtggcgc gqtattatcc cqtattgacg 120
gccgggcaag 130

<210> 14

<211> 132
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1. .132
<223> /mol_type= “ DNA”
/note=“sintético”
/organism=“Secuencia Artificial”

<400> 14
tacatcgaac tggatctgaa cagcgqtaag atccttgaga qttttcgccc cgaagaacqt 60
ttttccaatg atgagcactt ttaaaqttct gctatqtggc gcgqtattat cccqtattga 120
cggccgggca ag 132

<210> 15
<211> 132
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 132
<223> /mol_type=“ADN”
/note=“sintético”
/organism=“Secuencia Artificial”

<400> 15
tacatcgaac tggatctcaa cagcggtaag atccttgaga gttttcgccc cgaagaacgt 60
tttccaatga tgagcacttt taaaagttct gctatgtggc gcggtattat cccgtattga 120
cggccgggca ag 132

<210> 16
<211> 131
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222>'1 .. 131
<223> /mol_type=“ADN”
/note=“sintético”
/organism=“Secuencia Artificial”

<400> 16
tacatcgaac tggatctcaa cagcggtaag atccttgaga gttttcgccc cgaagaacgt 60
tttccaatga tgagcacttt taaagttctg ctatgtggcg cggtattatc ccgtattgac 120
ggccgggcaa g 131

<210> 17
<211> 86
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 86
<223> /mol_type=“ADN”
/note=“sintético”
/organism=“Secuencia Artificial”

<400> 17

tcgccccgaa gaacgttttc catgatgagc acttttaaag ttctgctatg tggcgcggta 60
ttatcccgta ttgacggccg ggcaag 86

<210> 18
<211> 65
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 65
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 18
accggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag 60
tttta 65

<210> 19
<211> 84
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 84
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 19
tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt 60
aaacttggtc tgacagttac caat 84

<210> 20
<211> 84
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 84
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 20
caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc actgattaag 60
cattggtaac tgtcagacca agtt 84

<210> 21
<211> 79
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 79
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 21
tcgttcatcc atagttgcct gactgcccgt cgtgtagata actacgatac gggagggctt 60
accatctggc cccagtgct 79

<210> 22
<211> 51
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<222> 1 .. 51
<223> /mol_type="ADN"
/note="sintético"
/organism="Secuencia Artificial"

<400> 22
gaaacgattc cggtcgagac cttgcggtat cattgcagca ctggggccag a 51

REIVINDICACIONES

1. Un método para fabricar por lo menos semiautomáticamente un polinucleótido con errores mínimos o reducidos a través de un proceso de eliminación de error iterativo, el método comprende las etapas de:

a. analizar y dividir una secuencia de polinucleótido objetivo in silico en secuencias de ADN de cadena sencilla propensas a errores suficientemente cortas para ser sintetizadas por una síntesis de oligonucleótidos convencional, en el que dicho análisis comprende adicionalmente la construcción de cada secuencia de ADN de cadena sencilla corta por separado;

b. combinar de forma recursiva las secuencias de ADN de cadena sencilla propensas a errores in vitro para formar secuencias de ADN de cadena sencilla alargadas propensas a errores, el proceso recursivo comprende las etapas de fosforilar una de las dos hebras de entrada de hebra sencilla, hibridar las hebras, alargar las hebras para producir un ADN de hebra doble alargado y digerir la hebra fosforilada del ADN de doble hebra mediante una exonucleasa para producir una salida de ADN de una hebra alargada;

c. clonar y secuenciar la salida de ADN de hebra sencilla alargada para identificar errores en el mismo, en el que si se identifica un polinucleótido objetivo libre de error, este forma la salida y finaliza el proceso; o en donde si se identifica una salida propensa a errores, se identifica un grupo de regiones libre de errores más grande de ADN de una hebra alargada y se utilizan como entradas más precisas para repetir la combinación recursiva para producir un ADN de una hebra alargado sin errores o con pocos errores que la iteración anterior; y

d. repetir el proceso hasta que se forma una molécula objetivo libre de errores.

2. El método de la reivindicación 1 en el que el grupo de fragmentos libres de errores más largo de ADN de una sola hebra alargada que se utiliza como entradas más precisas para repetir la combinación recursiva se encuentran mediante un corte mínimo de un árbol de construcción recursiva, el corte mínimo en el árbol se define como sigue:

se dice que un nodo en el árbol se cubre mediante un conjunto de secuencias de ADN si por lo menos se produce una secuencia libre de errores en por lo menos una de las secuencias de ADN y el corte mínimo comprende el grupo más superficial o más cerca de la raíz de los nodos del árbol en el árbol que se consideran cubiertos por las secuencias de ADN.

3. El método de la reivindicación 1, que comprende adicionalmente la construcción de una librería de polinucleótidos.

4. El método de cualquier reivindicación precedente, en el que dicho polinucleótido objetivo se fabrica o se sintetiza de forma automática.

5. El método de cualquier reivindicación precedente en código implementable legible por ordenador que comprende: un grupo de instrucciones para construir el polinucleótido objetivo, dicho grupo de instrucciones dirigen un ensamble del polinucleótido objetivo a partir de una pluralidad de secuencias de ADN de hebra sencilla.

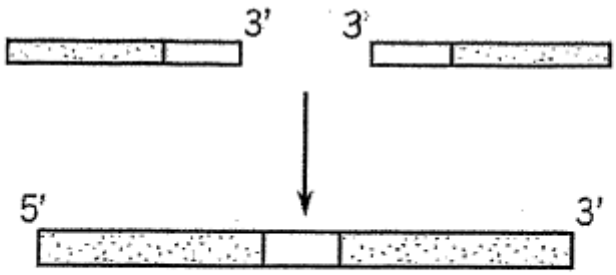


FIG. 1

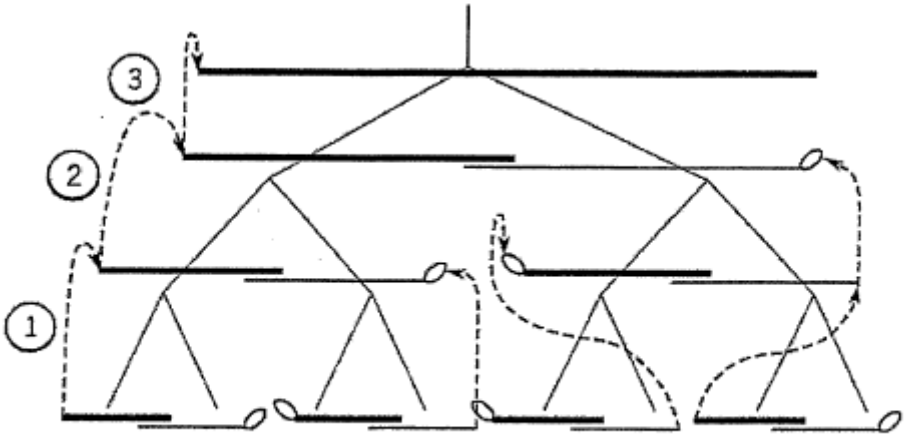


FIG. 2

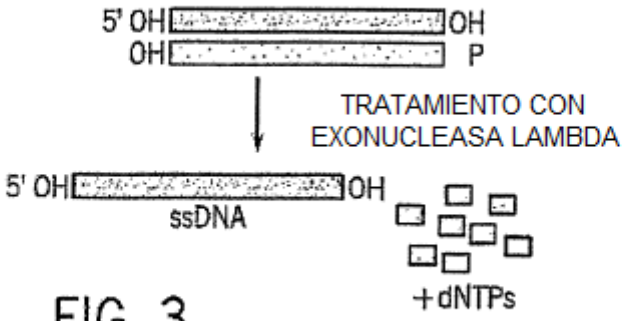


FIG. 3

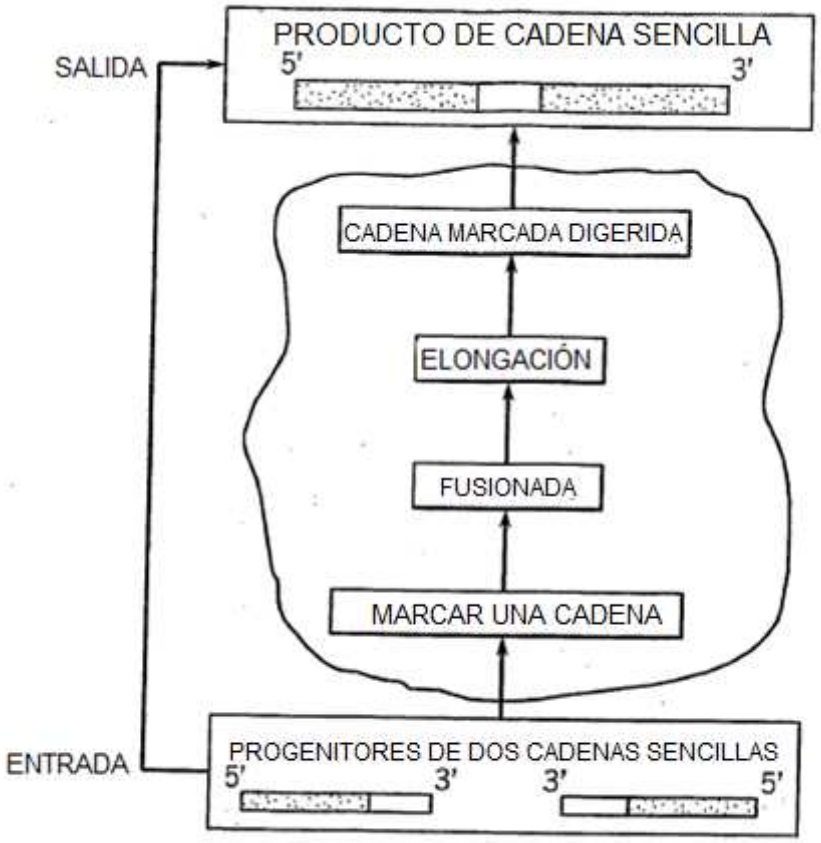

SALIDA
PRODUCTO DE CADENA SENCILLA
5'
3'
CADENA MARCADA DIGERIDA
ELONGACIÓN
FUSIONADA
MARCAR UNA CADENA
ENTRADA
PROGENITORES DE DOS CADENAS SENCILLAS
5'
3'
3'
5'

FIG. 4A

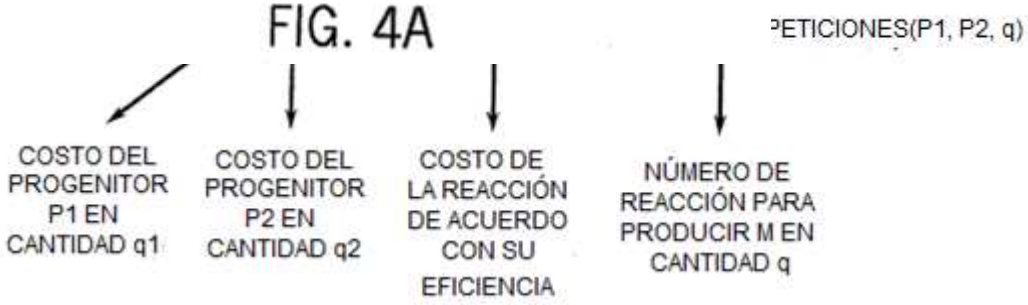

PETICIONES(P1, P2, q)
COSTO DEL PROGENITOR P1 EN CANTIDAD q1
COSTO DEL PROGENITOR P2 EN CANTIDAD q2
COSTO DE LA REACCIÓN DE ACUERDO CON SU EFICIENCIA
NÚMERO DE REACCIÓN PARA PRODUCIR M EN CANTIDAD q

FIG. 4B

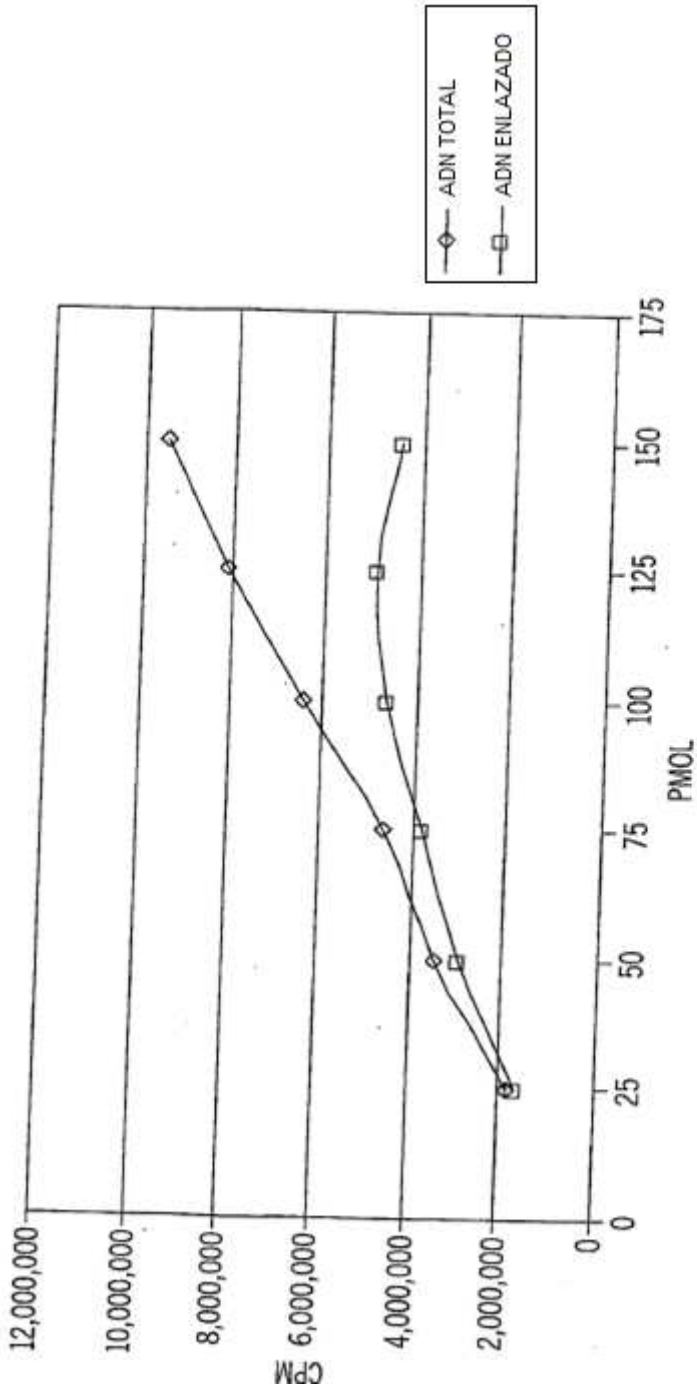

ADN TOTAL
ADN ENLAZADO
CPM
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
0
25
50
75
100
125
150
175
PMOL

FIG. 5A

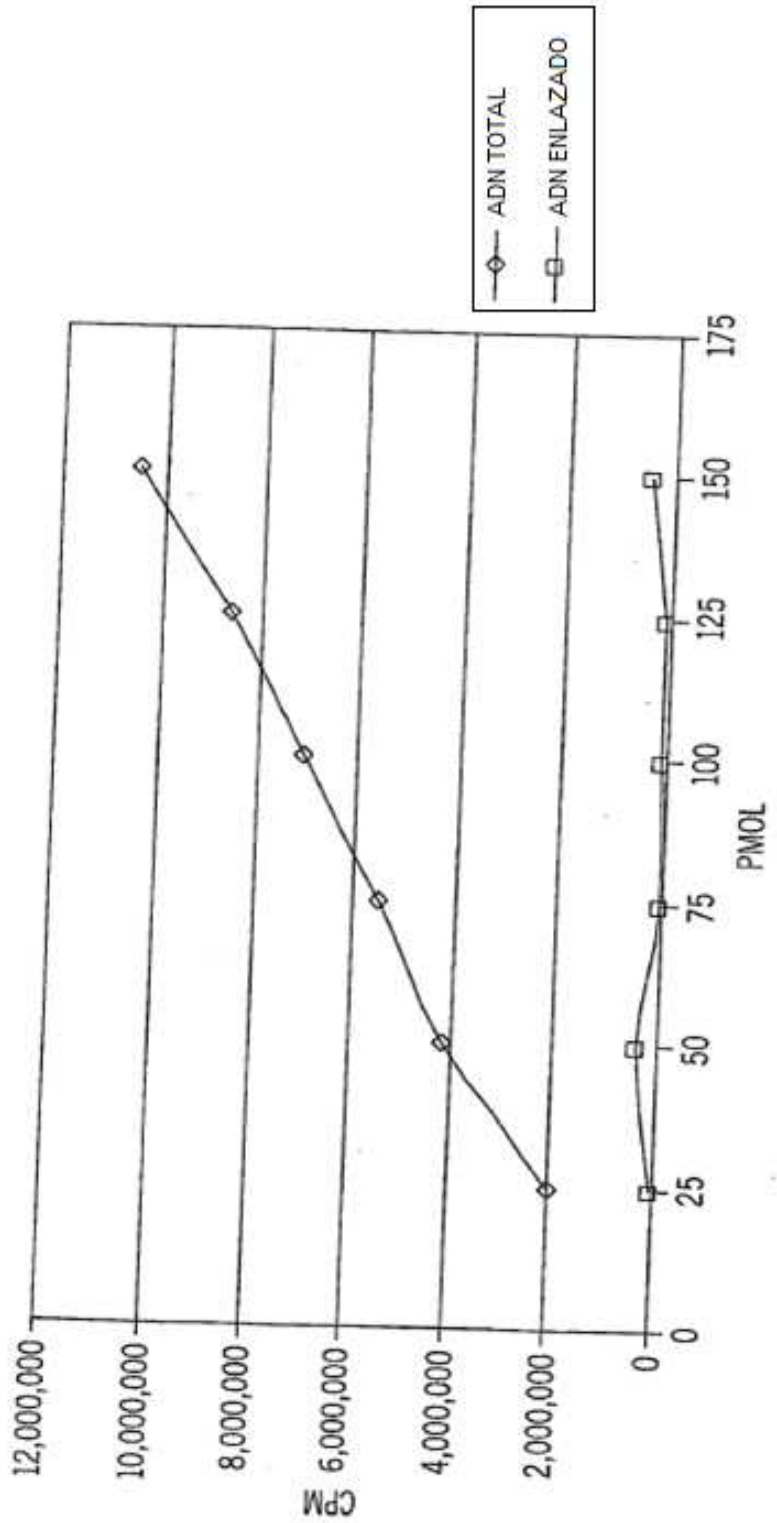

ADN TOTAL
ADN ENLAZADO
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
CPM
0
25
50
75
100
125
150
175
PMOL

FIG. 5B

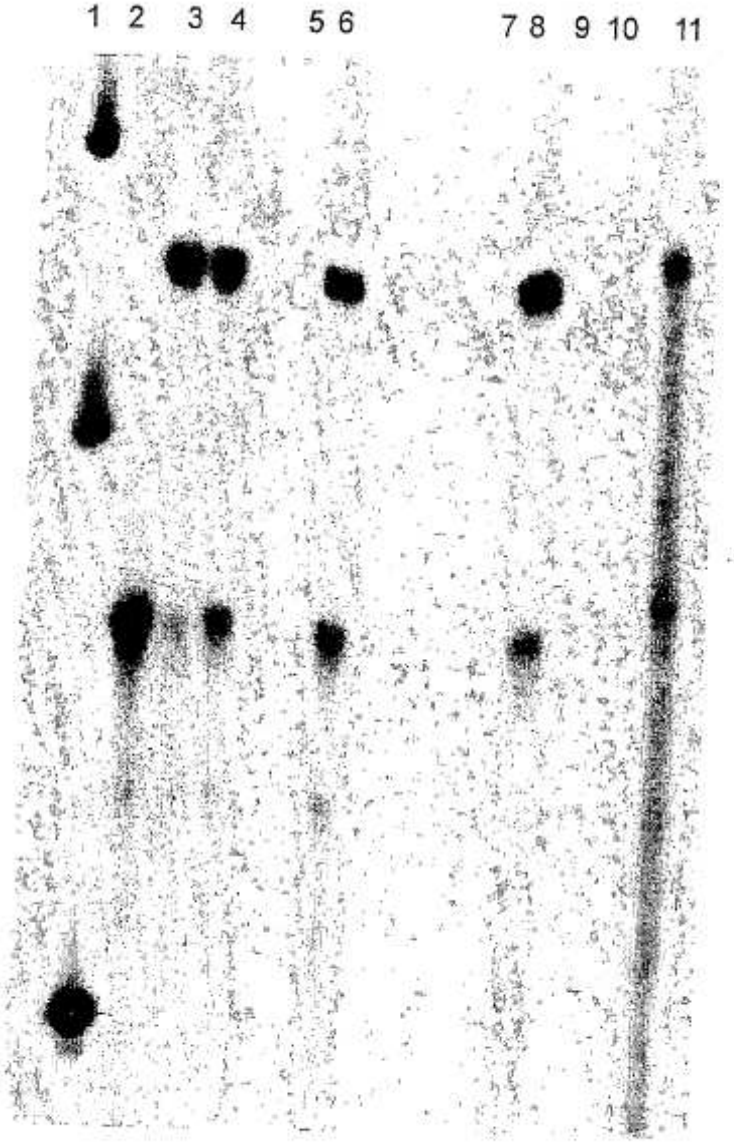


FIG. 6

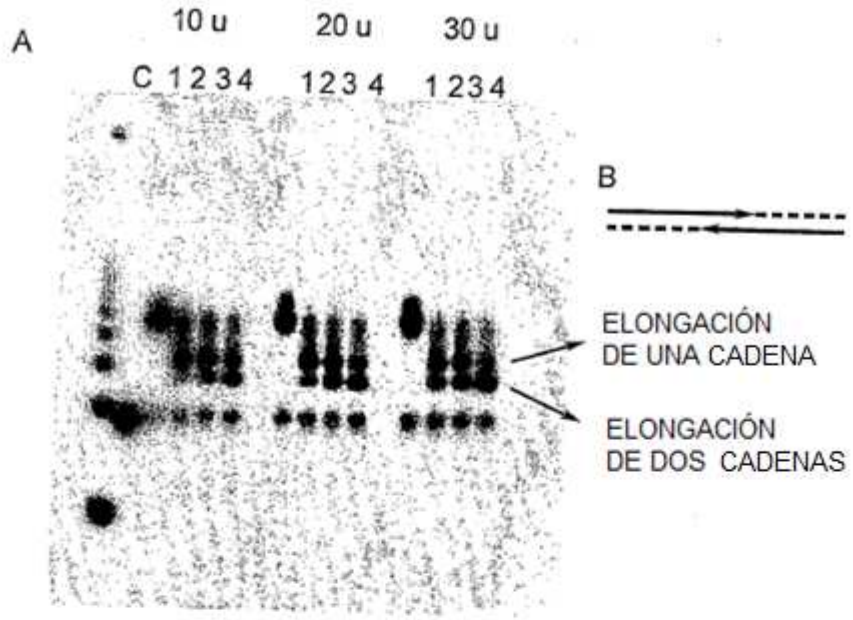

A
10 u
20 u
30 u
C 1 2 3 4
1 2 3 4
1 2 3 4
B
ELONGACIÓN
DE UNA CADENA
ELONGACIÓN
DE DOS CADENAS

FIG. 7

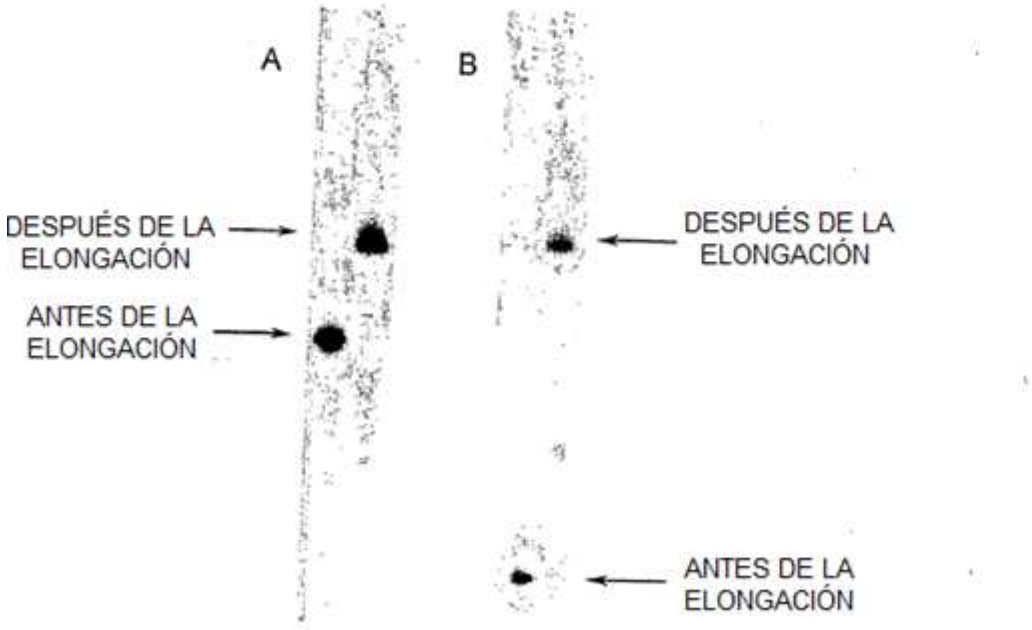
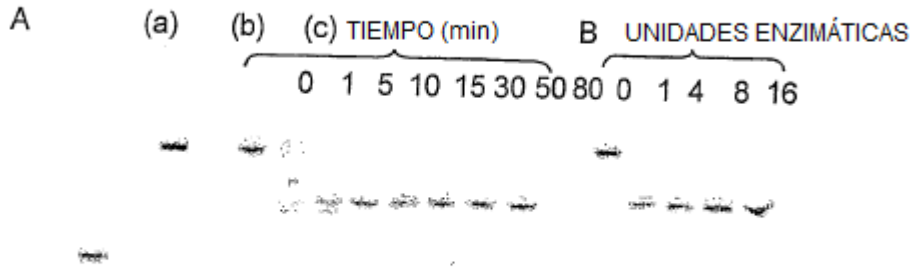

A
B
DESPUÉS DE LA
ELONGACIÓN
ANTES DE LA
ELONGACIÓN
DESPUÉS DE LA
ELONGACIÓN
ANTES DE LA
ELONGACIÓN

FIG. 8

FIG. 9


A
(a)
(b)
(c) TIEMPO (min)
0 1 5 10 15 30 50 80
B UNIDADES ENZIMÁTICAS
0 1 4 8 16

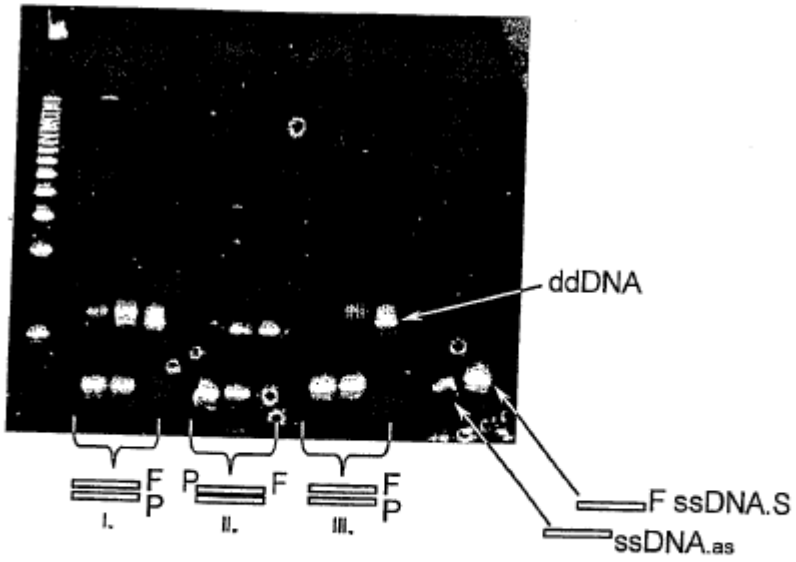

ddDNA
F
P
I.
P
F
II.
F
P
III.
F ssDNA.S
ssDNA.as

FIG. 10

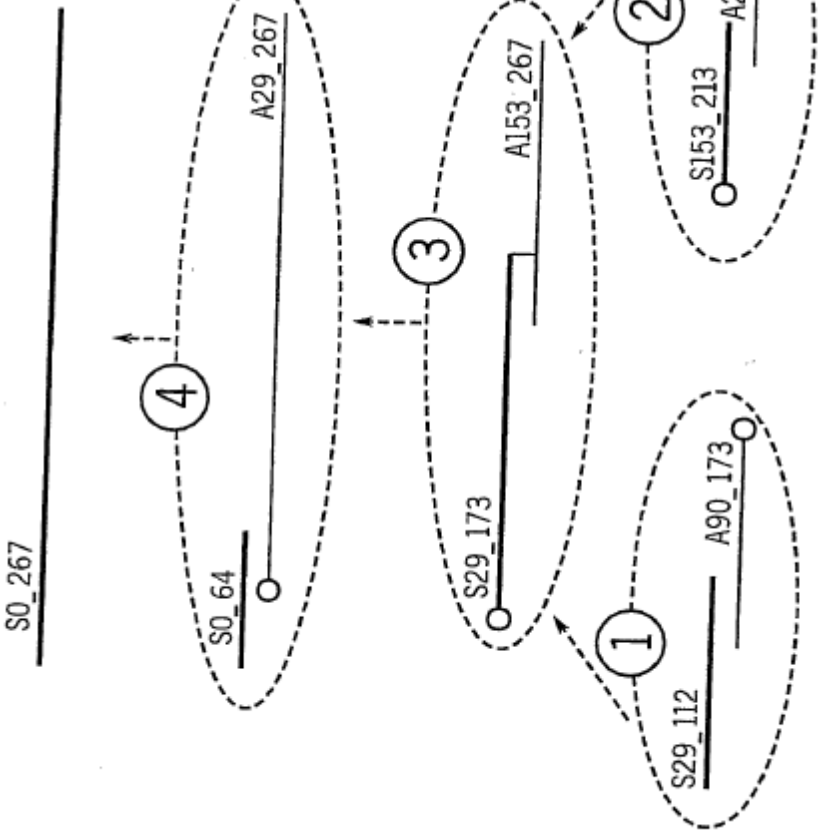

S0_267
4
S0_64
A29_267
3
S29_173
A153_267
2
S153_213
A217_267
1
S29_112
A90_173

FIG. 11

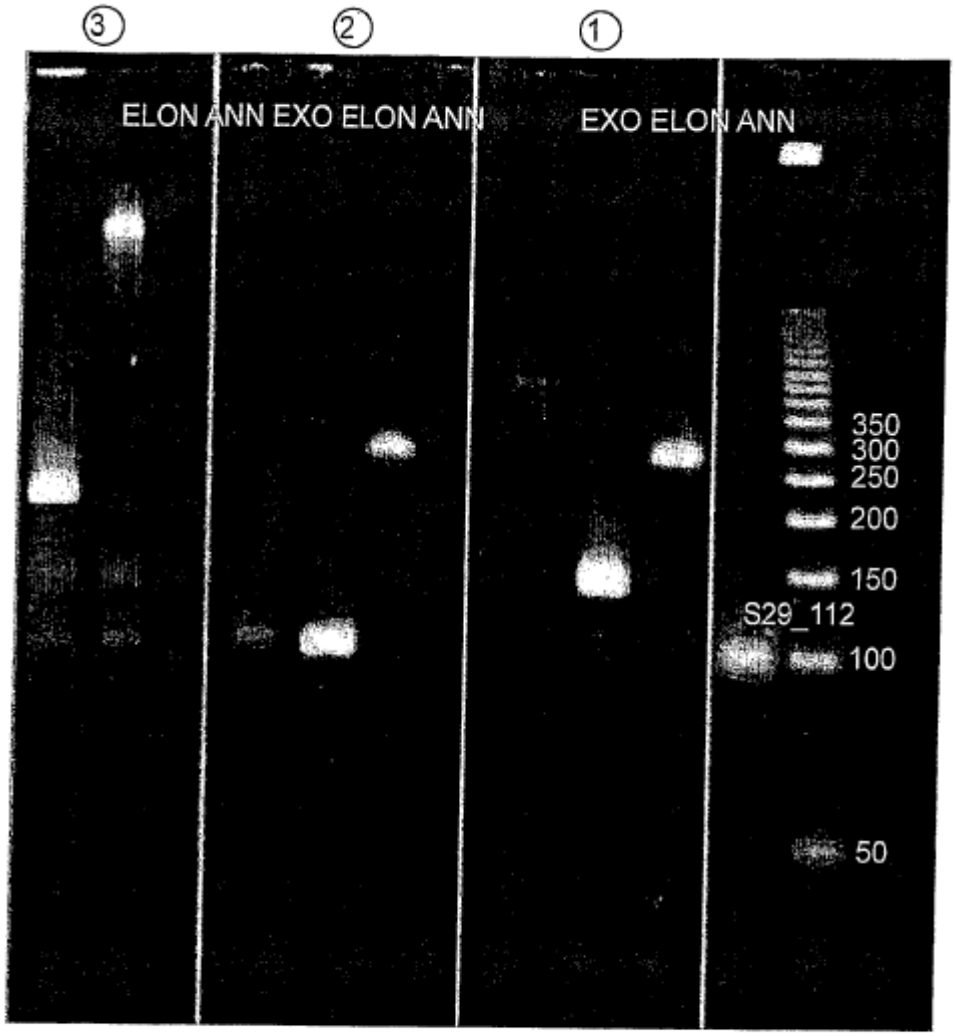

3
2
1
ELON ANN EXO ELON ANN
EXO ELON ANN
350
300
250
200
150
S29_112
100
50

FIG. 12

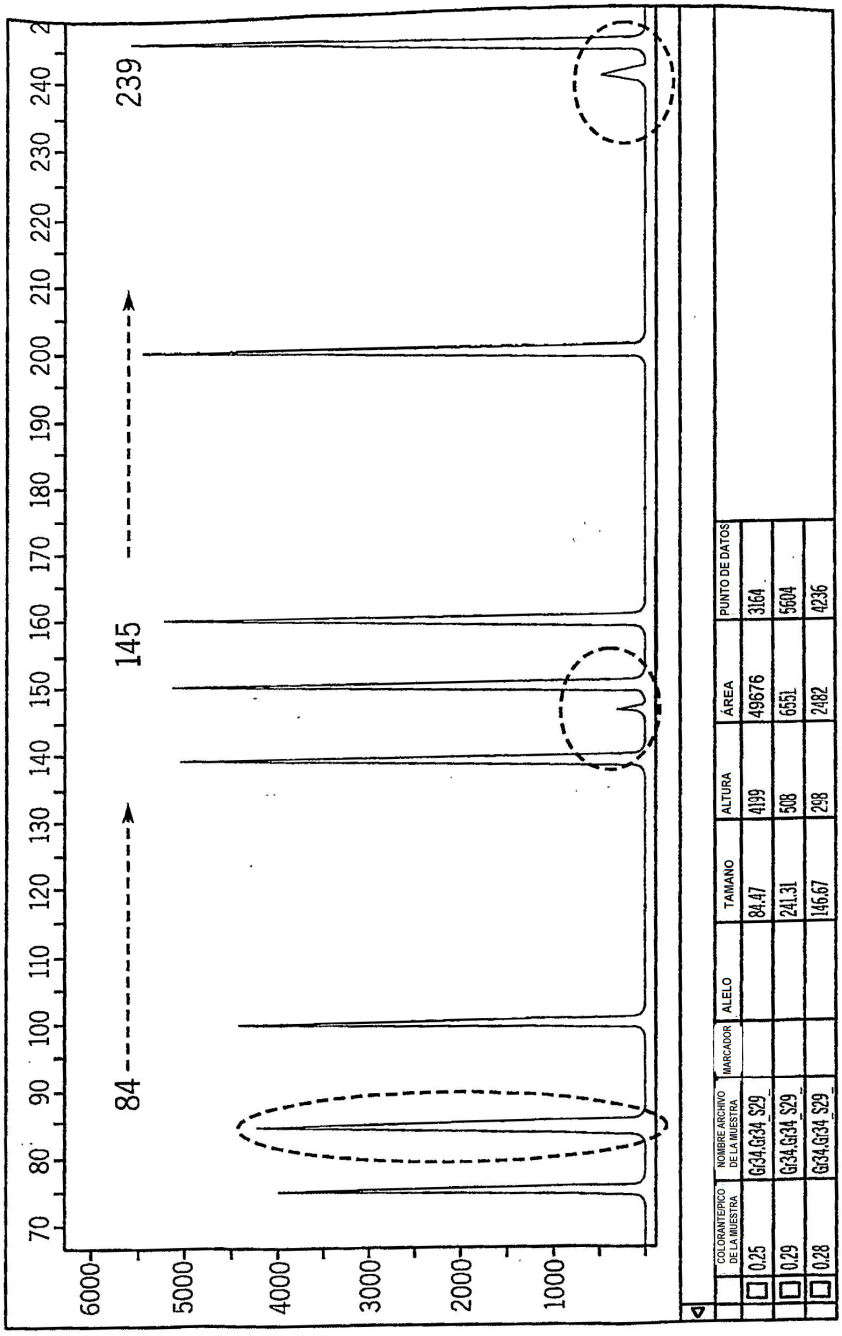

84
145
239
70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240
6000
5000
4000
3000
2000
1000
COLORANTE/PICO DE LA MUESTRA
NOMBRE ARCHIVO DE LA MUESTRA
MARCADOR
ALELO
TAMANO
ALTURA
ÁREA
PUNTO DE DATOS
0.25 G34.Gt34_S29 84.47 4199 49676 3164
0.29 Gt34.Gt34_S29_ 241.31 508 6551 5604
0.28 Gt34.Gt34_S29_ 146.67 298 2482 4236

FIG. 13

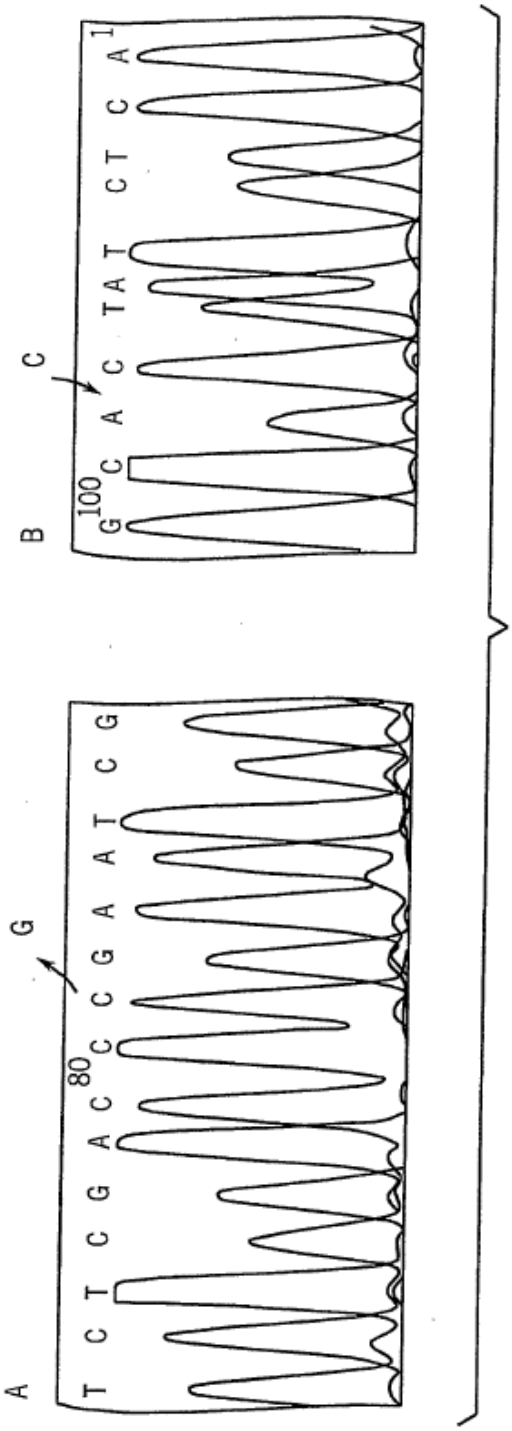


FIG. 14

FIG. 15

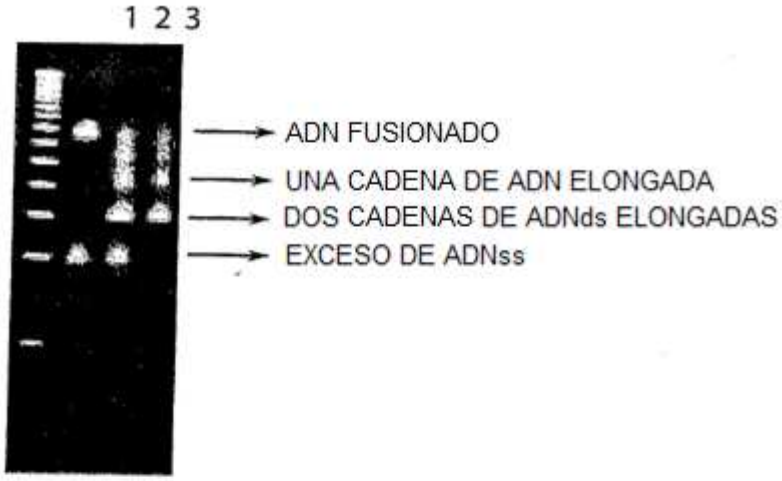
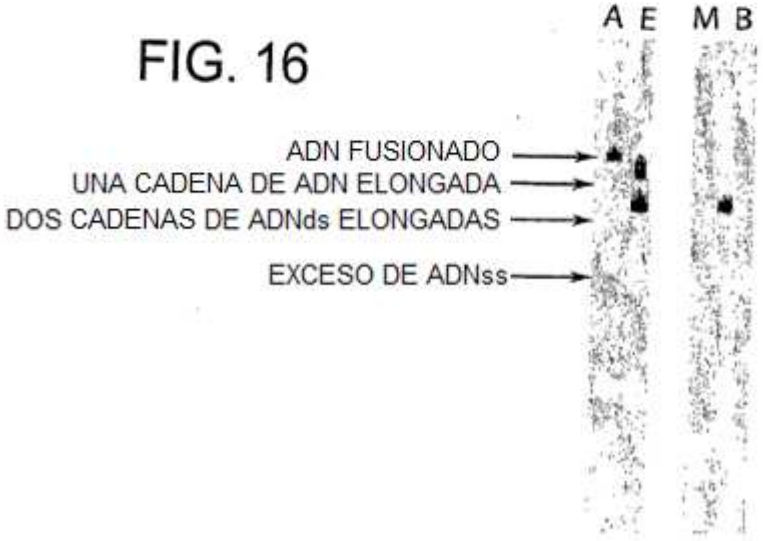

1 2 3
ADN FUSIONADO
UNA CADENA DE ADN ELONGADA
DOS CADENAS DE ADNds ELONGADAS
EXCESO DE ADNss

FIG. 16


A E M B
ADN FUSIONADO
UNA CADENA DE ADN ELONGADA
DOS CADENAS DE ADNds ELONGADAS
EXCESO DE ADNss

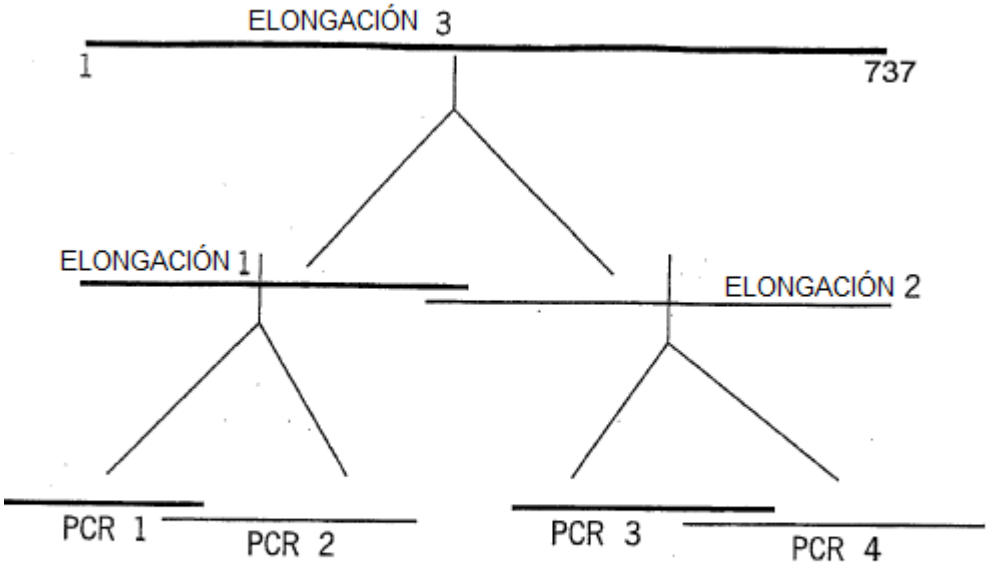
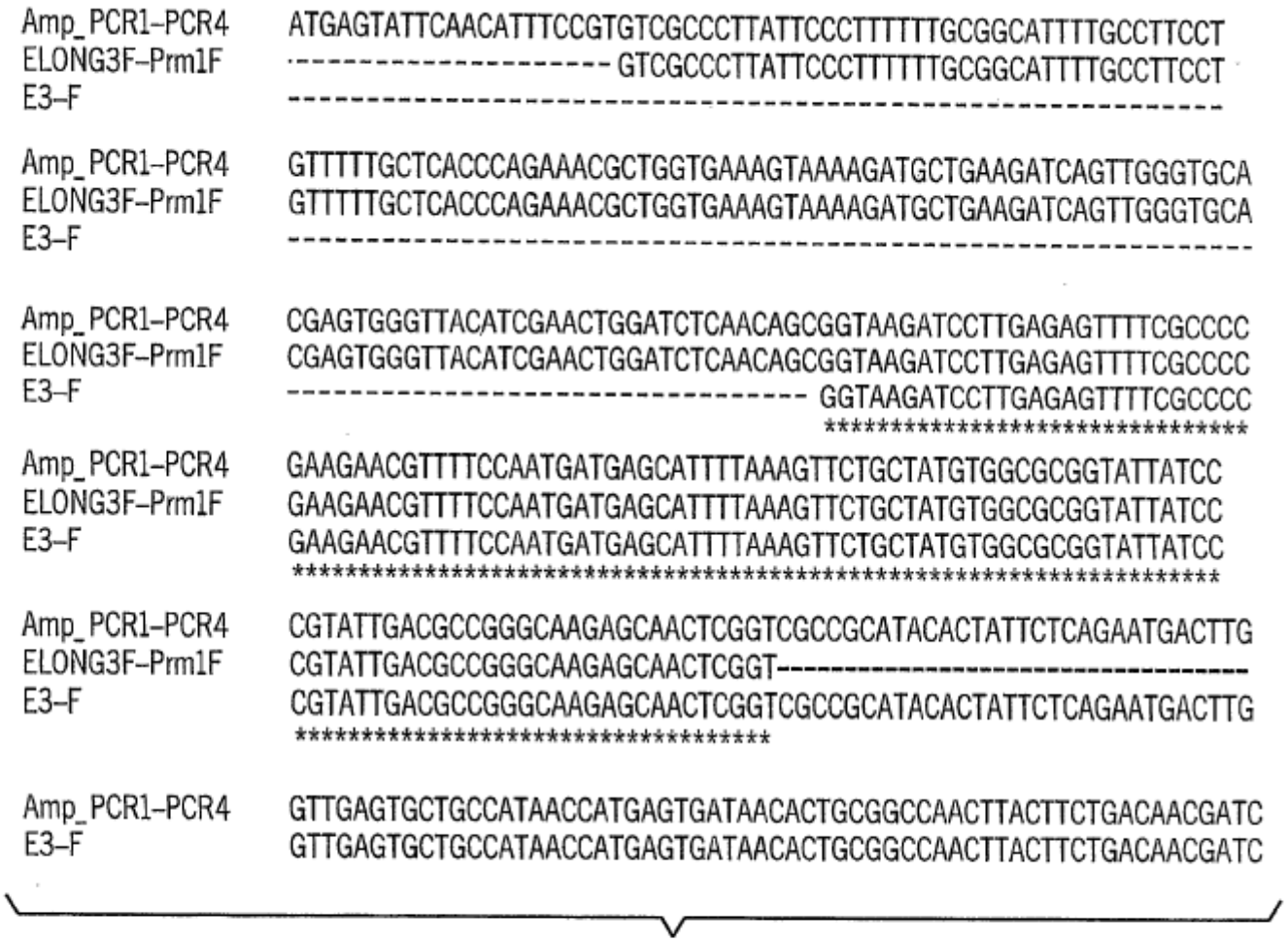

ELONGACIÓN 3
1
737
ELONGACIÓN 1
ELONGACIÓN 2
PCR 1
PCR 2
PCR 3
PCR 4

FIG. 17



A FIGURA 18B

FIG. 18A

FIG. 18B

DE LA FIGURA 18A

Amp_PCR1–PCR4 E3–F	TGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
	TGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
Amp_PCR1–PCR4 E3–F	GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT
	GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT
Amp_PCR1–PCR4 E3–F	GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATG
	GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATG
Amp_PCR1–PCR4 E3–F	CCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCT
	CCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCT
Amp_PCR1–PCR4 E3–F	TCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
	TCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
Amp_PCR1–PCR4 E3–F	TCGGCCCTTCCGGCTGGCTGGTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT
	TCGGCCCTTCCGGCTGGCTGGTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT
Amp_PCR1–PCR4 E3–F	CGCGGTATCATTGCAGC–
	CGCGGTATCATTGCAGCA

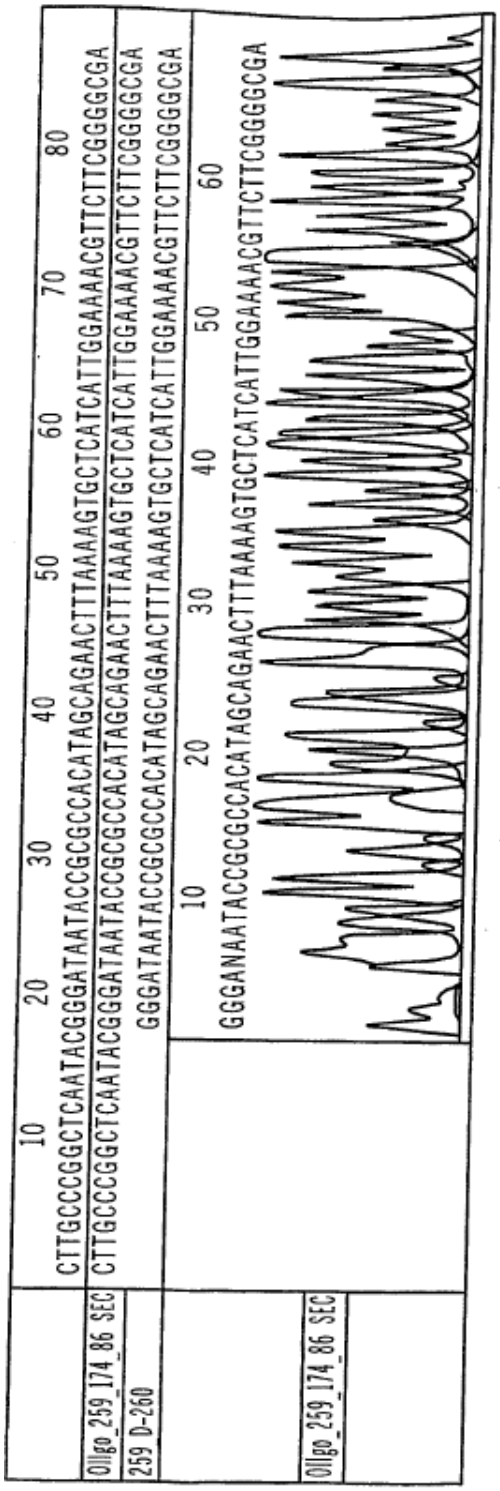

10 20 30 40 50 60 70 80
CTTGCCCGGCTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
Oligo_259_174_86 SEC
CTTGCCCGGCTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
259 D-260
GGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
10 20 30 40 50 60
GGGANAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
Oligo_259_174_86 SEC

FIG. 19

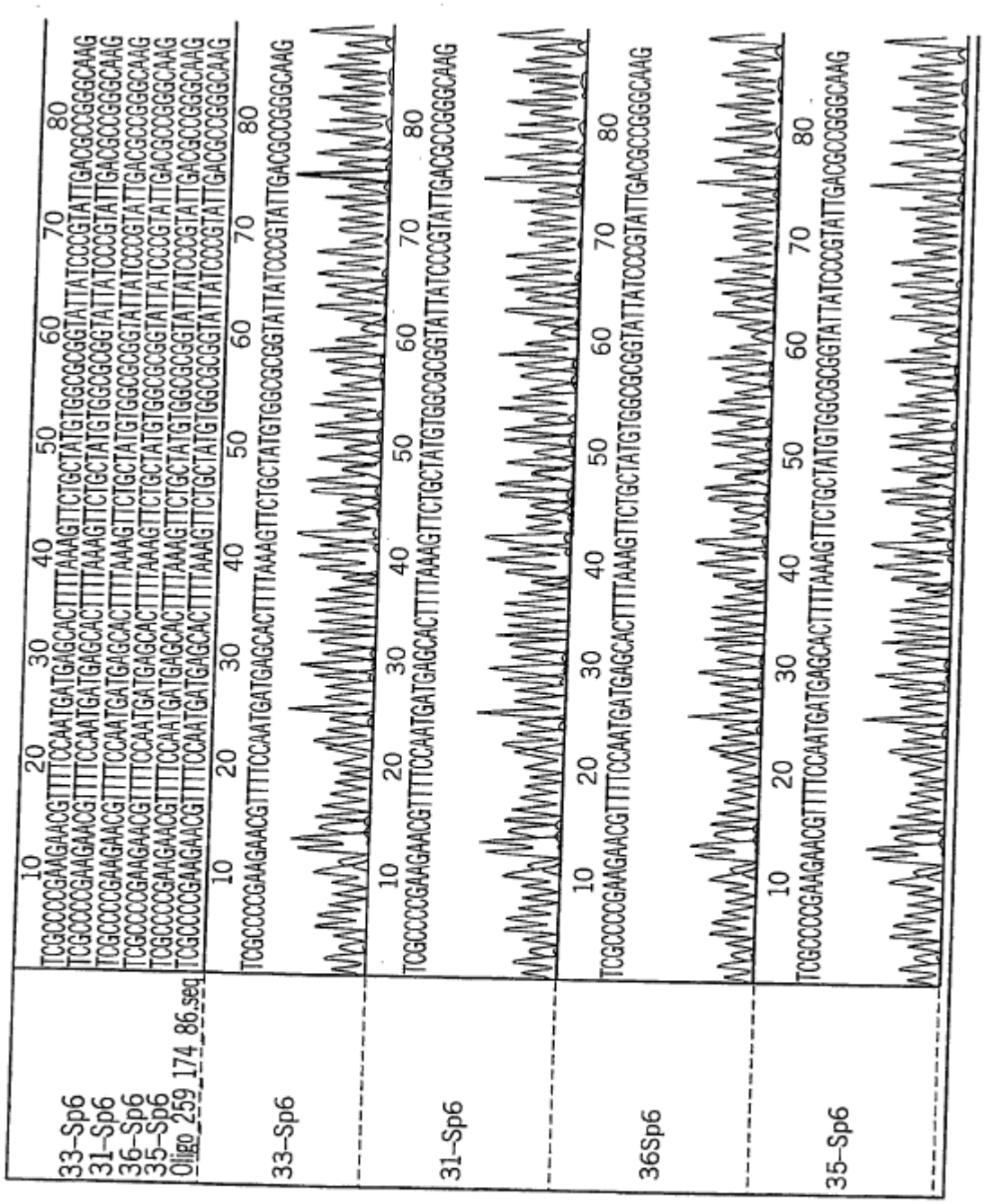

33–Sp6
31–Sp6
36–Sp6
35–Sp6
Oligo_259_174_86_seq
33–Sp6
31–Sp6
36Sp6
35–Sp6

FIG. 20

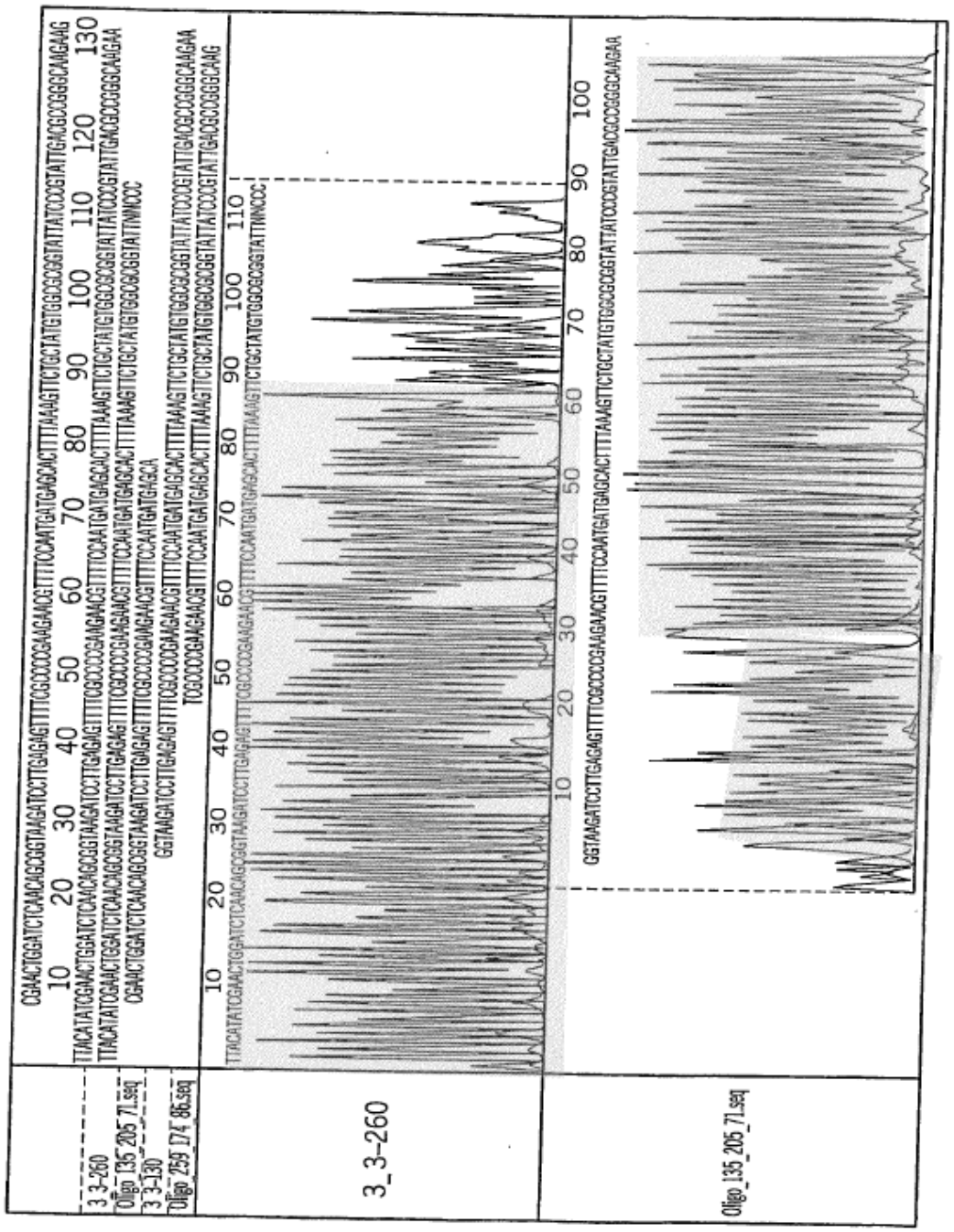


FIG. 21

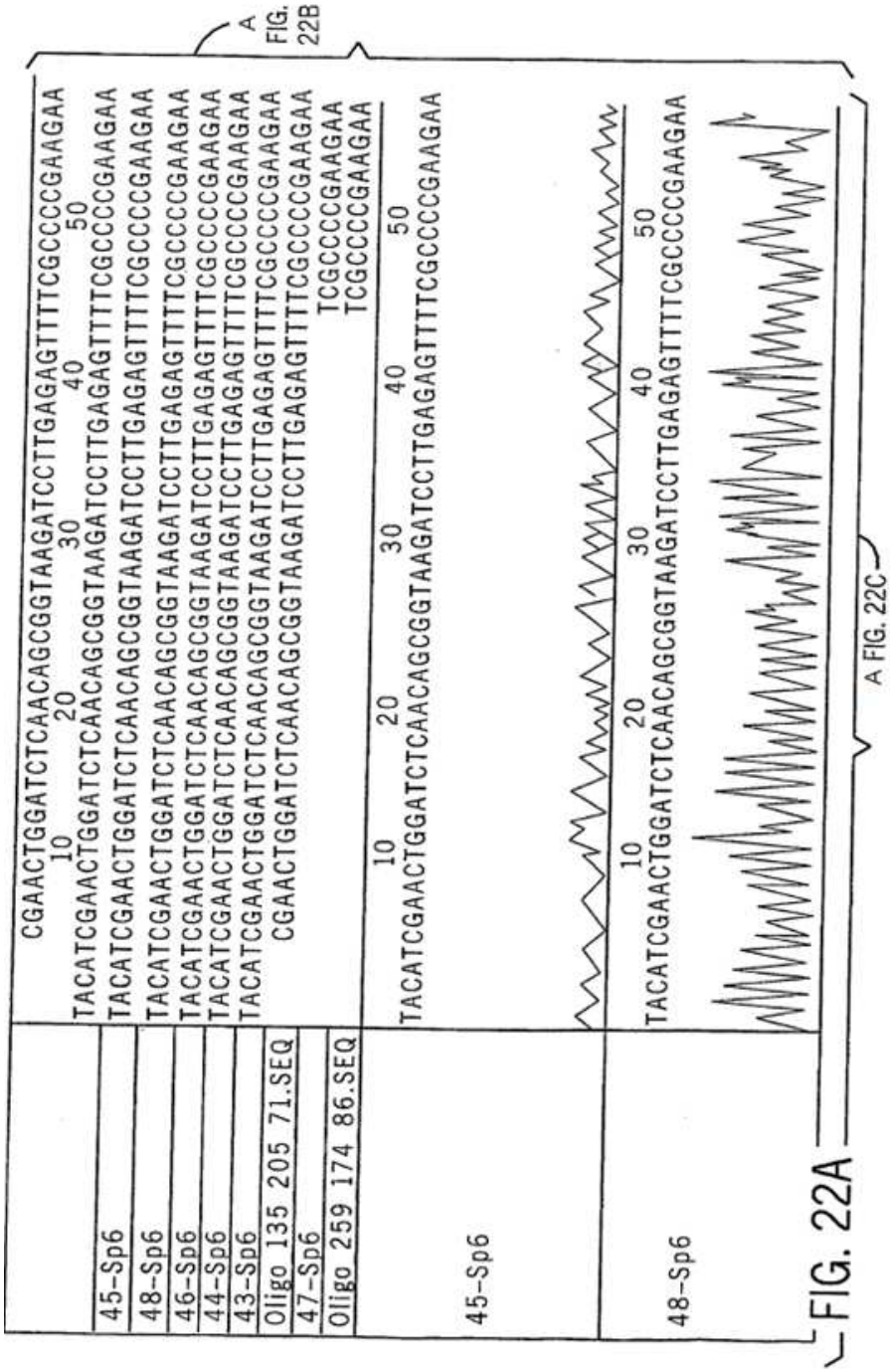
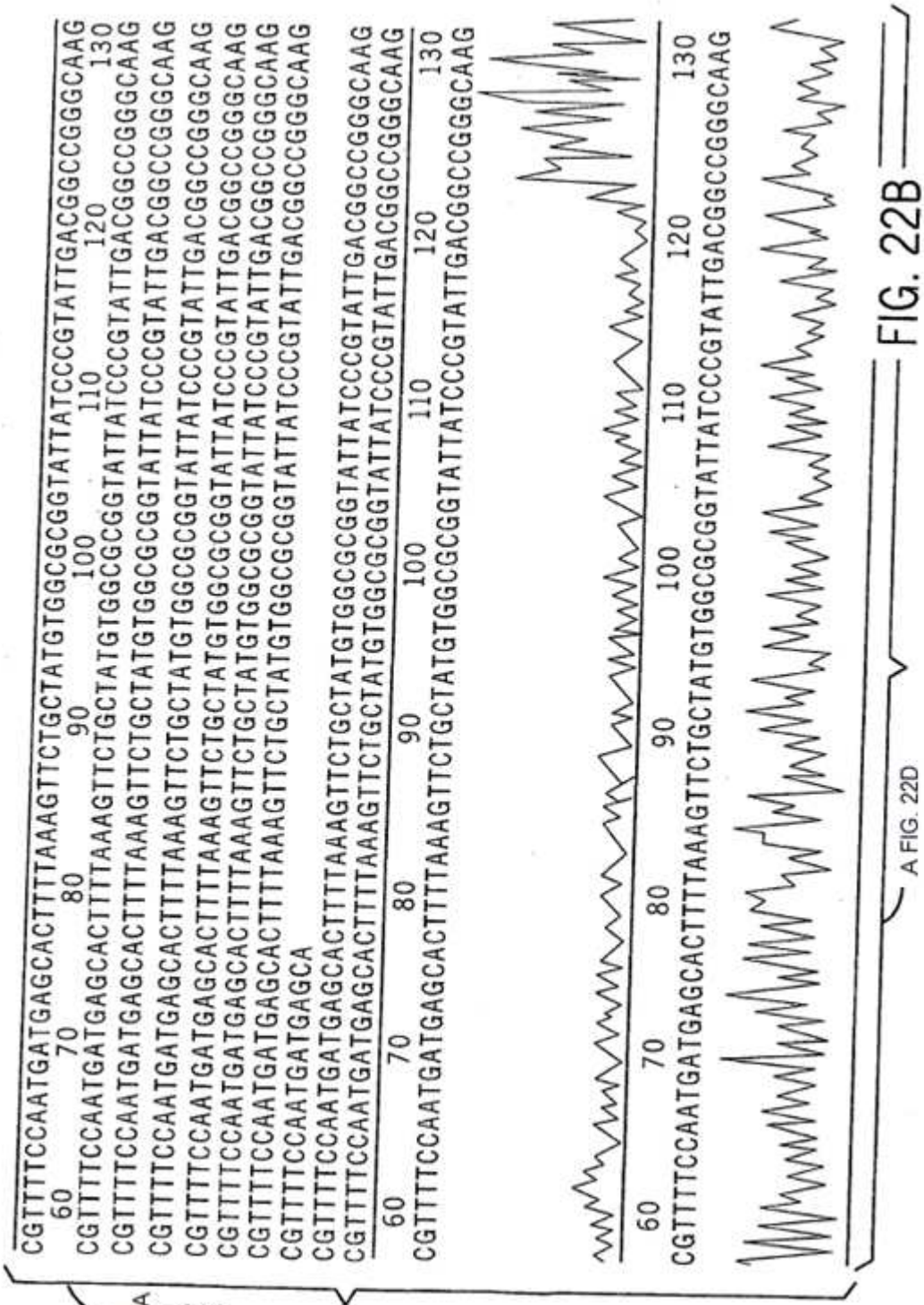

A FIG. 22B
45-Sp6
48-Sp6
46-Sp6
44-Sp6
43-Sp6
Oligo 135 205 71.SEQ
47-Sp6
Oligo 259 174 86.SEQ
45-Sp6
48-Sp6
10 20 30 40 50
CGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
CGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
TCGCCCCGAAGAA
TCGCCCCGAAGAA
10 20 30 40 50
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
10 20 30 40 50
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA
A FIG. 22C

FIG. 22A


DE LA FIG. 22A
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCA
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
A FIG. 22D
FIG. 22B

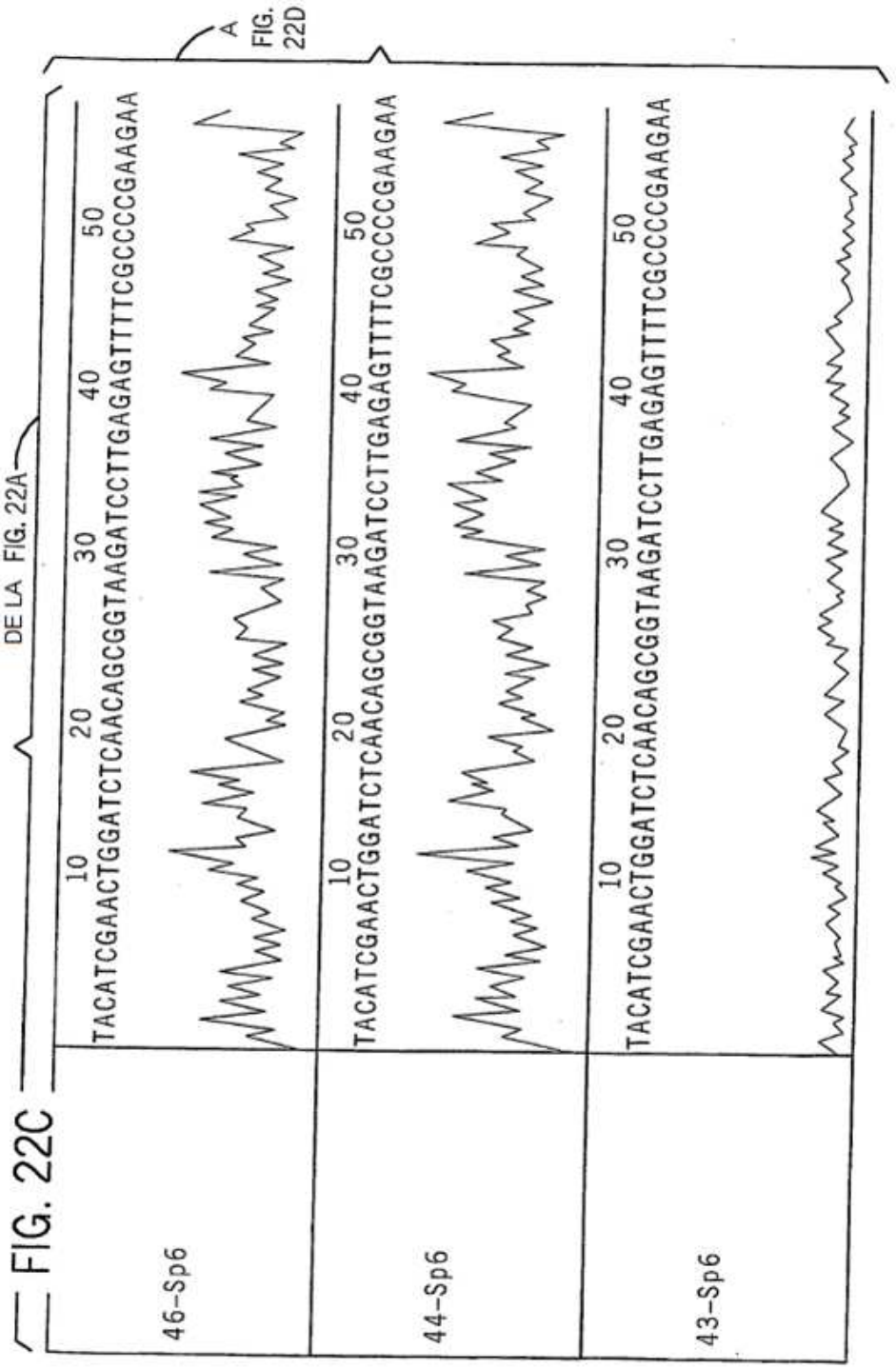
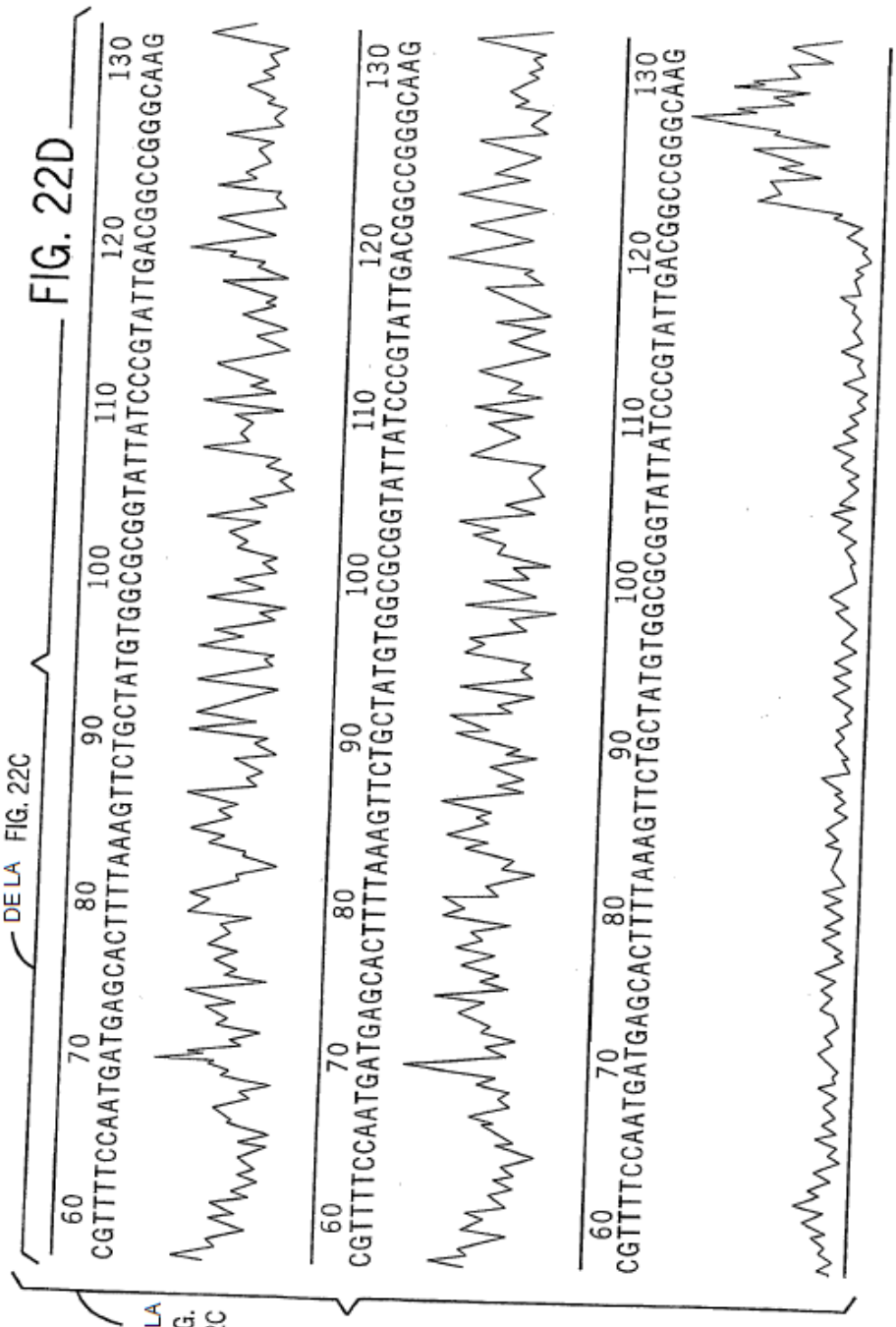


FIG. 22C
DE LA FIG. 22A
A
FIG. 22D
46-Sp6
44-Sp6
43-Sp6
10 20 30 40 50
TACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA


DE LA FIG. 22C
FIG. 22D
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
DE LA FIG. 22C
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG
60 70 80 90 100 110 120 130
CGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGGCCGGGCAAG

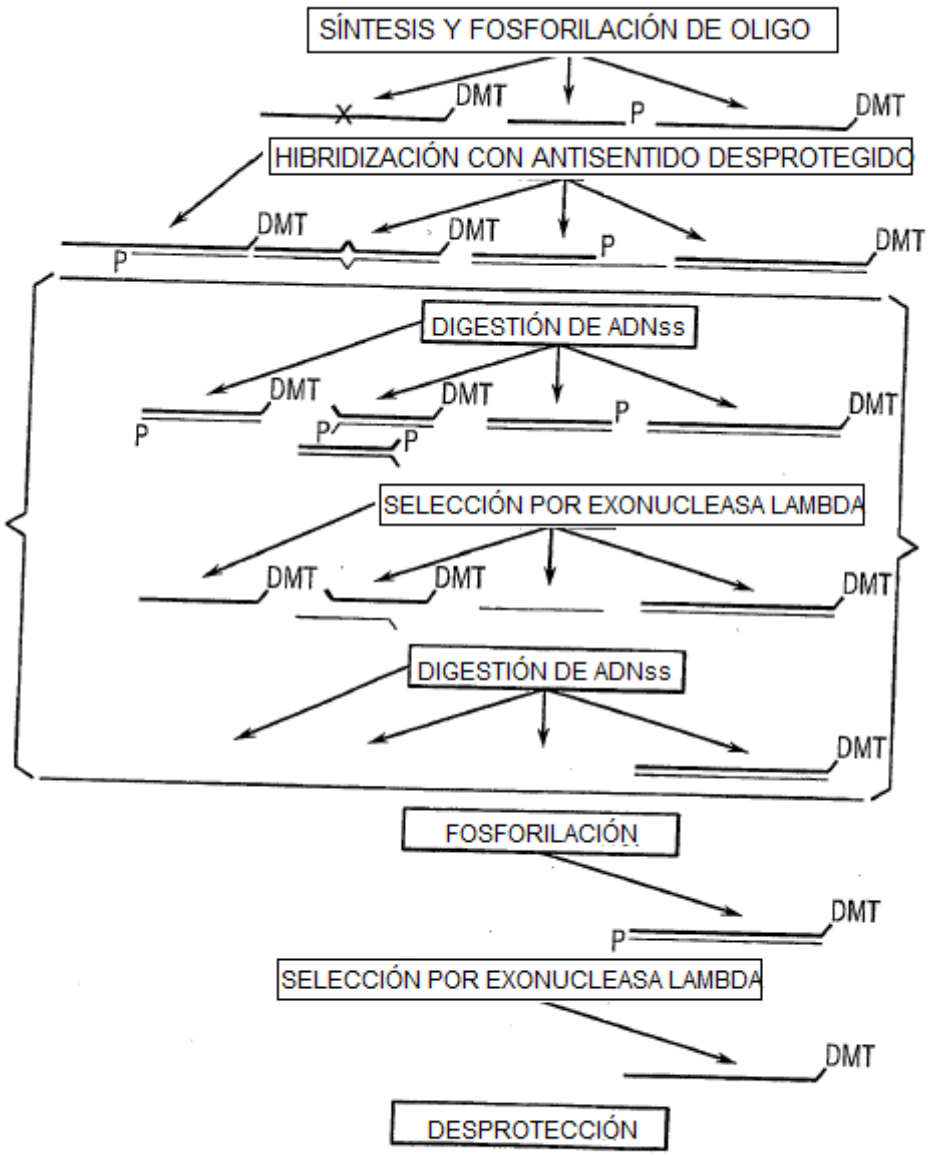

SÍNTESIS Y FOSFORILACIÓN DE OLIGO
DMT
P
DMT
HIBRIDIZACIÓN CON ANTISENTIDO DESPROTEGIDO
DMT
DMT
P
DMT
P
DIGESTIÓN DE ADNss
DMT
DMT
P
DMT
P
P
P
SELECCIÓN POR EXONUCLEASA LAMBDA
DMT
DMT
DMT
DIGESTIÓN DE ADNss
DMT
FOSFORILACIÓN
DMT
P
SELECCIÓN POR EXONUCLEASA LAMBDA
DMT
DESPROTECCIÓN

FIG. 23

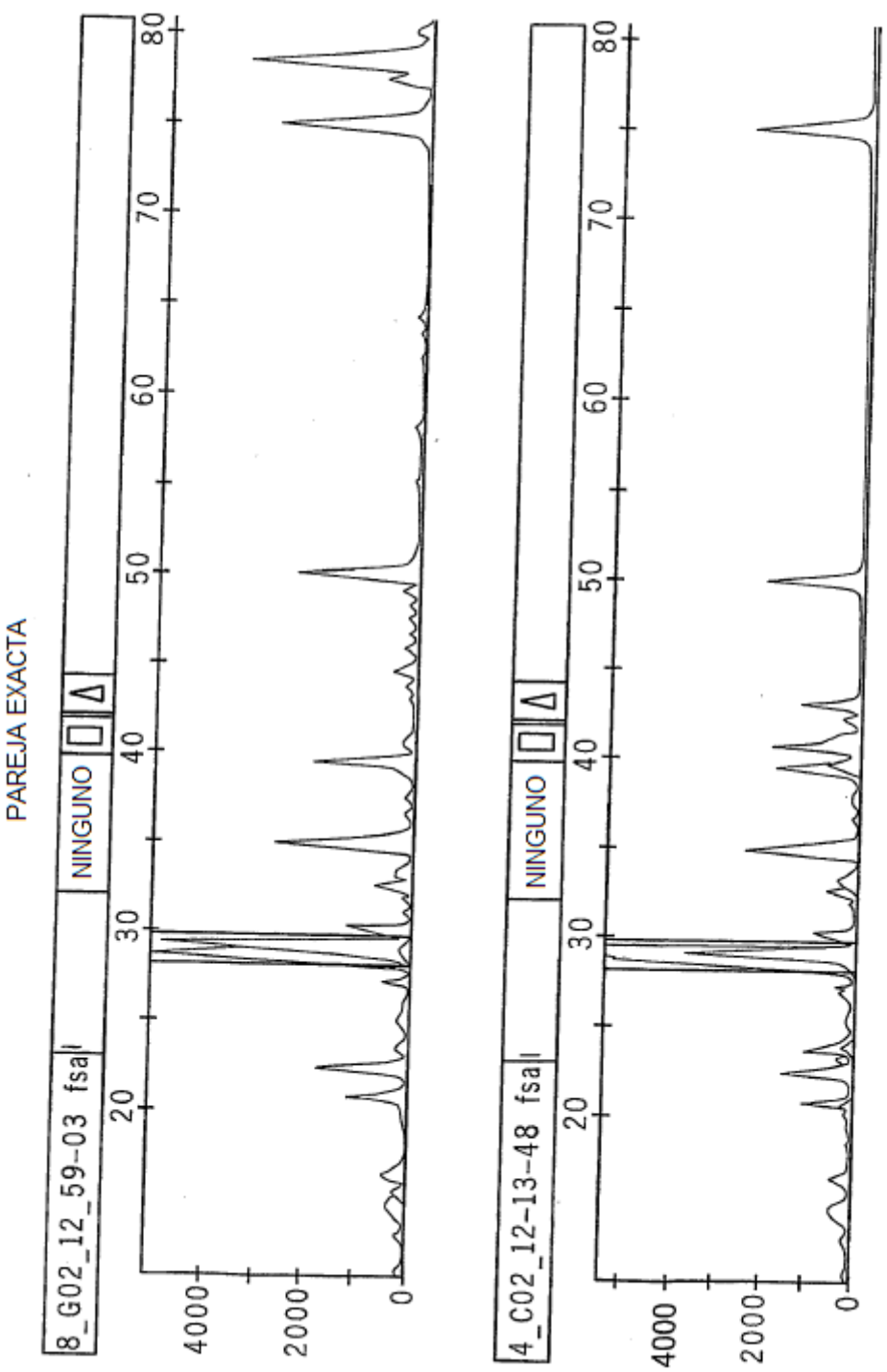

PAREJA EXACTA
8_G02_12_59-03 fsa
NINGUNO
Δ
20
30
40
50
60
70
80
4000
2000
0
4_C02_12-13-48 fsa
NINGUNO
Δ
20
30
40
50
60
70
80
4000
2000
0
DELECIÓN

FIG. 24

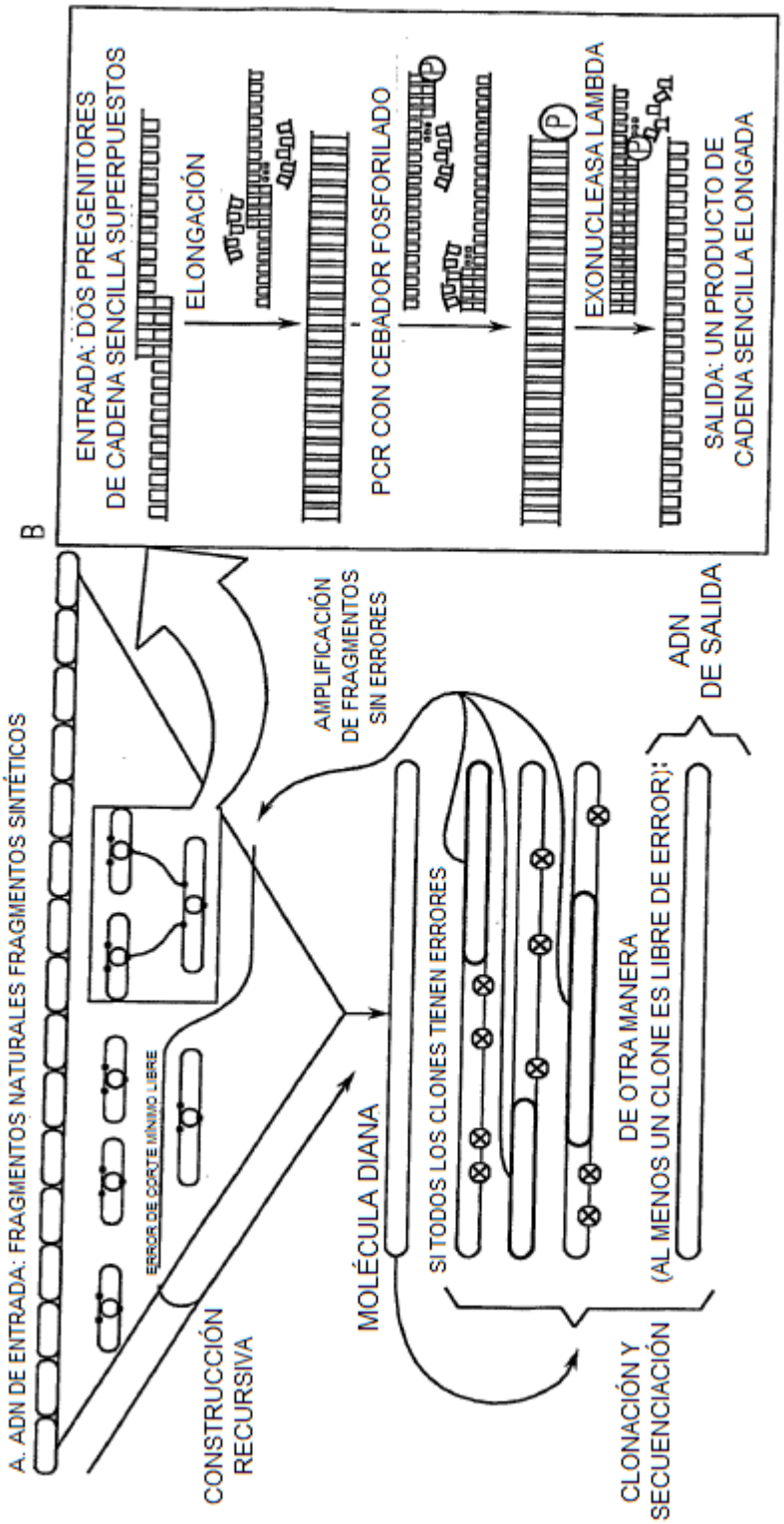
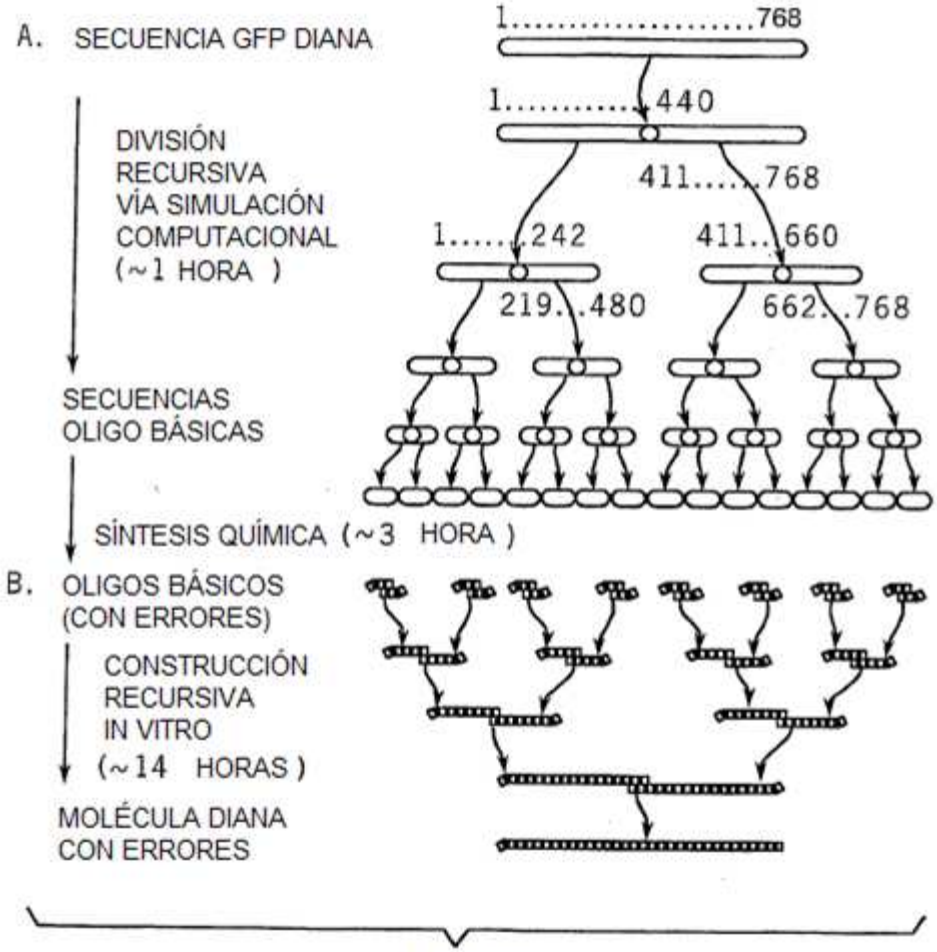

A. ADN DE ENTRADA: FRAGMENTOS NATURALES FRAGMENTOS SINTÉTICOS
ERROR DE CORTE MÍNIMO LIBRE
CONSTRUCCIÓN RECURSIVA
MOLÉCULA DIANA
SI TODOS LOS CLONES TIENEN ERRORES
DE OTRA MANERA
(AL MENOS UN CLONE ES LIBRE DE ERROR);
CLONACIÓN Y SECUENCIACIÓN
AMPLIFICACIÓN DE FRAGMENTOS SIN ERRORES
ADN DE SALIDA
B
ENTRADA: DOS PREGENITORES DE CADENA SENCILLA SUPERPUESTOS
ELONGACIÓN
PCR CON CEBADOR FOSFORILADO
EXONUCLEASA LAMBDA
SALIDA: UN PRODUCTO DE CADENA SENCILLA ELONGADA

FIG. 25

FIG. 26A



A FIGURA 26B

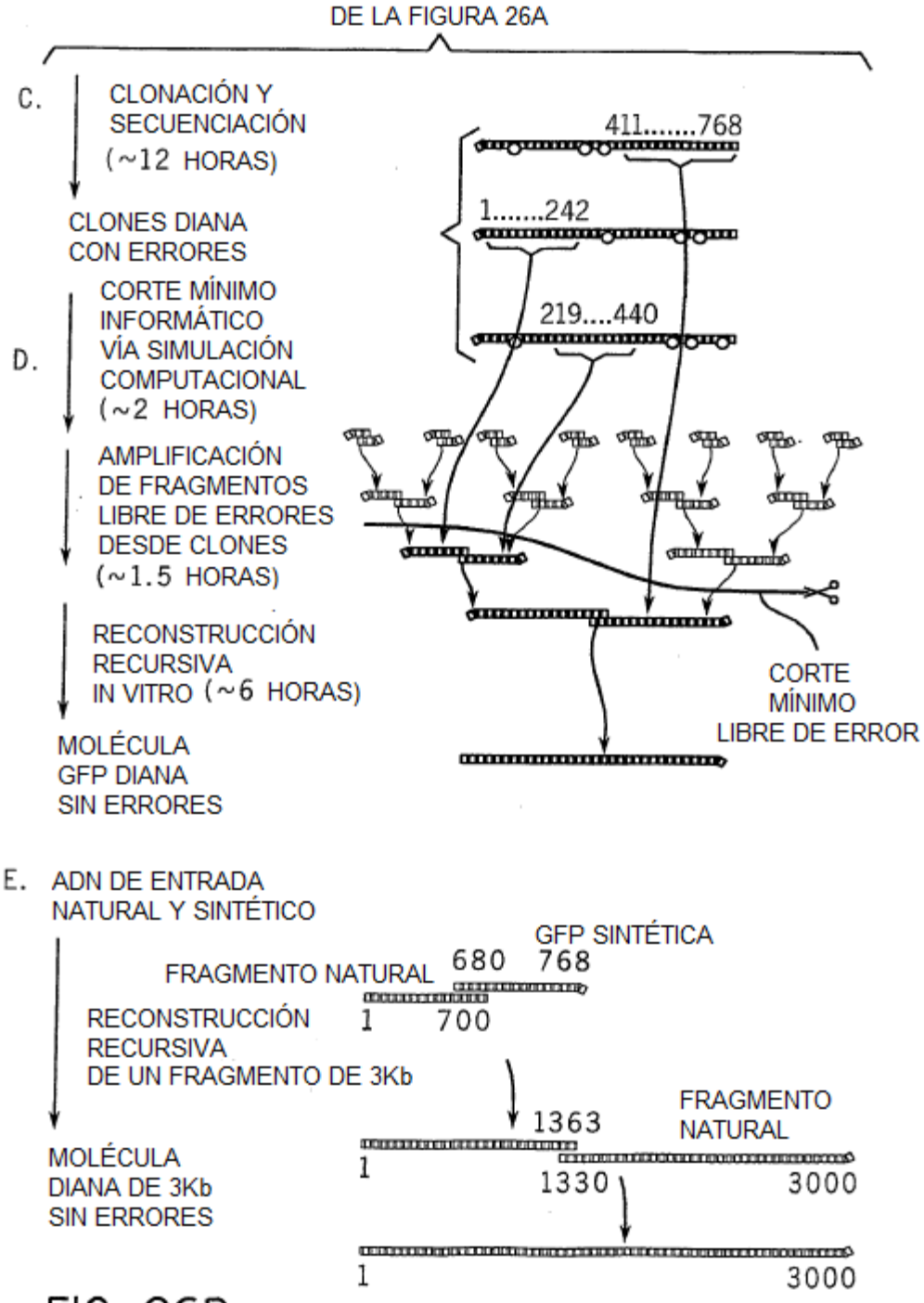

DE LA FIGURA 26A
C.
CLONACIÓN Y SECUENCIACIÓN (~12 HORAS)
CLONES DIANA CON ERRORES
D.
CORTE MÍNIMO INFORMÁTICO VÍA SIMULACIÓN COMPUTACIONAL (~2 HORAS)
AMPLIFICACIÓN DE FRAGMENTOS LIBRE DE ERRORES DESDE CLONES (~1.5 HORAS)
RECONSTRUCCIÓN RECURSIVA IN VITRO (~6 HORAS)
MOLÉCULA GFP DIANA SIN ERRORES
411.....768
1.....242
219....440
CORTE MÍNIMO LIBRE DE ERROR
E.
ADN DE ENTRADA NATURAL Y SINTÉTICO
GFP SINTÉTICA
FRAGMENTO NATURAL
680
768
1
700
RECONSTRUCCIÓN RECURSIVA DE UN FRAGMENTO DE 3Kb
1363
FRAGMENTO NATURAL
1
1330
3000
MOLÉCULA DIANA DE 3Kb SIN ERRORES
1
3000

FIG. 26B

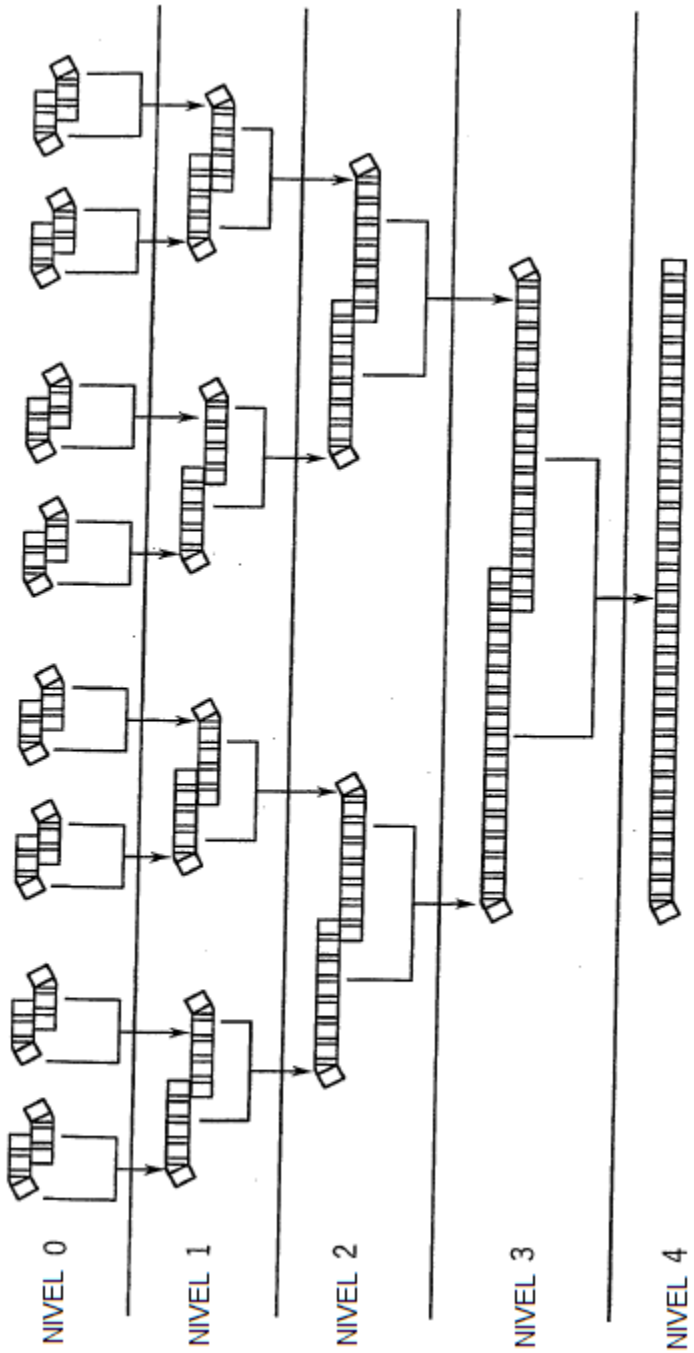


FIG. 27A

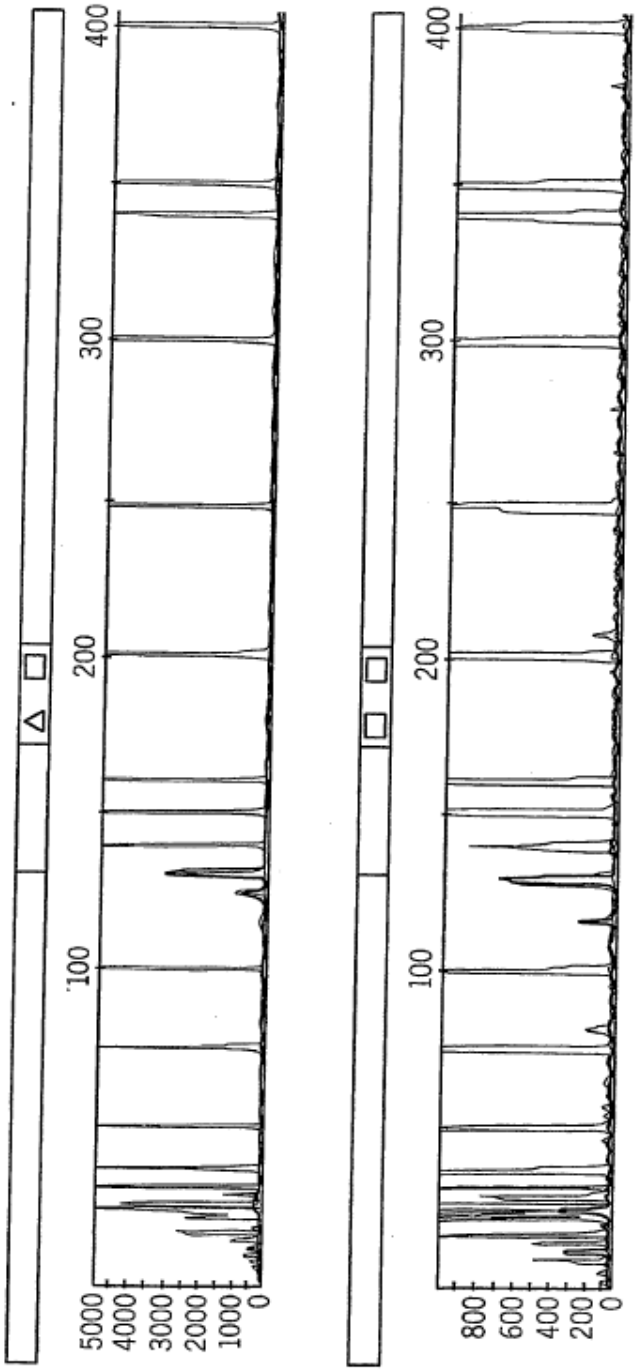

100
200
300
400
5000
4000
3000
2000
1000
0
100
200
300
400
800
600
400
200
0

FIG. 27Ba

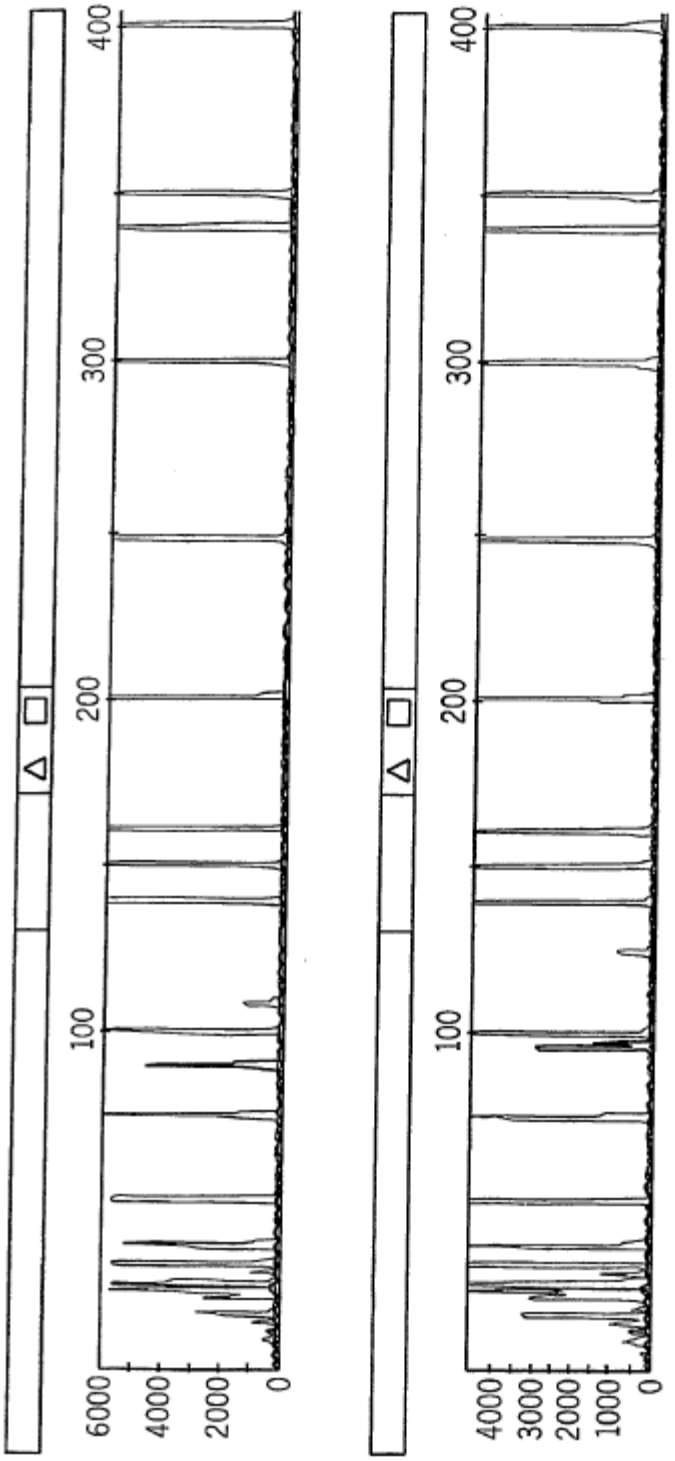

100
200
300
400
6000
4000
2000
0
100
200
300
400
4000
3000
2000
1000
0

FIG. 27Bb

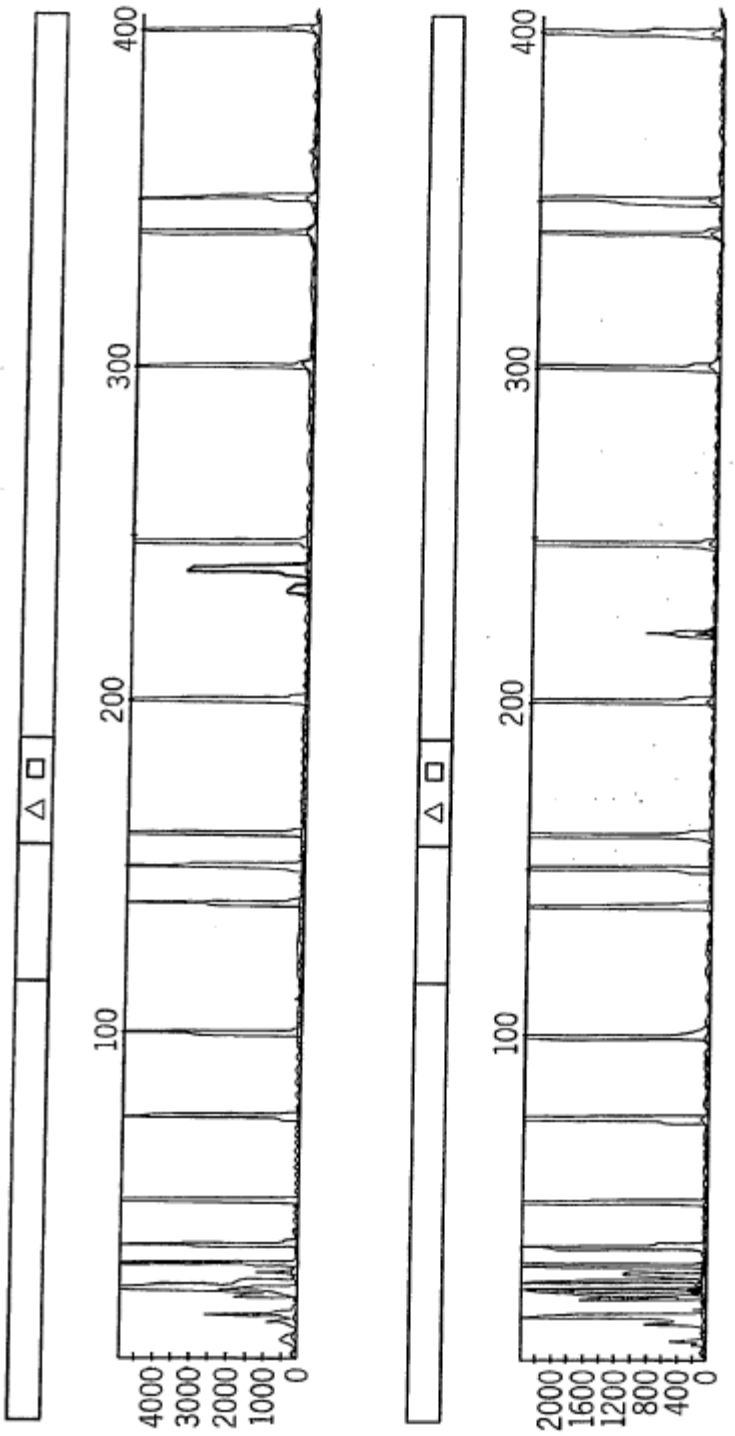

100
200
300
400
4000
3000
2000
1000
0
100
200
300
400
2000
1600
1200
800
400
0

FIG. 27Ca

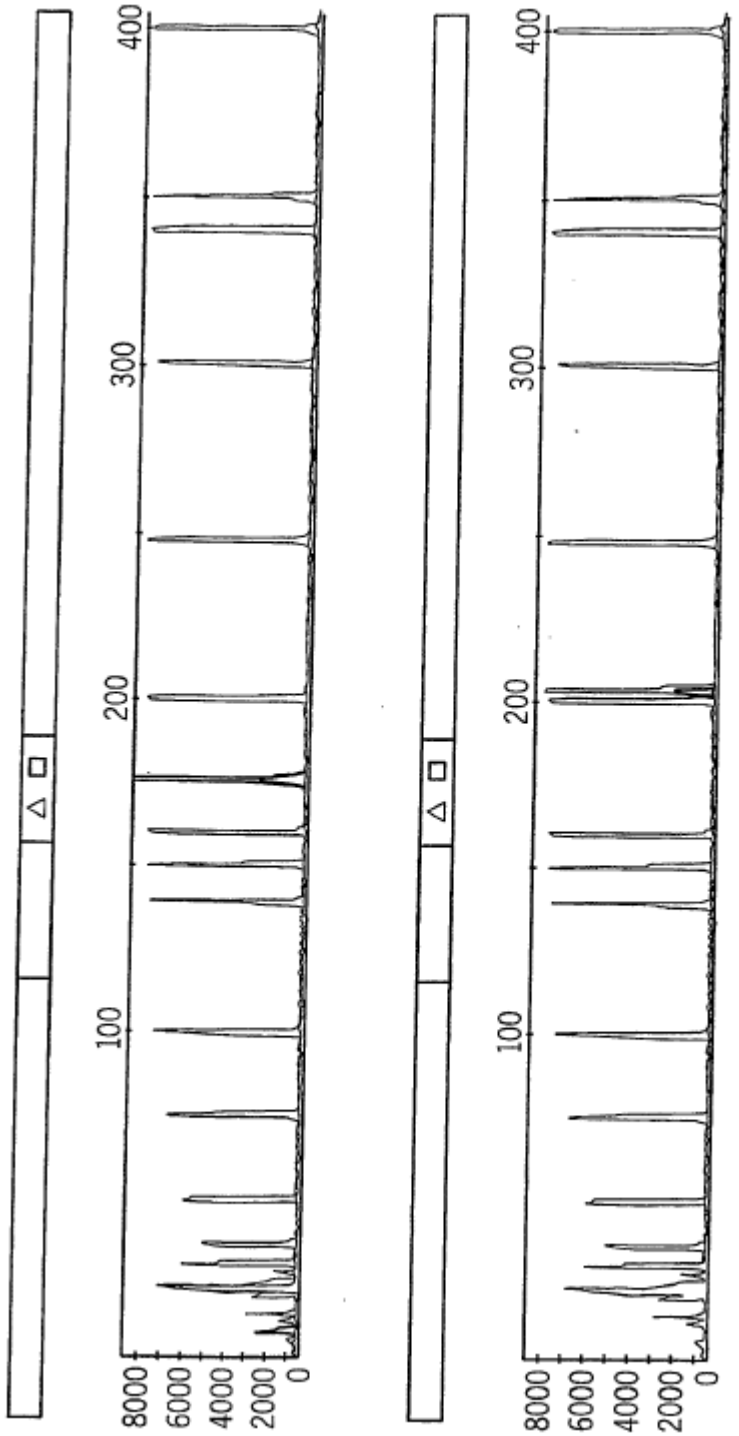


FIG. 27Cb

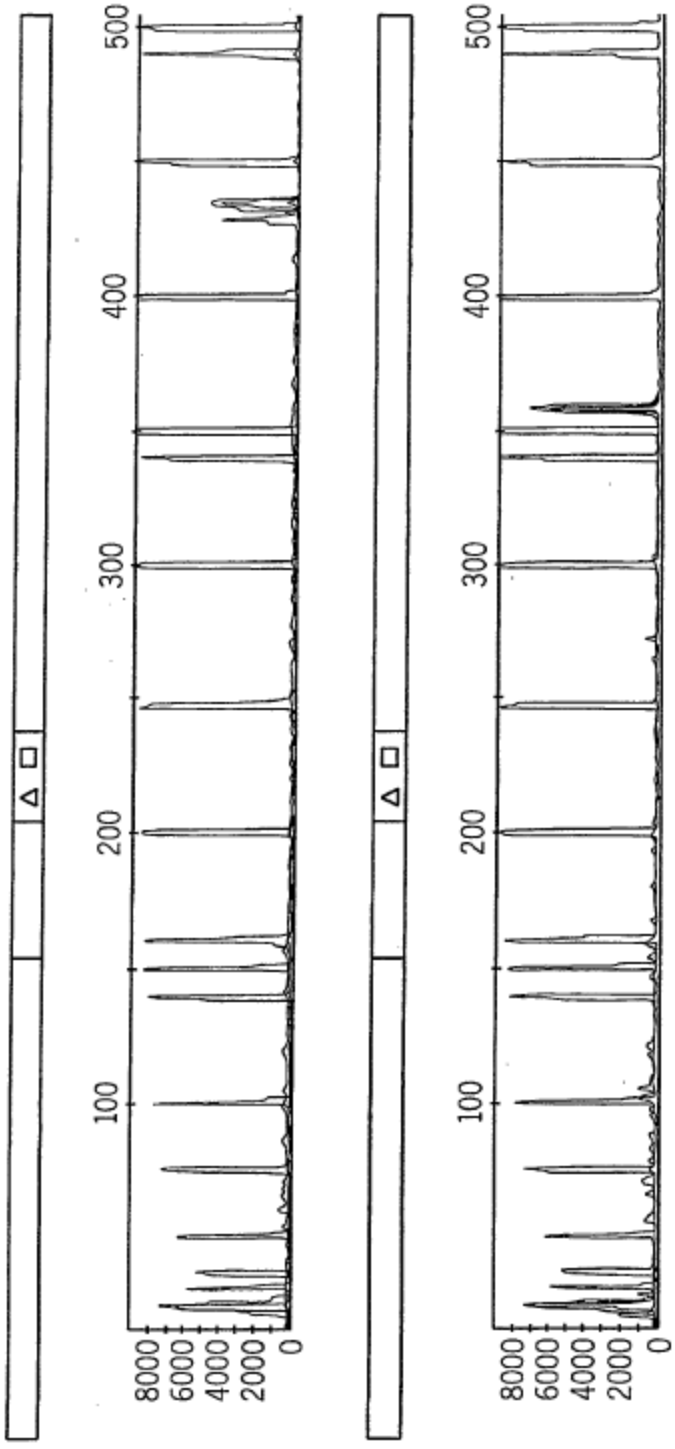


FIG. 27D

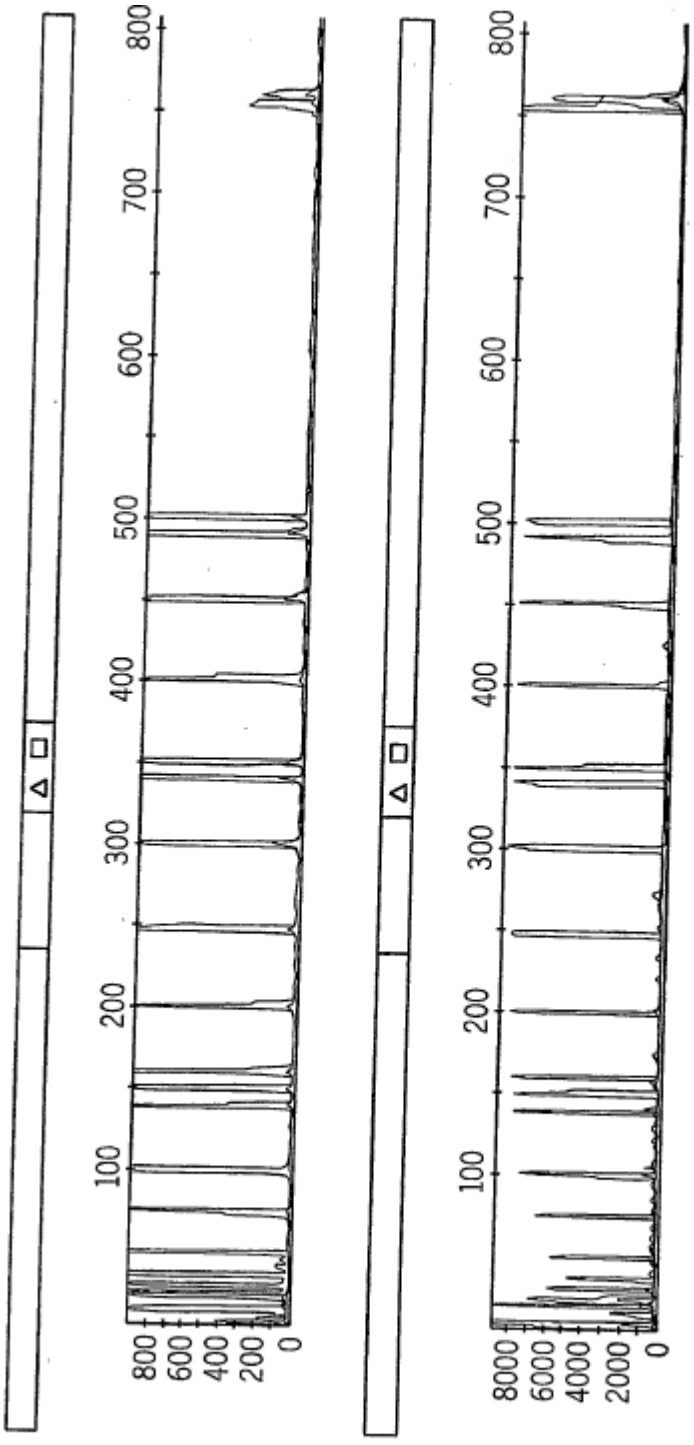

800
700
600
500
400
300
200
100
800
600
400
200
0
8000
6000
4000
2000
0

FIG. 27E

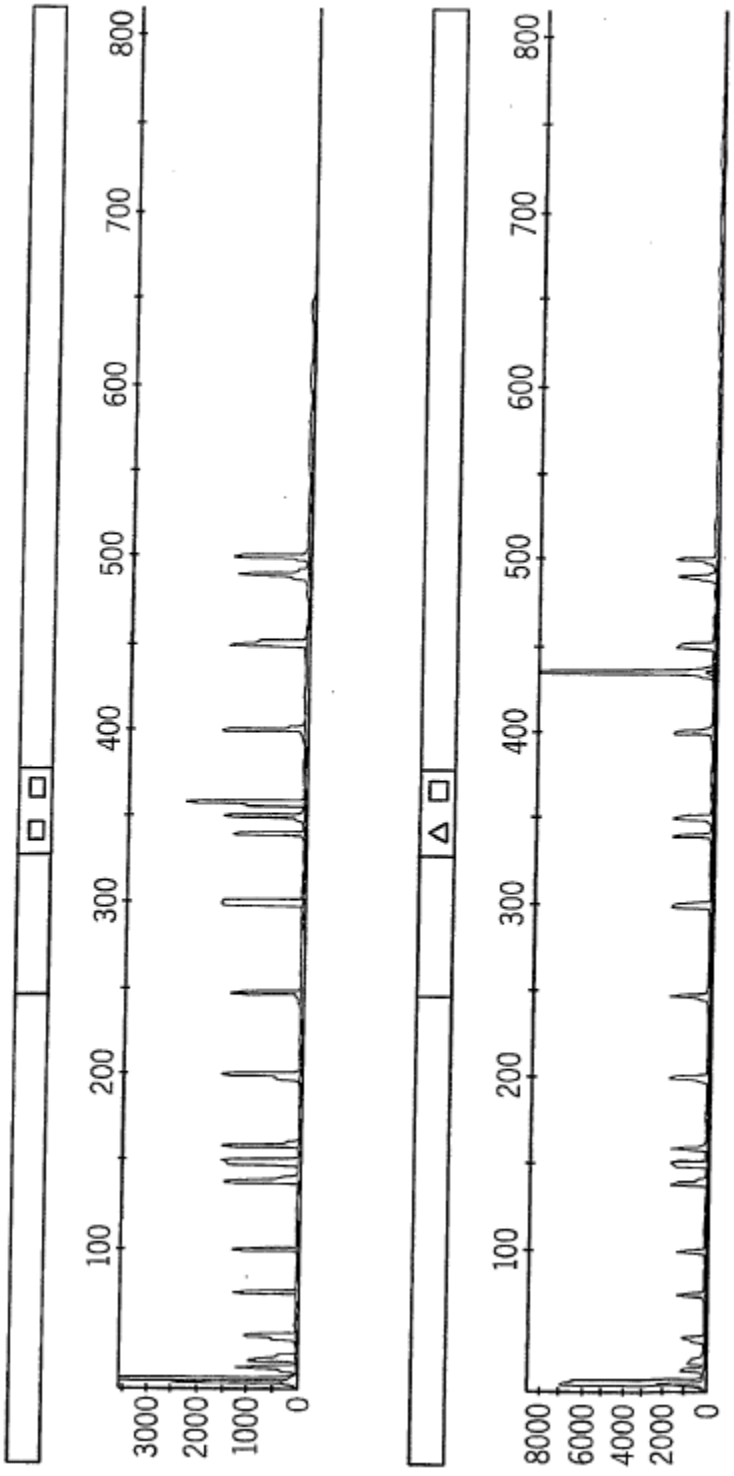

100
200
300
400
500
600
700
800
3000
2000
1000
0
100
200
300
400
500
600
700
800
8000
6000
4000
2000
0

FIG. 28A

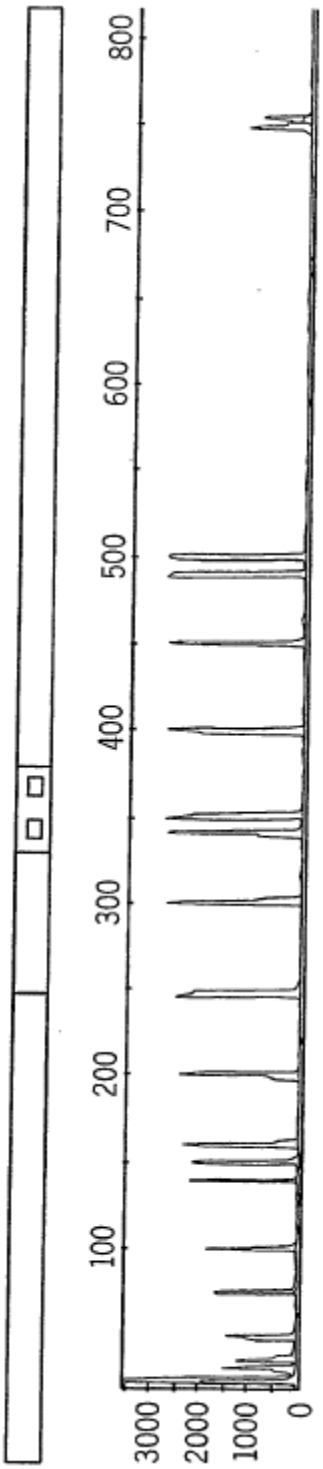


FIG. 28B

NÚMERO DE CLONES
LONGITUD DEL FRAGMENTO
SIN CORRECCIÓN DE ERRORES
PURIFICACIÓN DEL GEL
HIBRIDIZACIÓN PURIFICACIÓN EN CHIP DE ADN
CONSTRUCCIÓN RECURSIVA Y
CORRECCIÓN DE ERROR ITERATIVA
0
5
10
15
20
25
30
35
40
45
50
0
1000
2000
3000
4000
5000

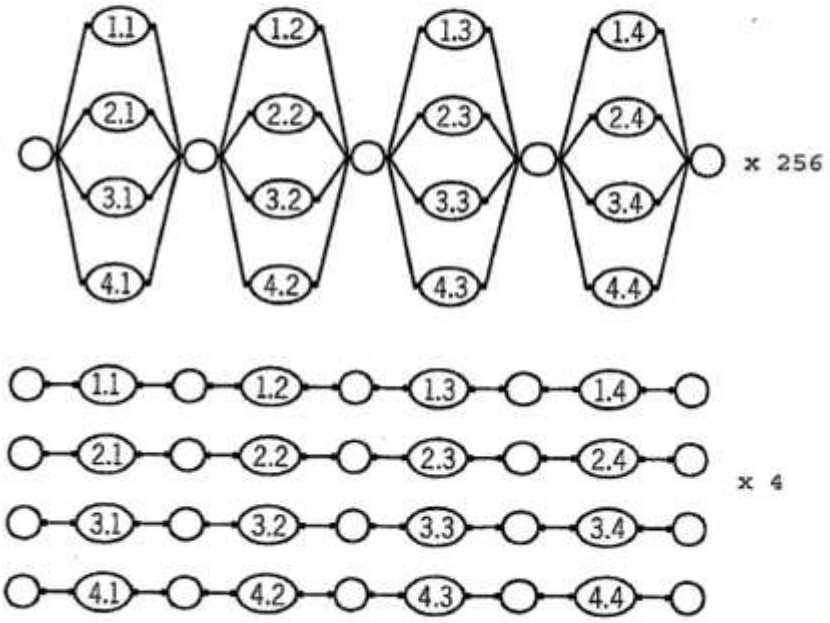

1.1
1.2
1.3
1.4
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
x 256
x 4

FIG. 29

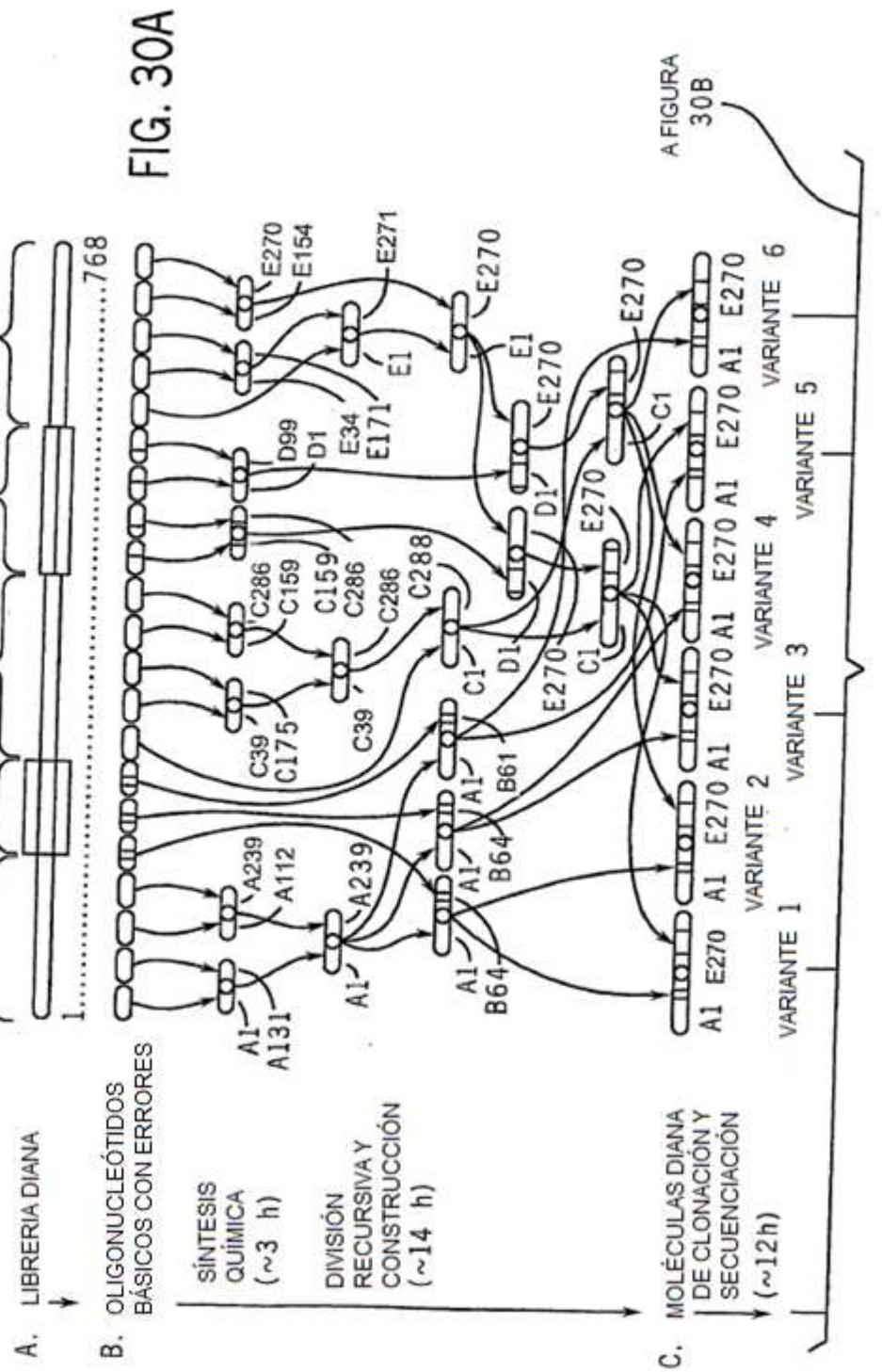

A. LIBRERIA DIANA
B. OLIGONUCLEÓTIDOS BÁSICOS CON ERRORES
SÍNTESIS QUÍMICA (~3 h)
DIVISIÓN RECURSIVA Y CONSTRUCCIÓN (~14 h)
C. MOLÉCULAS DIANA DE CLONACIÓN Y SECUENCIACIÓN (~12h)
A
B
C
D
E
1
768
A1
A131
A239
A112
A239
A1
A1
B64
A1
B64
A1
B61
C1
C39
C175
C286
C159
C39
C286
C159
C286
C288
D1
D99
D1
E34
E171
E1
E271
E270
E154
E270
E1
E270
E270
E270
C1
C1
E270
E270
A1
E270
A1
E270
A1
E270
A1
E270
A1
E270
A1
E270
A1
E270
VARIANTE 1
VARIANTE 2
VARIANTE 3
VARIANTE 4
VARIANTE 5
VARIANTE 6
A FIGURA 30B

FIG. 30A

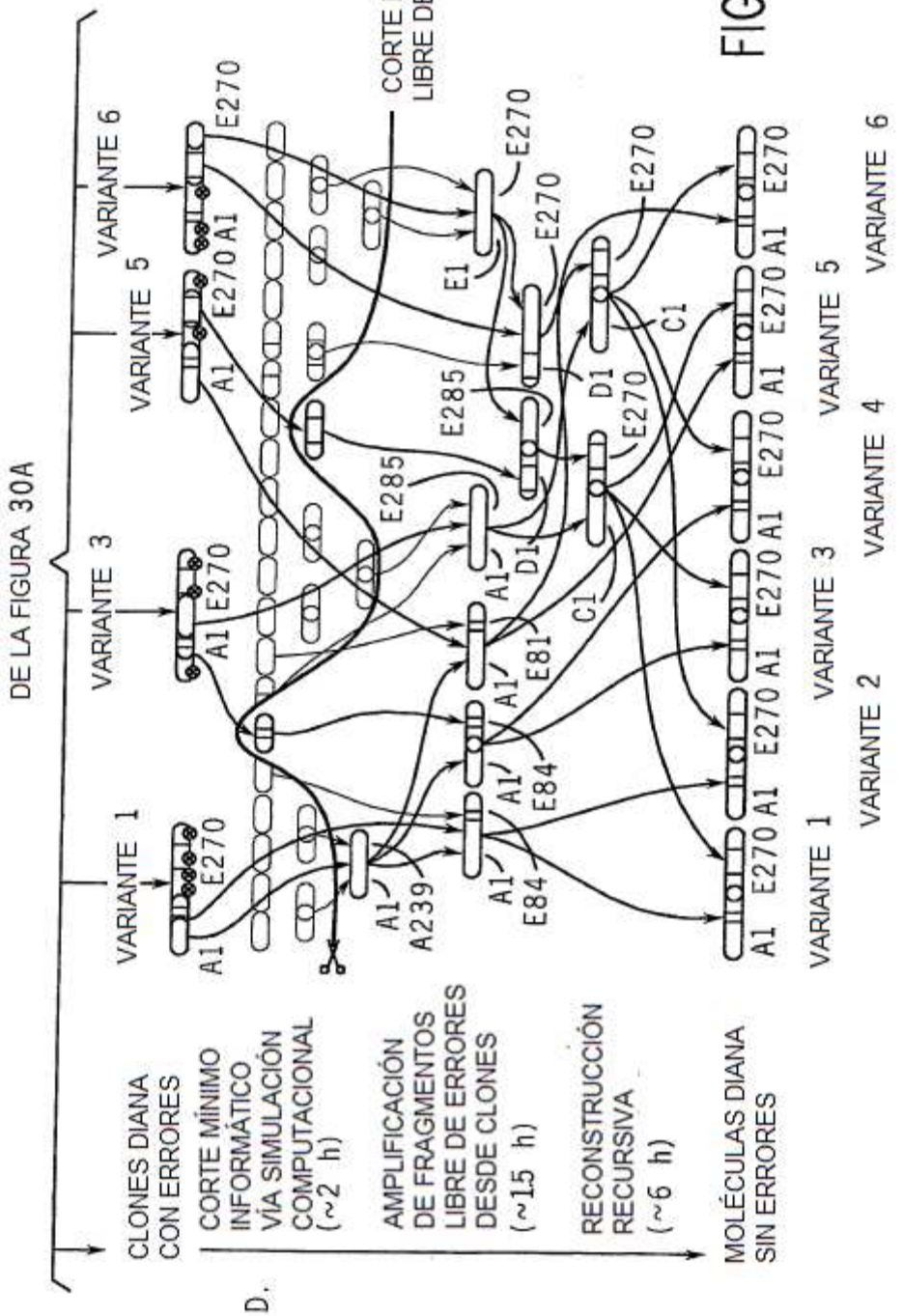

DE LA FIGURA 30A
CLONES DIANA CON ERRORES
CORTE MÍNIMO INFORMÁTICO VÍA SIMULACIÓN COMPUTACIONAL (~2 h)
AMPLIFICACIÓN DE FRAGMENTOS LIBRE DE ERRORES DESDE CLONES (~15 h)
RECONSTRUCCIÓN RECURSIVA (~6 h)
MOLÉCULAS DIANA SIN ERRORES
D.
VARIANTE 1
VARIANTE 2
VARIANTE 3
VARIANTE 4
VARIANTE 5
VARIANTE 6
CORTE MÍNIMO LIBRE DE ERROR
A1
E270
A239
E84
E81
E285
E1
D1
C1

FIG. 30B

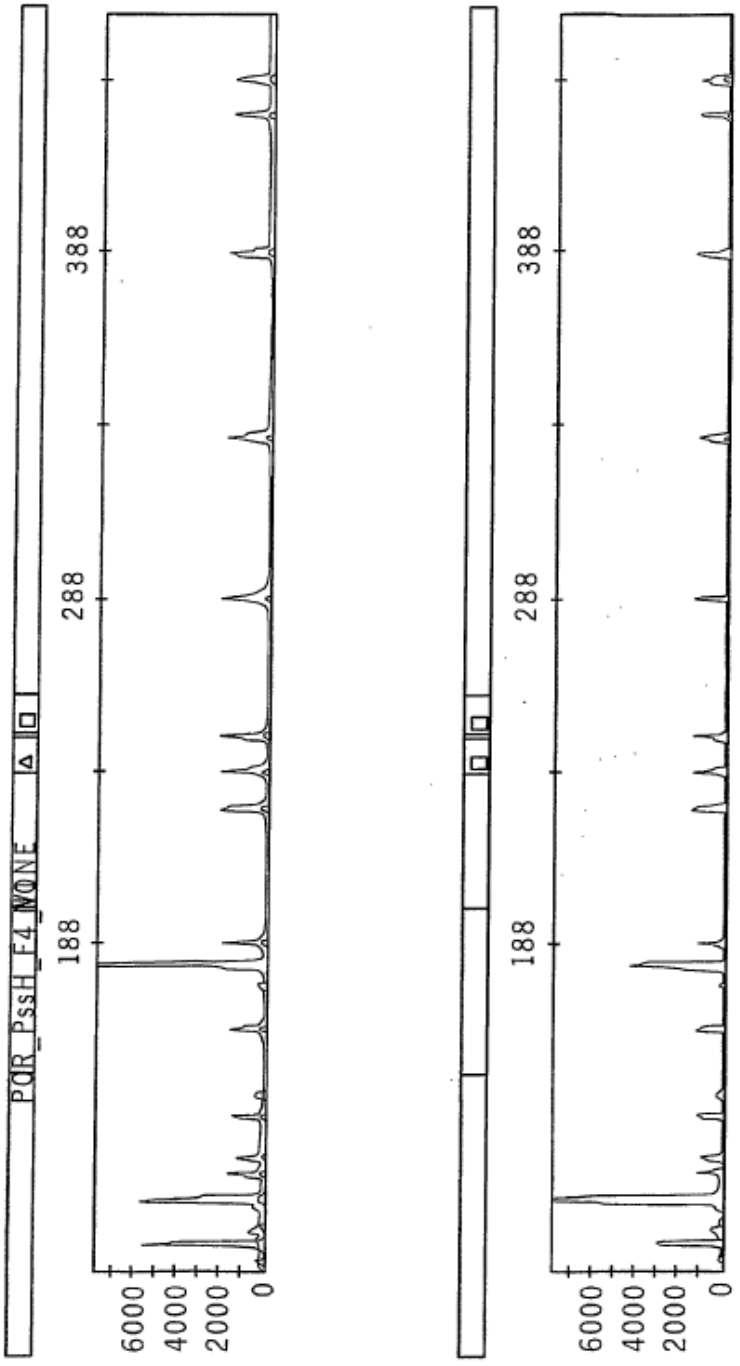

POR_PssH_F4_NONE
188
288
388
6000
4000
2000
0
188
288
388
6000
4000
2000
0

FIG. 31Aa

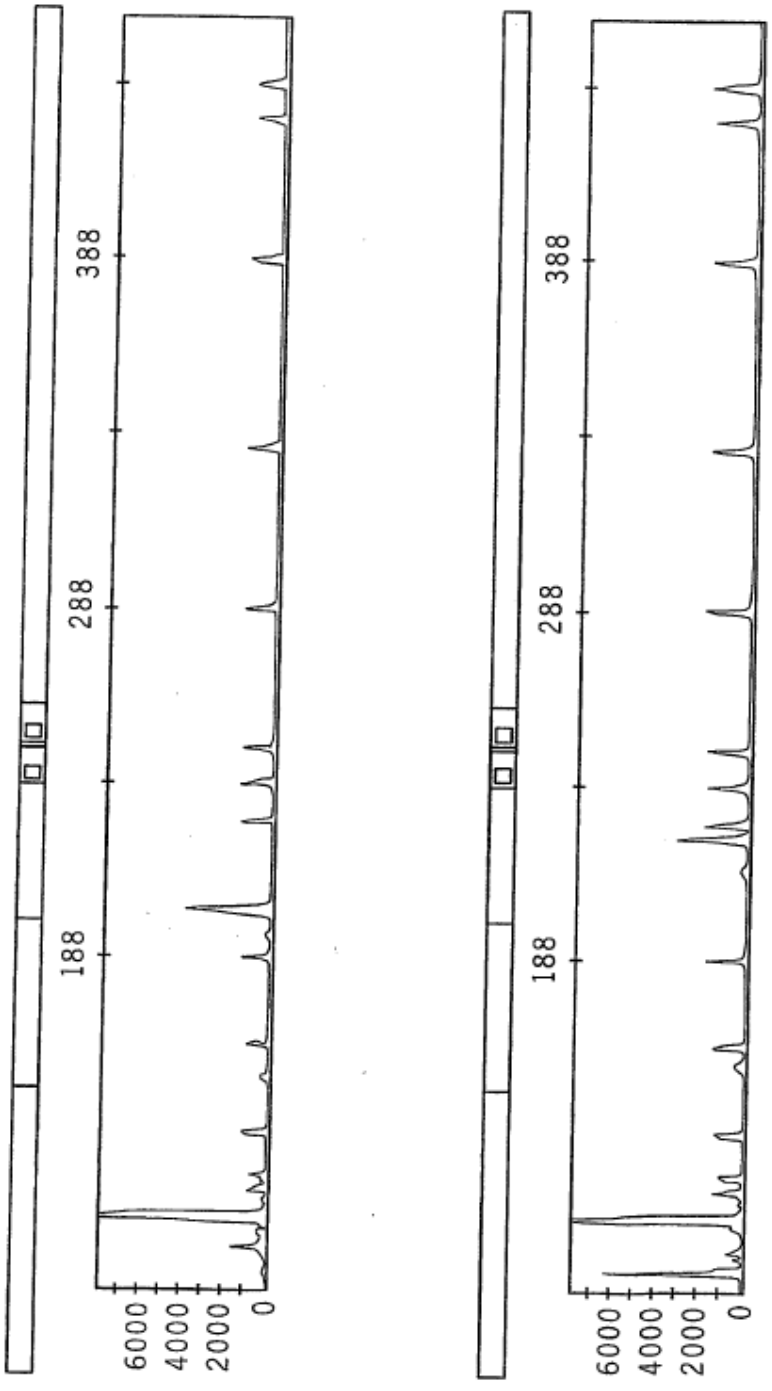


FIG. 31Ab

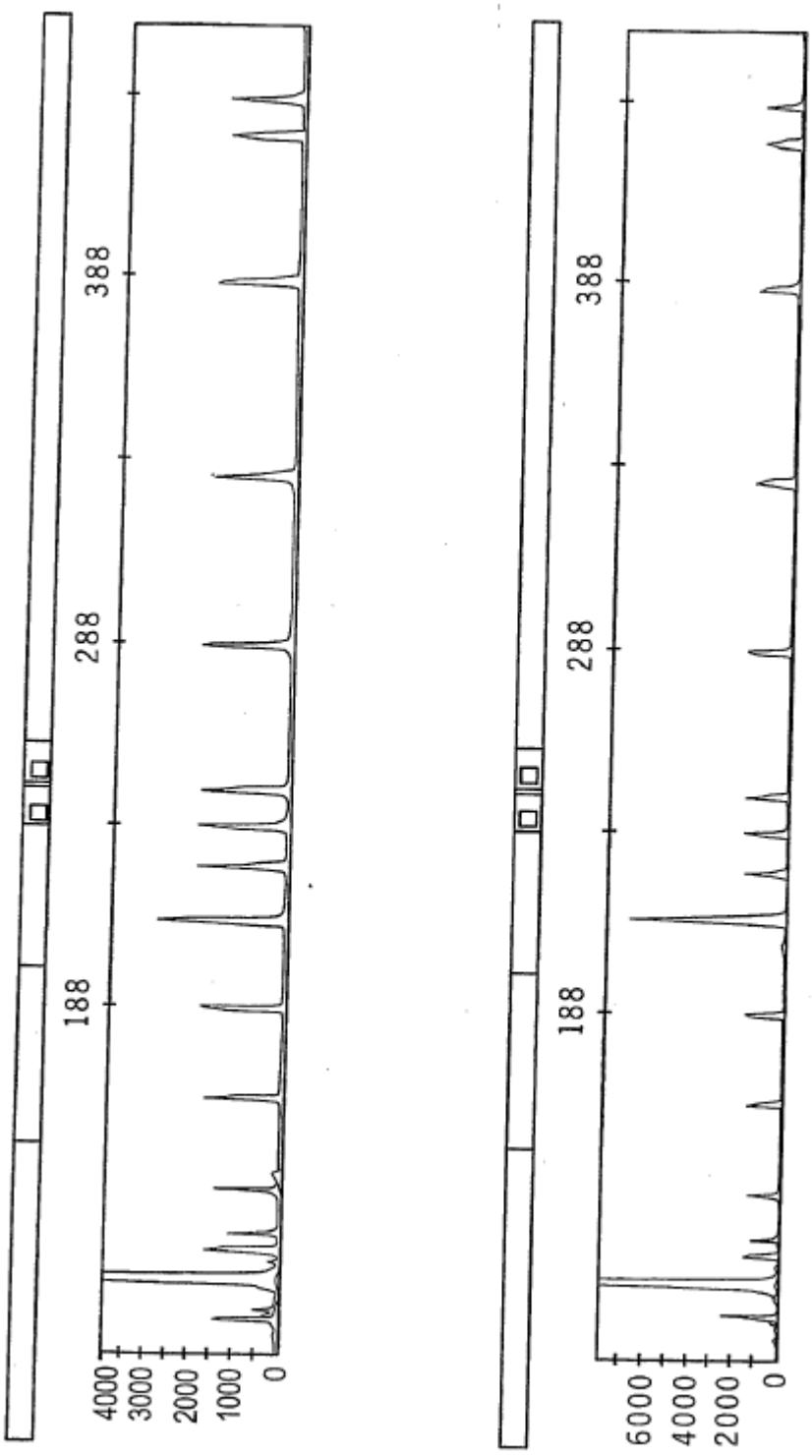


FIG. 31Ac

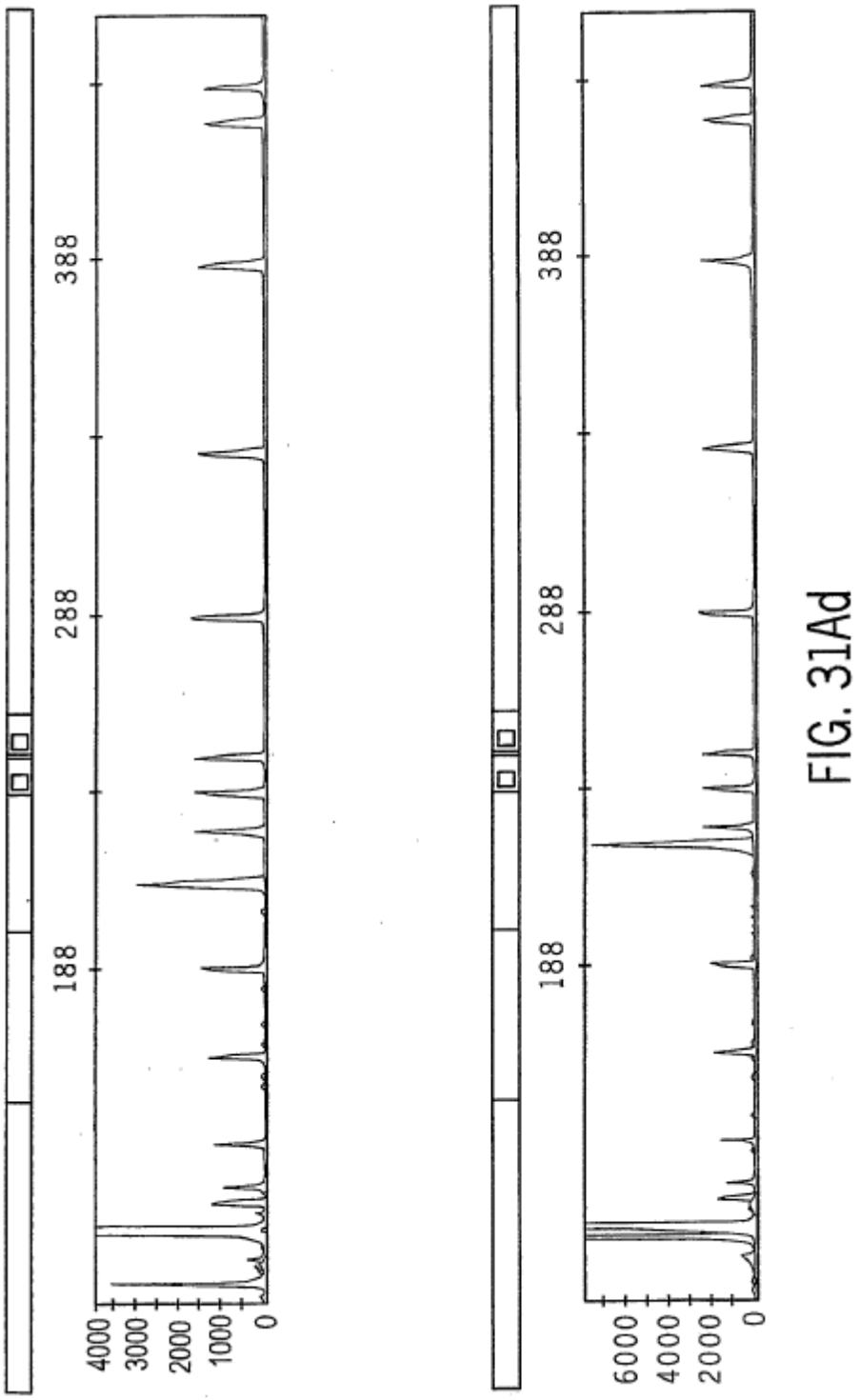


FIG. 31Ad

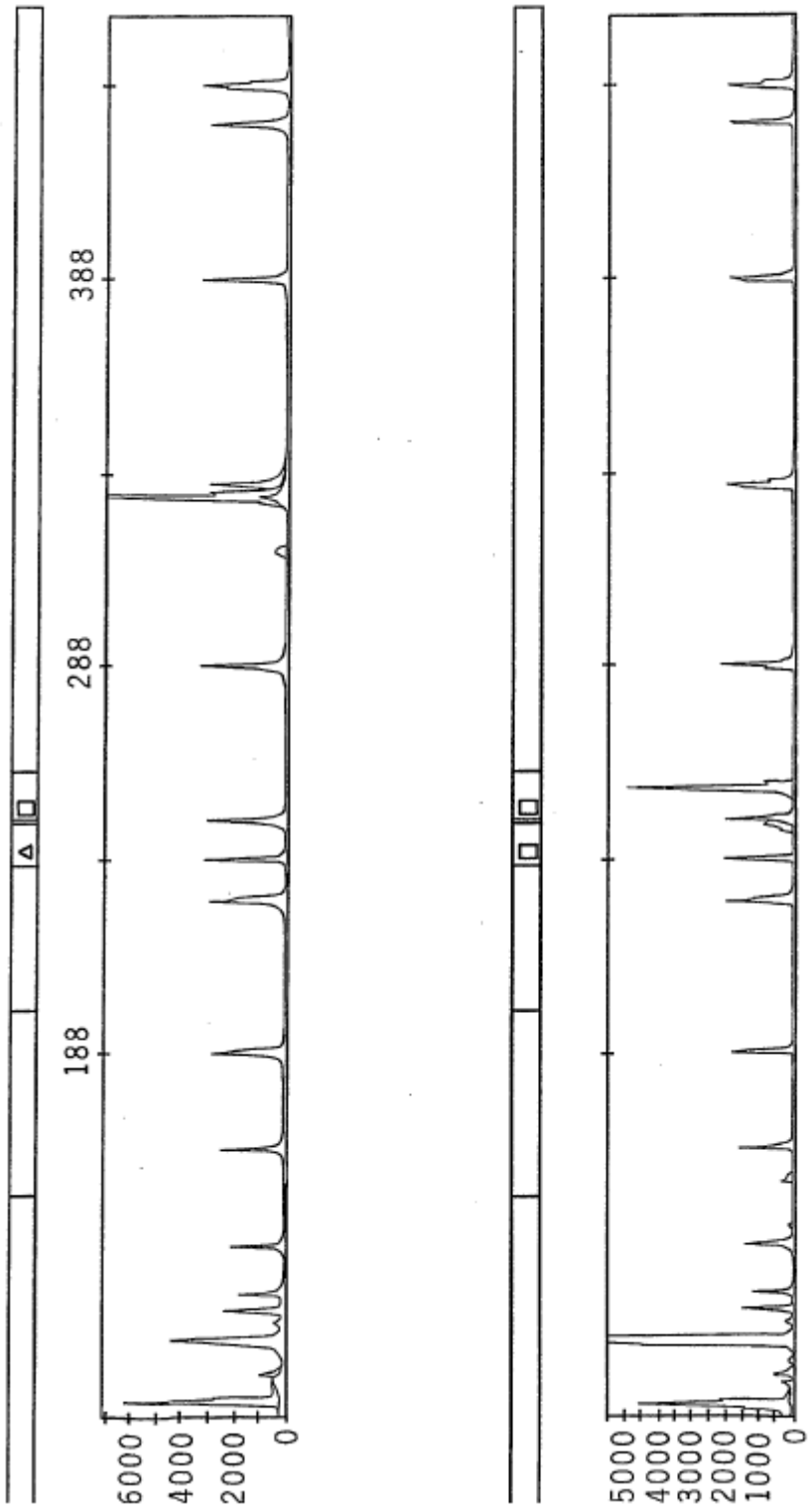


FIG. 31Ba

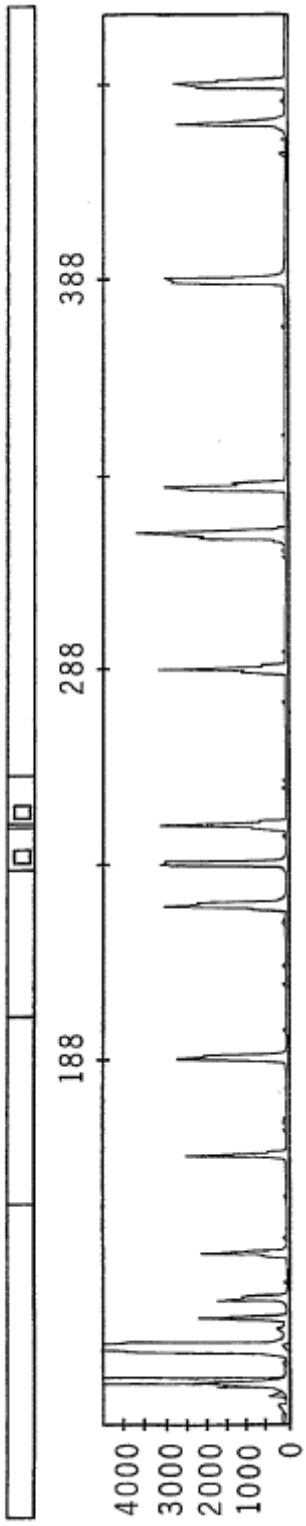
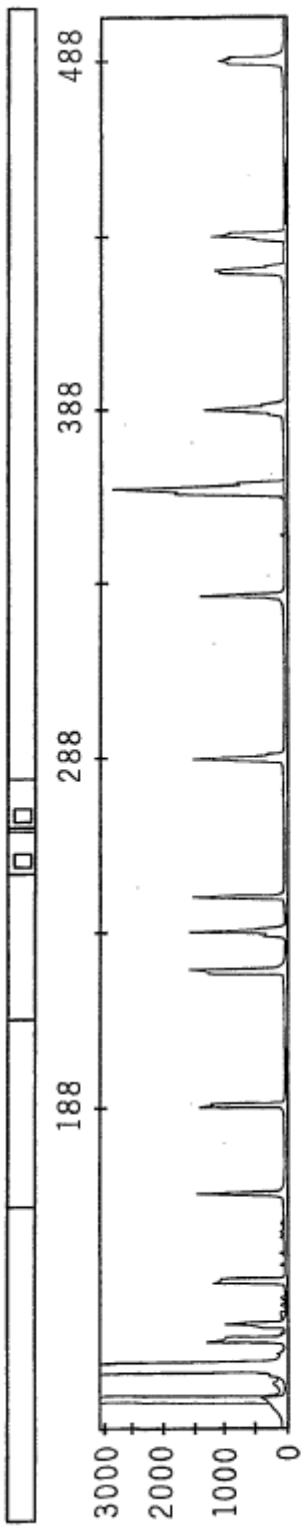


FIG. 31Bb


188
288
388
488
3000
2000
1000
0

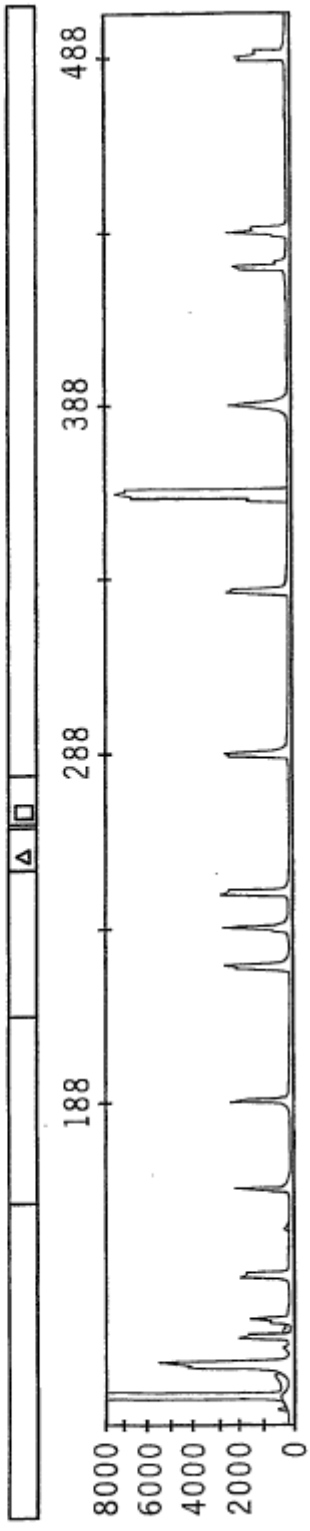

188
288
388
488
8000
6000
4000
2000
0

FIG. 31Ca

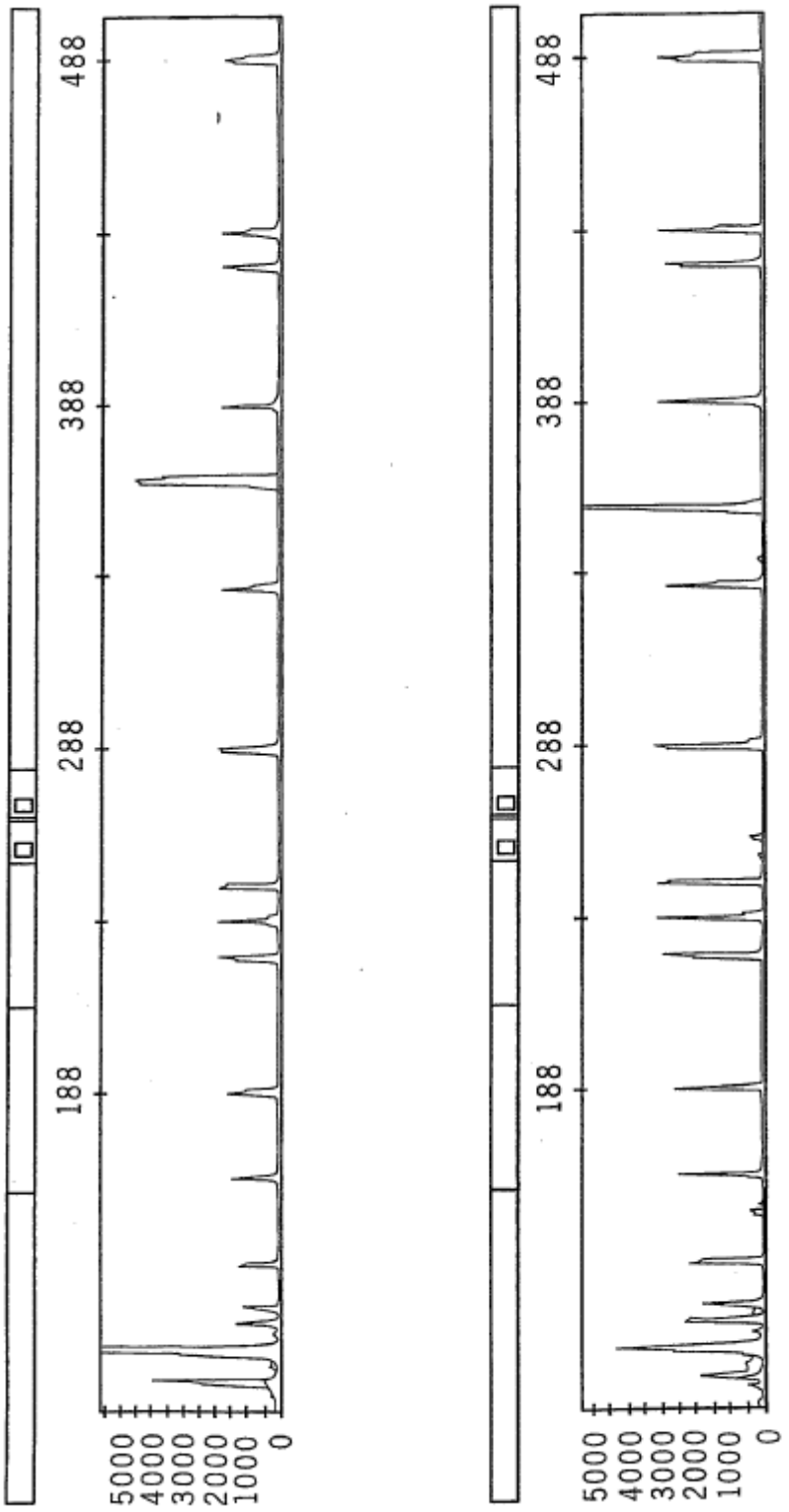

188
288
388
488
5000
4000
3000
2000
1000
0
188
288
388
488
5000
4000
3000
2000
1000
0

FIG. 31Cb

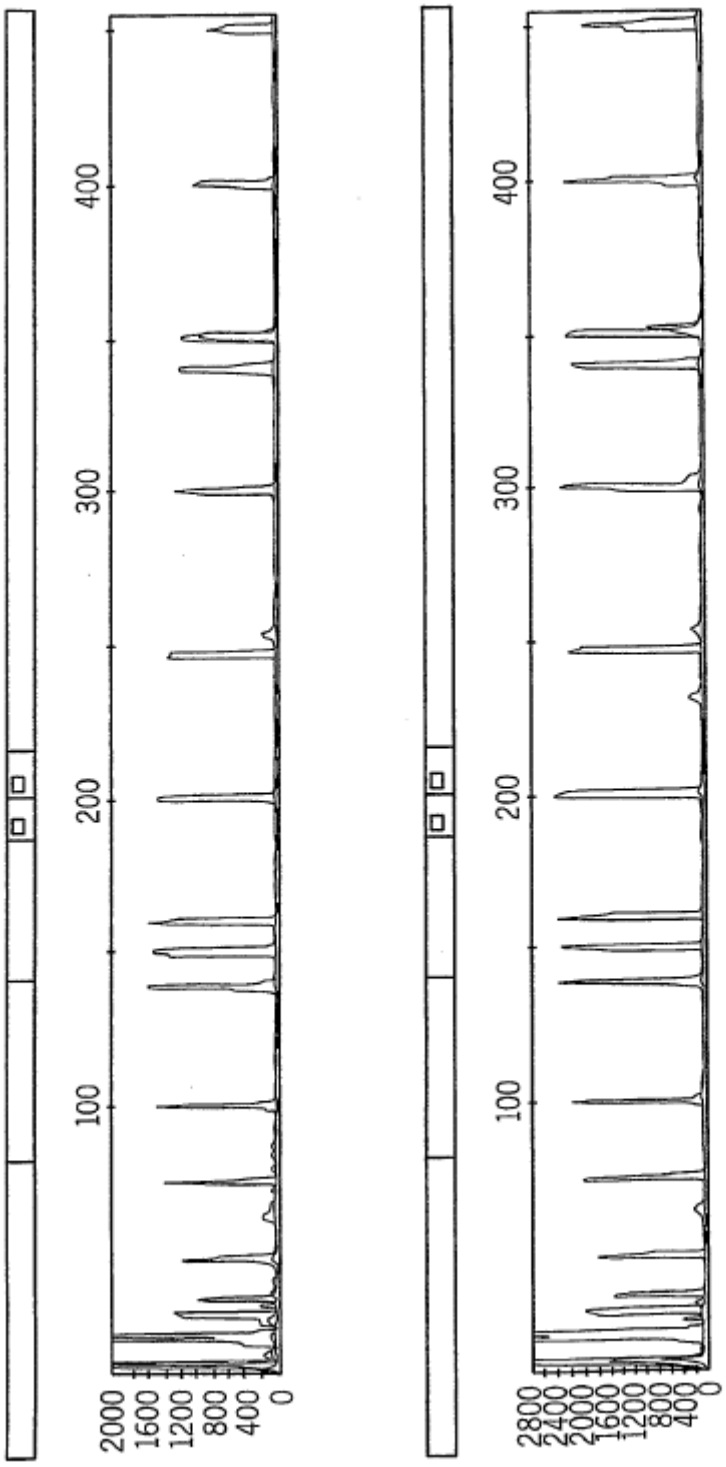

100
200
300
400
2000
1600
1200
800
400
0
100
200
300
400
2800
2400
2000
1600
1200
800
400
0

FIG. 31D

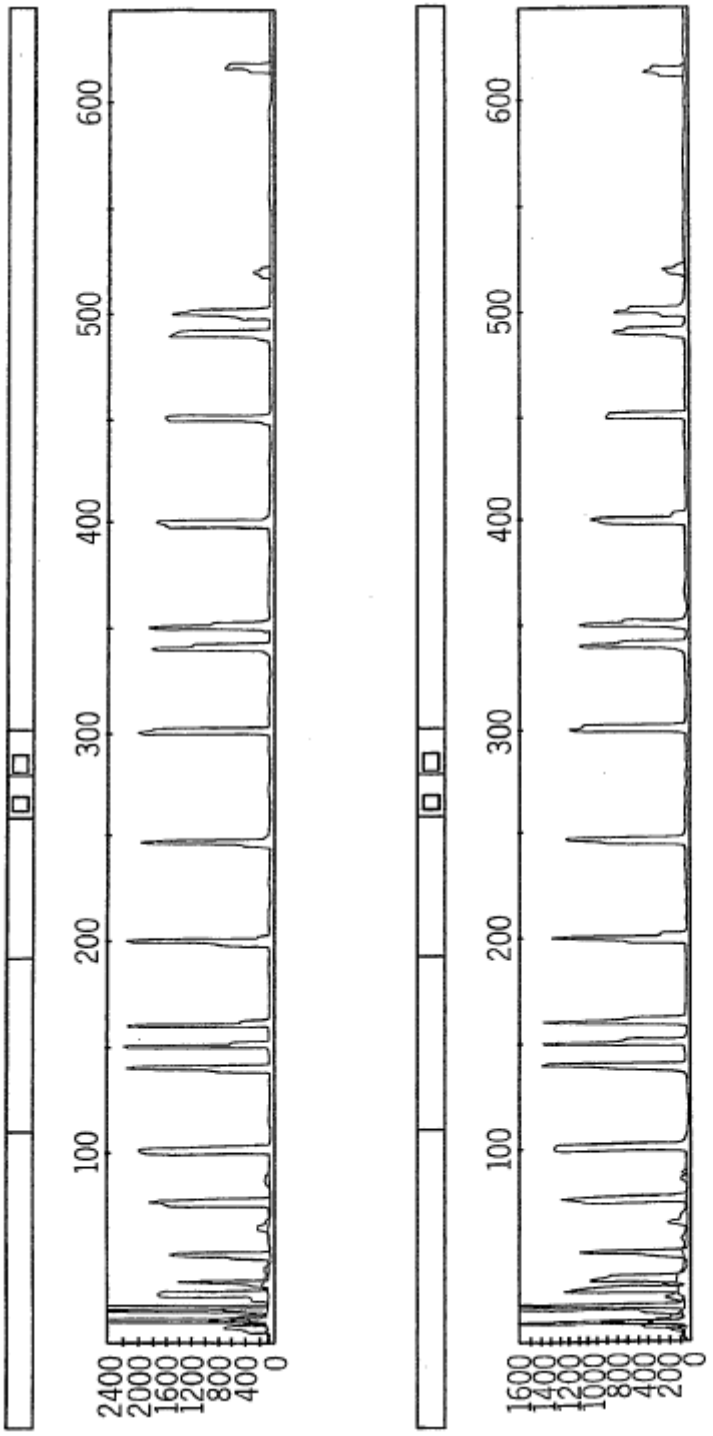

100
200
300
400
500
600
2400
2000
1600
1200
800
400
0
100
200
300
400
500
600
1600
1400
1200
1000
800
400
200
0

FIG. 31E

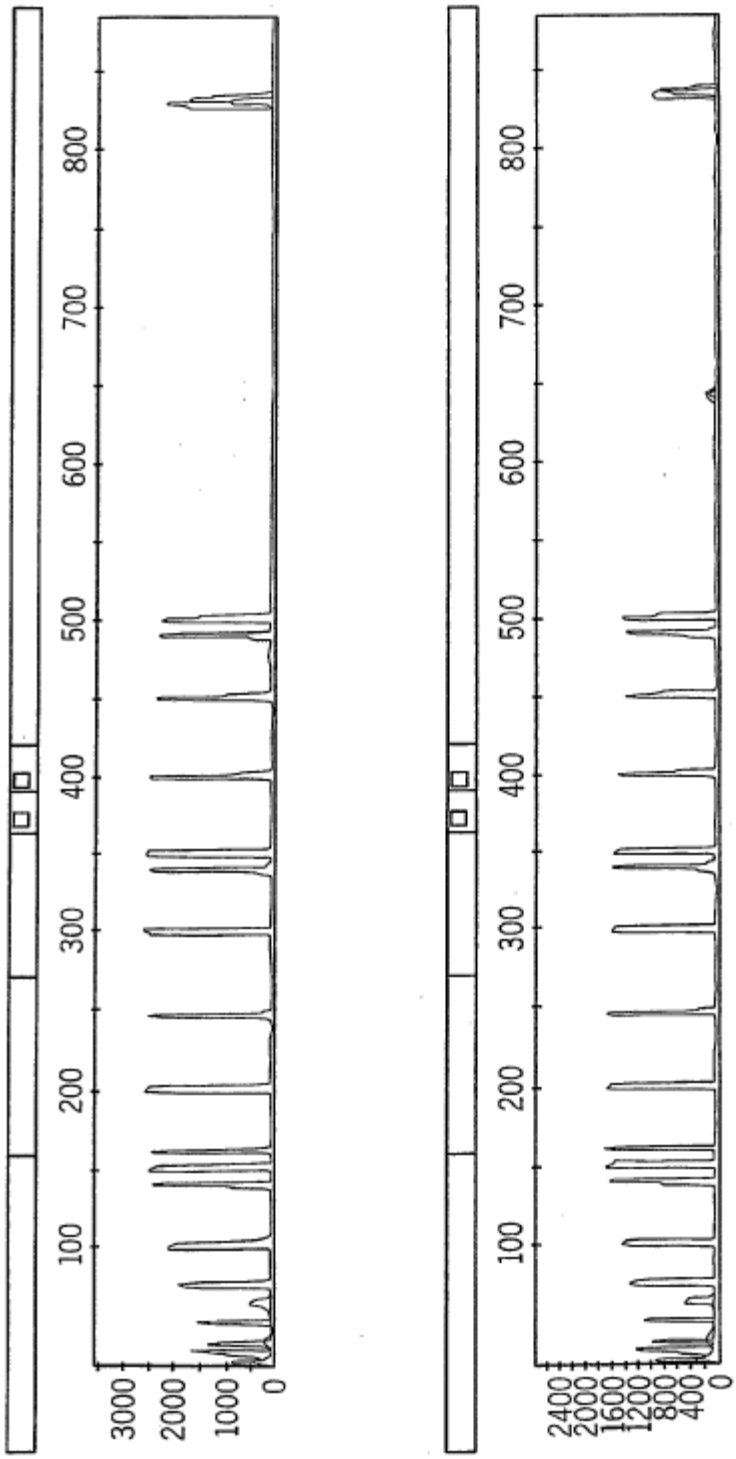

100
200
300
400
500
600
700
800
3000
2000
1000
0
100
200
300
400
500
600
700
800
2400
2000
1600
1200
800
400
0

FIG. 31Fa

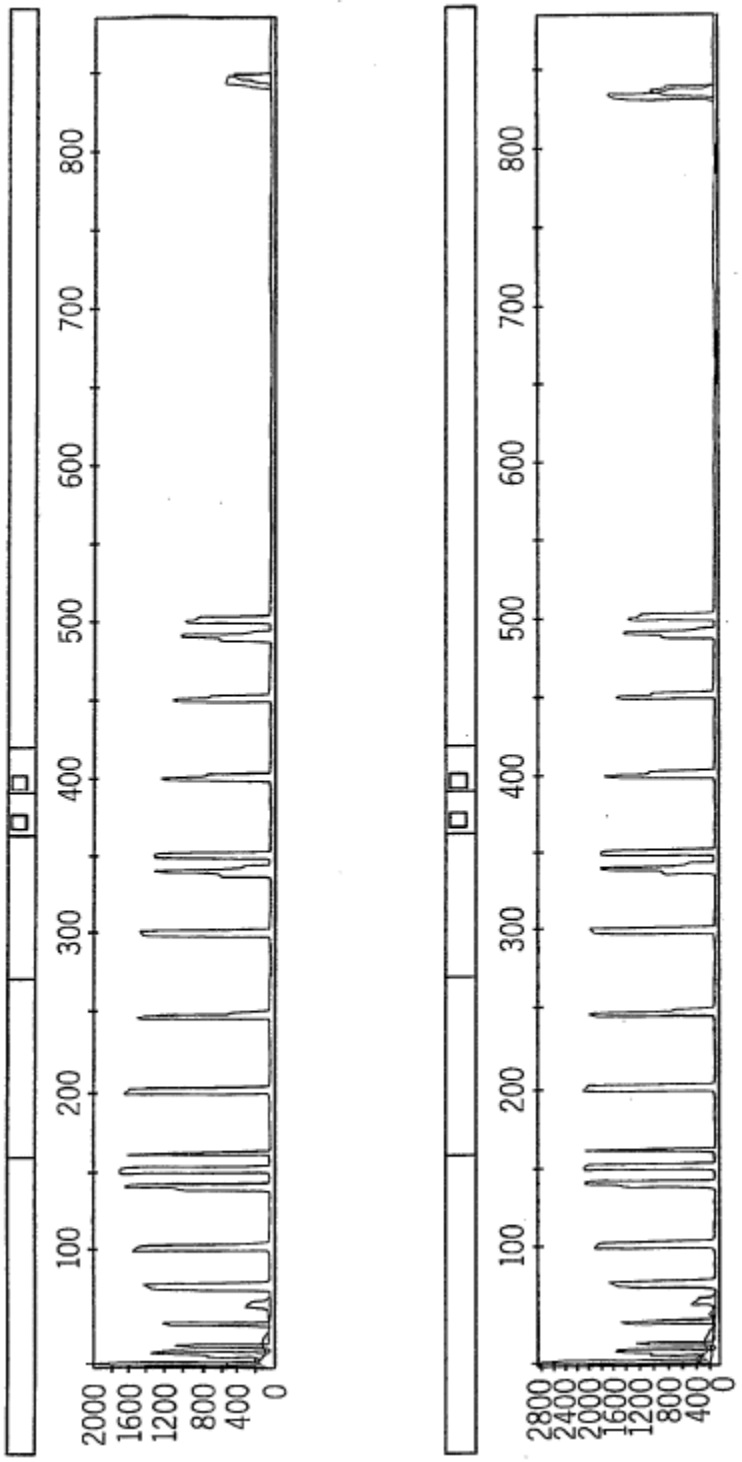

100
200
300
400
500
600
700
800
2000
1600
1200
800
400
0
100
200
300
400
500
600
700
800
2800
2400
2000
1600
1200
800
400
0

FIG. 31Fb

FIG. 32A

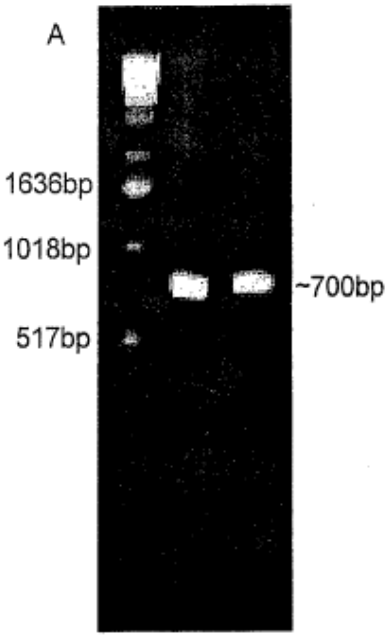


FIG. 32B

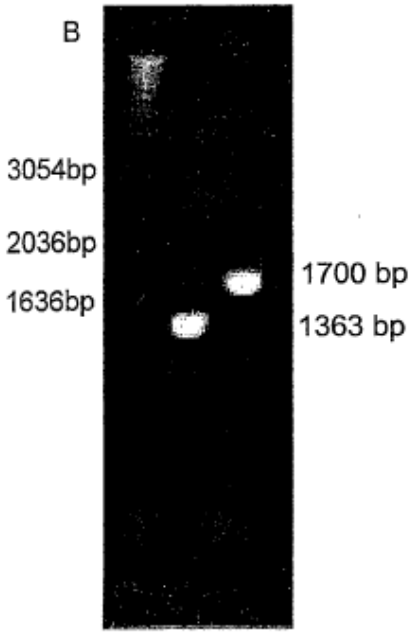
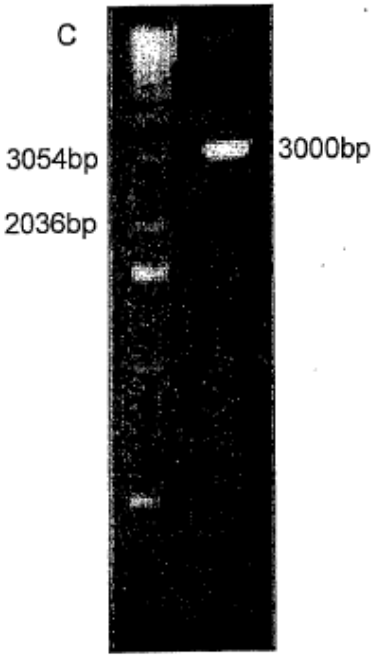


FIG. 32C



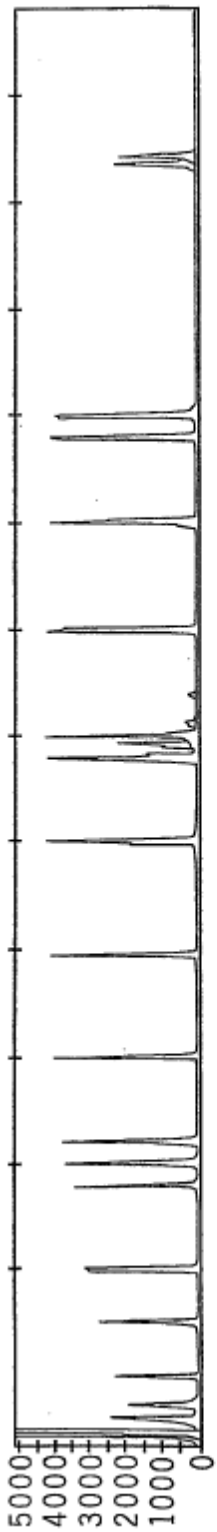


FIG. 33

FIG. 34

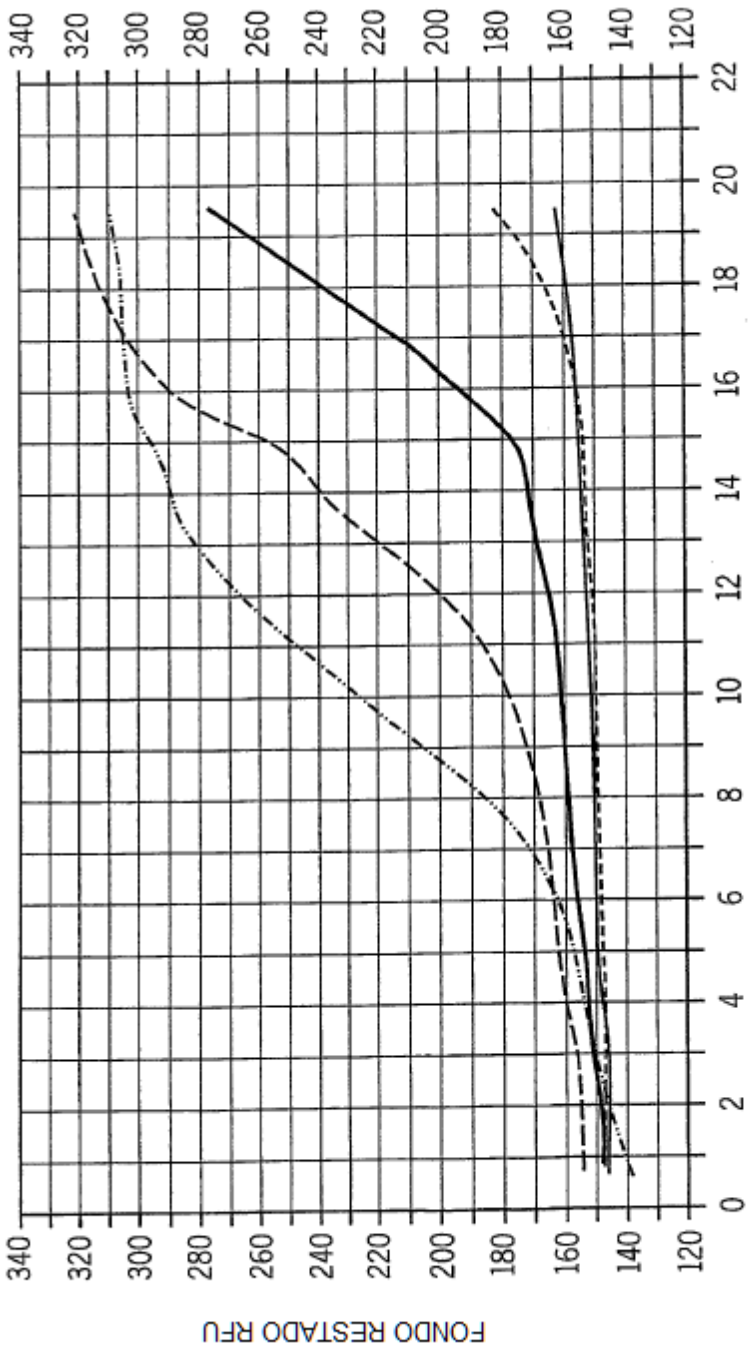

FONDO RESTADO RFU
340
320
300
280
260
240
220
200
180
160
140
120
0
2
4
6
8
10
12
14
16
18
20
22
CICLO

FIG. 35

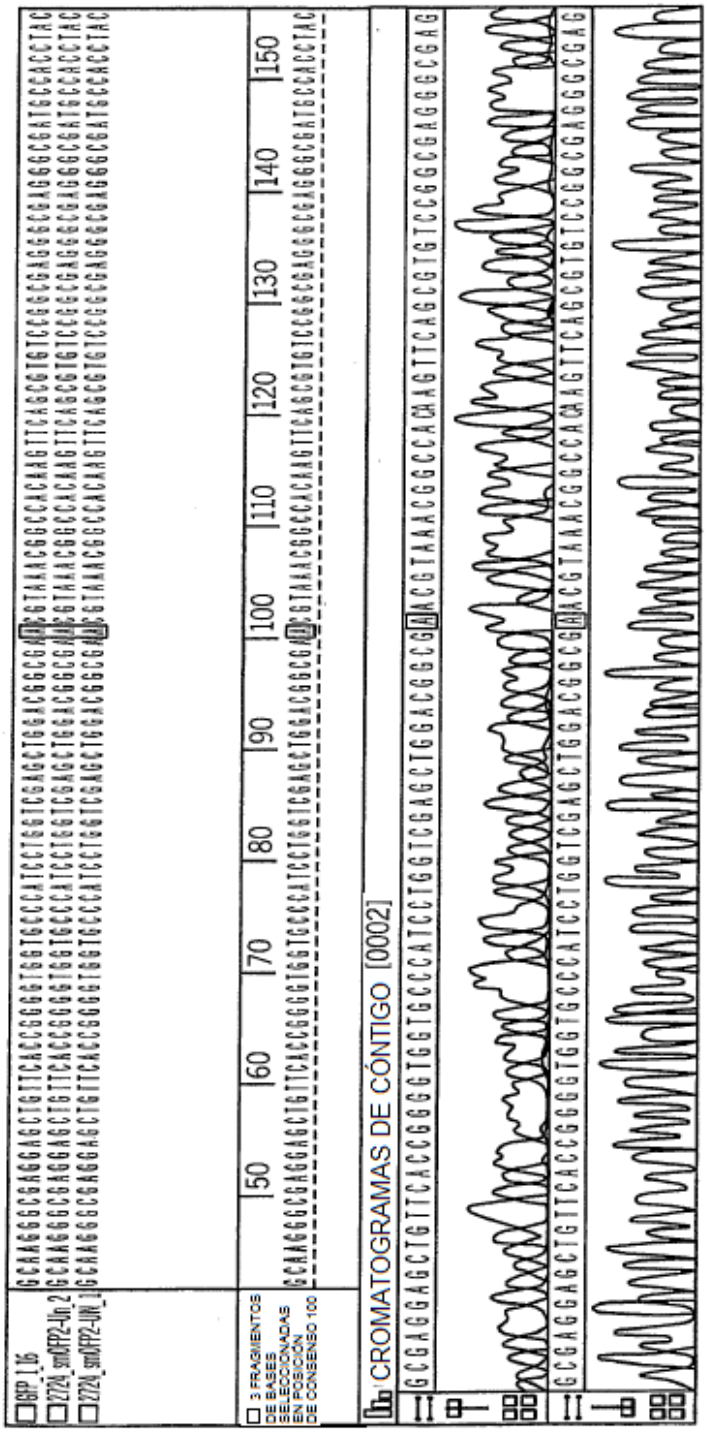

GFP_1.ab1
2724_smAFP2-Un_2
2724_smAFP2-UN_1
3 FRAGMENTOS DE BASES SELECCIONADAS EN POSICIÓN DE CONSENSO 100
50
60
70
80
90
100
110
120
130
140
150
CROMATOGRAMAS DE CÓNTIGO [0002]

FIG. 36

FIG. 37A

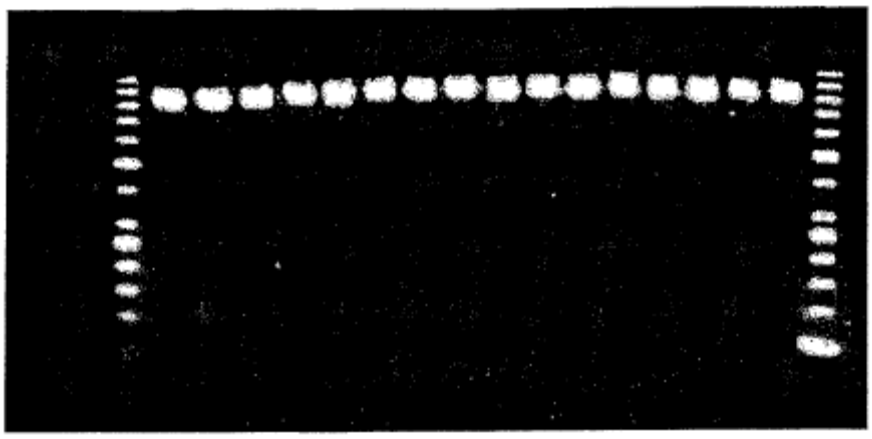


FIG. 37B

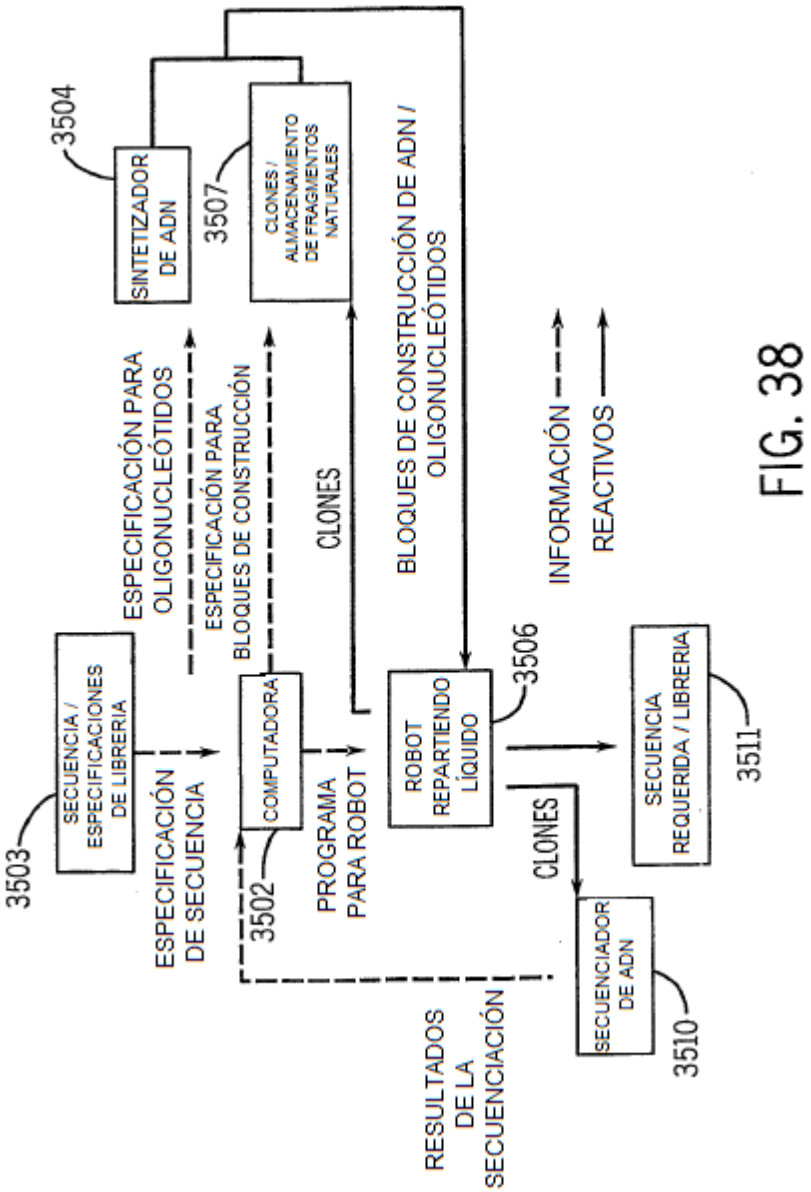

3504
SINTETIZADOR DE ADN
3507
CLONES / ALMACENAMIENTO DE FRAGMENTOS NATURALES
ESPECIFICACIÓN PARA OLIGONUCLEÓTIDOS
ESPECIFICACIÓN PARA BLOQUES DE CONSTRUCCIÓN
CLONES
BLOQUES DE CONSTRUCCIÓN DE ADN / OLIGONUCLEÓTIDOS
INFORMACIÓN
REACTIVOS
3503
SECUENCIA / ESPECIFICACIONES DE LIBRERIA
ESPECIFICACIÓN DE SECUENCIA
3502
COMPUTADORA
PROGRAMA PARA ROBOT
ROBOT REPARTIENDO LÍQUIDO
3506
SECUENCIA REQUERIDA / LIBRERIA
3511
CLONES
SECUENCIADOR DE ADN
3510
RESULTADOS DE LA SECUENCIACIÓN

FIG. 38

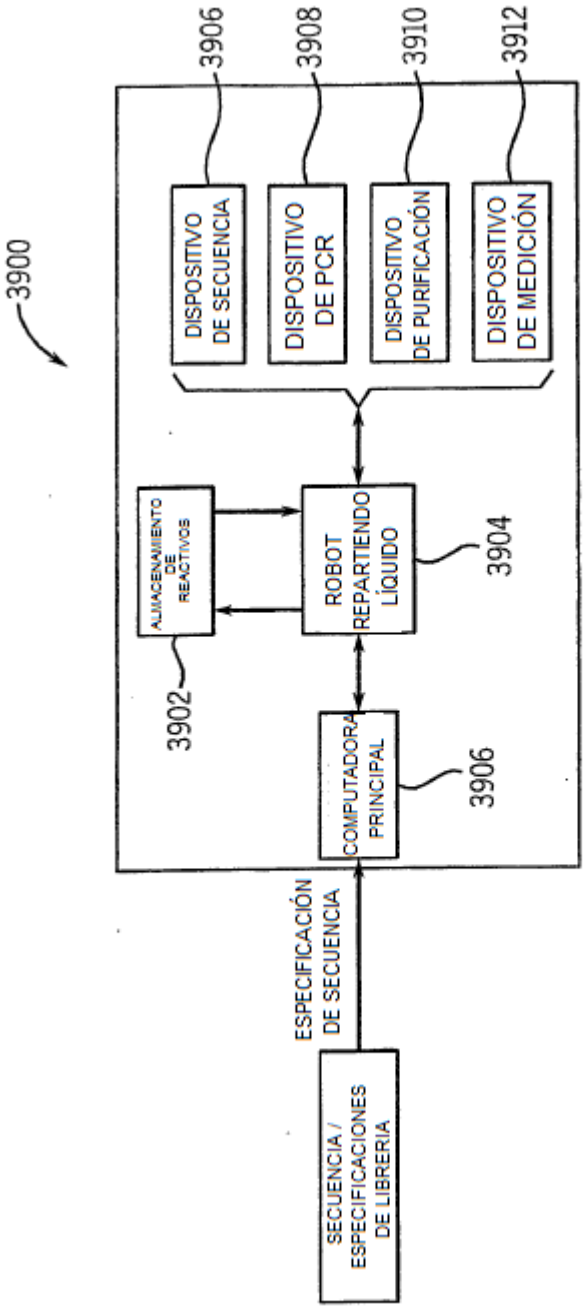

3900
3906
3908
3910
3912
DISPOSITIVO DE SECUENCIA
DISPOSITIVO DE PCR
DISPOSITIVO DE PURIFICACIÓN
DISPOSITIVO DE MEDICIÓN
ALMACENAMIENTO DE REACTIVOS
ROBOT REPARTIENDO LÍQUIDO
3904
3902
COMPUTADORA PRINCIPAL
3906
ESPECIFICACIÓN DE SECUENCIA
SECUENCIA / ESPECIFICACIONES DE LIBRERIA

FIG. 39

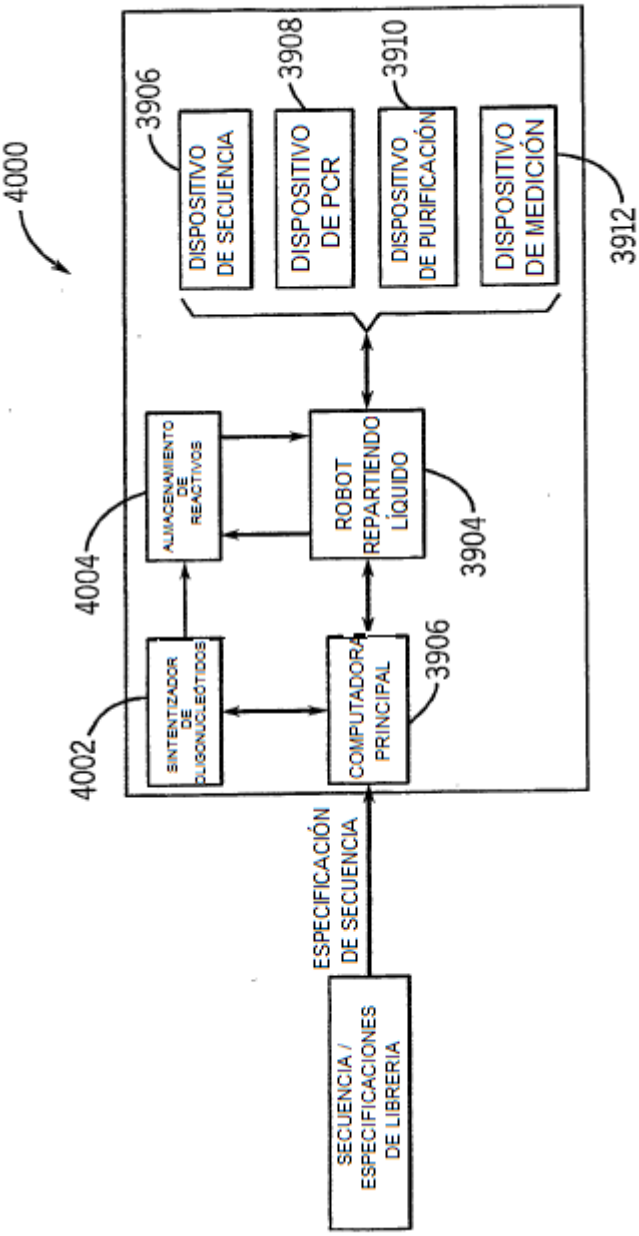

4000
3906
DISPOSITIVO DE SECUENCIA
3908
DISPOSITIVO DE PCR
3910
DISPOSITIVO DE PURIFICACIÓN
DISPOSITIVO DE MEDICIÓN
3912
ALMACENAMIENTO DE REACTIVOS
ROBOT REPARTIENDO LÍQUIDO
3904
4004
SINTENTIZADOR DE OLIGONUCLEÓTIDOS
COMPUTADORA PRINCIPAL
3906
4002
ESPECIFICACIÓN DE SECUENCIA
SECUENCIA / ESPECIFICACIONES DE LIBRERIA

FIG. 40

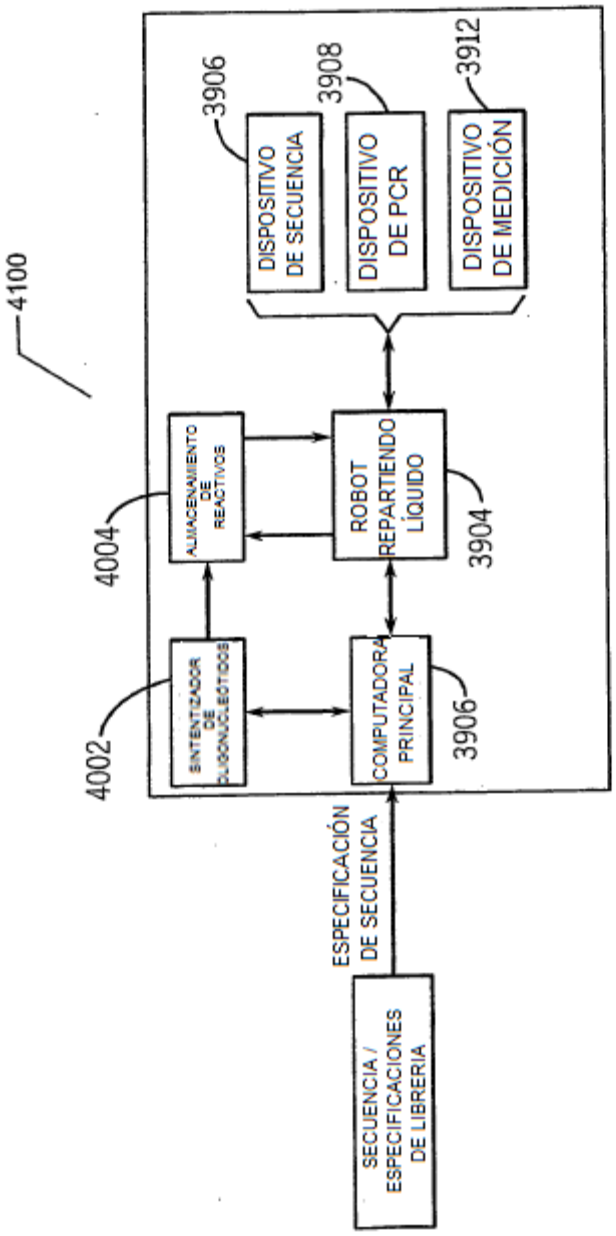

4100
3906
DISPOSITIVO DE SECUENCIA
3908
DISPOSITIVO DE PCR
3912
DISPOSITIVO DE MEDICIÓN
4004
ALMACENAMIENTO DE REACTIVOS
ROBOT REPARTIENDO LÍQUIDO
3904
4002
ENTENTIZADOR DE OLIGONUCLEÓTIDOS
COMPUTADORA PRINCIPAL
3906
ESPECIFICACIÓN DE SECUENCIA
SECUENCIA / ESPECIFICACIONES DE LIBRERIA

FIG. 41

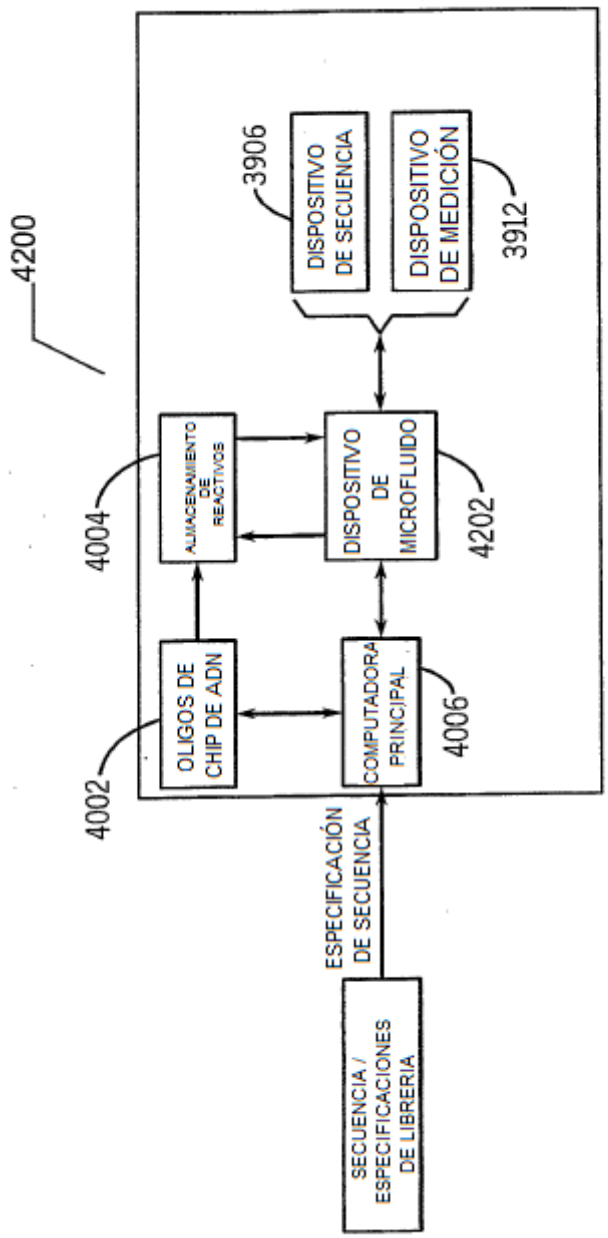

4200
3906
DISPOSITIVO DE SECUENCIA
DISPOSITIVO DE MEDICIÓN
3912
4004
ALMACENAMIENTO DE REACTIVOS
DISPOSITIVO DE MICROFLUIDO
4202
OLIGOS DE CHIP DE ADN
COMPUTADORA PRINCIPAL
4006
4002
ESPECIFICACIÓN DE SECUENCIA
SECUENCIA / ESPECIFICACIONES DE LIBRERIA

FIG. 42