



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 301 553**

(51) Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/64 (2006.01)

C12N 9/02 (2006.01)

C12Q 1/68 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **01945414 .9**

(86) Fecha de presentación : **13.06.2001**

(87) Número de publicación de la solicitud: **1299532**

(87) Fecha de publicación de la solicitud: **09.04.2003**

(54) Título: **Bancos combinatorios mejorados por recombinación en levadura y procedimiento de análisis.**

(30) Prioridad: **14.06.2000 FR 00 07555**

(45) Fecha de publicación de la mención BOPI:
01.07.2008

(45) Fecha de la publicación del folleto de la patente:
01.07.2008

(73) Titular/es: **Aventis Pharma S.A.**
20, avenue Raymond Aron
92160 Antony, FR
CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS)

(72) Inventor/es: **Truan, Gilles;**
Abecassis, Valérie y
Pompon, Denis

(74) Agente: **Elzaburu Márquez, Alberto**

ES 2 301 553 T3

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

DESCRIPCIÓN

Bancos combinatorios mejorados por recombinación en levadura y procedimiento de análisis.

5 La presente invención se refiere a un procedimiento de fabricación de bancos combinatorios de expresión funcional a partir de

La presente invención se refiere a un procedimiento de fabricación de bancos combinatorios de expresión funcional a partir de un banco combinatorio de ácidos nucleicos que pertenecen a una misma familia génica, que comprende una etapa de clonación por recombinación en levadura. La invención se refiere también a un procedimiento de producción de proteínas mosaicas funcionales, y de un análisis de un banco combinatorio de expresión funcional, por determinación de una huella secuencial para cada una de las proteínas mosaicas del banco.

15 La diversidad de las funciones de las proteínas puede verse con el resultado de la evolución de los genes por acontecimientos de mutación, recombinación y selección (1, 2). Se han desarrollado diferentes técnicas para intentar reproducir, a escala de laboratorio, las diferentes etapas de los procesos de evolución natural. Las aproximaciones clásicas de evolución molecular utilizan etapas de mutación aleatoria y recombinación por amplificación en cadena con polimerasa (denominado abreviadamente PCR por sus iniciales en inglés) (2-5). La evolución molecular es una aproximación que se ha utilizado con éxito en biotecnología para la modificación de funciones proteicas (5-12) y para permitir una mejor comprensión de los mecanismos de reconocimiento de sustratos (13). La evolución molecular constituye una aproximación eficaz para la comprensión del papel de las zonas de secuencias para la función proteica cuando dichas secuencias no están comprendidas en zonas altamente conservadas, cuando la estructura tridimensional no se conoce o cuando no está disponible ninguna información derivada de técnicas de modelización (29).

25 Un método para generar proteínas quiméricas que derivan del citocromo P₄₅₀, por recombinación *in vivo* en la levadura, ha permitido anteriormente obtener quimeras bipartitas que presentan una actividad monooxigenasa aumentada (Bellamine *et al.*, 1994). En otro estudio que pretende hacer evolucionar una secuencia nucleótida, en función de una característica determinada, también se ha utilizado la recombinación *in vivo* (EP 0934999). Esta vez, el ADN se fragmenta, previamente a la etapa de recombinación, para permitir una reordenamiento entre secuencias homólogas.

30 Con el fin de proceder a experimentos de evolución molecular o mezcla recombinatoria de ADN (DNA-shuffling en inglés), se parte de un banco de genes que puede generarse por mutagénesis de una secuencia única (14) o que puede estar constituida por un grupo que pertenece a una misma familia o subfamilia de genes (15). La técnica dicha de mezcla recombinatoria de familia (family-shuffling en inglés) se ha descrito como un medio de acelerar los procesos de evolución (16), que permite la emergencia de actividades o de propiedades inesperadas en las nuevas proteínas generadas (14). Esta técnica ha permitido así la creación de enzimas que presentan una asociación de propiedades parentales de interés (17, 18), que tienen una estabilidad térmica acrecentada (14) o que presentan nuevas especificidades de sustrato (19).

40 Sin embargo, incluso si la mezcla recombinatoria de genes de una misma familia (family-shuffling) permite obtener mejoras que imitan *in vitro* los procesos de evolución, la construcción de bancos aleatorios de estructuras mosaicas desprovistas de errores respecto del ensamblaje de una mayoría de estructuras parentales siempre es un punto crítico.

45 Las dificultades para obtener banco homogéneo por "family-shuffling" aumentan fuertemente cuando las similitudes entre las secuencias de partida utilizadas disminuyen (30, 31). Así, un número relativamente pequeño (del orden de 10%) de quimeras ha sido frecuentemente descrito (Kikuchi describe 1% de estructuras quimeras para 2 genes que tienen 84% de identidad a nivel proteico al utilizar las técnicas clásicas de DNA-shuffling (32)).

50 Se han desarrollado diferentes técnicas para disminuir la tasa de estructuras parentales cuya utilización de ADN de hélice simple como punto de partida para la mezcla recombinatoria (dando 14% de estructuras quiméricas para 2 genes que tienen 84% de identidad a nivel proteico (33) o fragmentaciones enzimáticas limitadas (32,34) dando, ellas, tasas de quimeras mucho más importantes. Sin embargo, este último método presenta el inconveniente que los fragmentos generados de manera enzimática no son fragmentos aleatorios, lo que induce una limitación en el número de nuevas estructuras génicas que pueden así introducirse.

55 Otros grupos han utilizado la recombinación *in vivo* en sistemas procariotas para obtener quimeras (30, 35, 36). Estos métodos presentan sin embargo el inconveniente de que la expresión funcional de proteínas en *E. coli* no es siempre la más adaptada cuando se trata de proteínas eucariotas, en particular los complejos multiproteicos, las proteínas membranas, o cualquier proteína que necesita la maquinaria celular eucariota para su actividad. En particular, ciertas proteínas eucariotas presentan modificaciones post-traduccionales (glicosilación...) que no pueden efectuarse en huéspedes procariotas.

65 Por lo tanto es un objetivo de la presente invención proporcionar un procedimiento de construcción de bancos combinatorios de expresión funcional a partir de ácidos nucleicos que pertenecen a una misma familia génica, que permite obtener bancos que presentan la complejidad necesaria, es decir que presentan una gran parte de estructuras quiméricas posibles, y con una tasa de presencia de estructuras parentales relativamente baja. Además, el procedimiento de la presente invención permite obtener bancos que permiten una mejor expresión de proteínas eucariotas.

La presente invención divulga igualmente un procedimiento de análisis de secuencias génicas de un banco combinatorio, en particular obtenido por el procedimiento según la invención, que permite asociar una “huella” a cada variante de secuencia presente en dicho banco. Este procedimiento de análisis permite, en combinación con un procedimiento de análisis de las funciones y/o actividades de las proteínas de dicho banco, de poder unir dichas estructuras
5 secuenciales y dichas estructuras funcionales. Así, la combinación de estos dos procedimientos puede utilizarse para “pilotar” la mezcla de informaciones genéticas, con el fin de obtener proteínas de interés de manera dirigida, más controlada, más rápidamente y con un menor coste.

Así, la presente invención se refiere a un procedimiento de construcción de un combinatorio de expresión funcional
10 a partir de un banco de ácidos nucleicos que pertenecen a una misma familia génica, caracterizada porque comprende las etapas que consisten en:

- a. introducir dicho banco de ácidos nucleicos en la levadura, simultáneamente con un vector de expresión,
- 15 b. obtener dicho banco de expresión funcional por recombinación de dicho banco combinatorio de ácidos nucleicos con dicho vector de expresión en dicha levadura.

Un banco combinatorio de expresión funcional obtenido mediante tal procedimiento según la invención es igualmente un objetivo de la invención.

Preferentemente, el vector de expresión con el que se efectúa la recombinación en la levadura se linealiza en el sitio de la clonación normal del ADNc y posee secuencias promotoras y terminadoras de transcripción, efectuándose la recombinación a nivel de dichas secuencias.

Los fragmentos de ácidos nucleicos que pertenecen al banco introducido en la levadura en la etapa a. pueden fragmentarse o no. Cuando estos fragmentos se fragmentan, esto permite aumentar la eficacia de recombinación *in vivo*, lo que aumenta la diversidad del banco, en la medida en la que un acontecimiento de recombinación es obligatorio antes de la clonación en el vector de expresión. Estos puntos se discutirán más adelante.

Los acontecimientos de recombinación que se efectúan en la levadura pueden ser de recombinación homóloga (entre secuencias idénticas) u homeólogas (entre secuencias que presentan un nivel de identidad suficiente).

El procedimiento según la invención igualmente es muy interesante porque no necesita de etapa de paso a un procariota, para la obtención del banco combinatorio.

Así, el procedimiento según la presente invención permite la obtención de un banco combinatorio de expresión directamente en un huésped eucariota, lo que presenta una ventaja cierta para la expresión de proteínas eucariotas, en particular las proteínas membranas, o que pertenecen a complejos multiproteicos.

El procedimiento según la presente invención se refiere por lo tanto a un método de producción de bancos combinatorios mejorados por recombinación en una levadura (CLERY, por Combinational Library Enhanced by Recombination in Yeast).

La levadura (que puede modificarse a nivel genómico) se utiliza igualmente ventajosamente como instrumento de expresión (39) de genes quiméricos, lo que permite mejorar la expresión funcional de nuevas proteínas eucariotas obtenidas por este método (en particular los complejos multiproteicos o las proteínas membranas). Además, la modificación genómica de la cepa de levadura utilizada puede permitir recrear el entorno natural de funcionamiento (y por lo tanto la optimización de las posibilidades de cribado), por producción de otras proteínas eucariotas esenciales para la actividad de las nuevas proteínas creadas, en particular en el caso de complejos multiproteicos.

El procedimiento según la invención permite la obtención de un banco combinatorio de expresión funcional por el hecho de dos etapas diferentes:

- la clonación del banco de ácidos nucleicos en el vector de expresión introducido simultáneamente en la levadura, por recombinación homóloga *in vivo*, permite obtener un banco de expresión funcional;
- la recombinación homóloga u homeóloga (entre las secuencias similares pero no idénticas), que se pueden producir *in vivo* en levadura, entre los diferentes ácidos nucleicos del banco combinatorio introducido en dicha levadura, permite aumentar la complejidad y la diversidad del banco combinatorio de expresión funcional obtenido.

Así, cuando los fragmentos de ácidos nucleicos del banco combinatorio introducido en dicha levadura se fragmentan y no poseen los dos extremos recombinógenos que permiten la clonación en el vector de expresión, es esencial que tenga lugar un acontecimiento de recombinación entre dos fragmentos adecuados previamente a dicha clonación.

Igualmente en un caso particular de realización del procedimiento según la invención, se observa la realización de al menos un acontecimiento de recombinación homeóloga en el banco obtenido, principalmente por el hecho de pertenecer a una misma familia génica de los ácidos nucleicos del banco introducido inicialmente en la levadura.

Por “ácidos nucleicos que pertenecen a una misma familia génica”, se entiende, en el sentido de la invención, ácidos nucleicos que poseen un mínimo de 35% de identidad, de manera preferida 40%, de manera más preferida 50%, o también 70%. Estos ácidos nucleicos se dirá que pertenecen a la misma familia génica si presentan los porcentajes de identidades anteriores, y pueden codificar proteínas que presentan actividades y/o funciones diferentes. Estos aminoácidos pueden codificar proteínas encontradas naturalmente, o pueden ser ácidos nucleicos “artificiales”, es decir que codifican proteínas que no se encuentran en la naturaleza. En particular tales ácidos nucleicos “artificiales” engloban proteínas de fusión, o proteínas ya obtenidas por métodos de mezcla recombinatoria de ADN.

Por “porcentaje de identidad” entre dos secuencias de ácidos nucleicos o de aminoácidos en el sentido de la presente invención, se entiende designar un porcentaje de nucleótidos o de residuos de aminoácidos idénticos entre las dos secuencias a comparar, obtenido después del mejor alineamiento, siendo este porcentaje puramente estadístico y estando repartidas las diferencias entre las dos secuencias al azar y sobre toda su longitud. Se entiende designar por “mejor alineamiento” o “alineamiento óptimo”, el alineamiento para el que el porcentaje de identidad determinado como se designa a continuación es el más elevado. Las comparaciones de secuencias entre dos secuencias de ácidos nucleicos o de aminoácidos se realizan tradicionalmente al comparar estas secuencias después de haberlas alineado de manera óptima, realizándose dicha comparación por segmento o por “ventana de comparación” para identificar y comparar las regiones locales de similitud de secuencia. El alineamiento óptimo de las secuencias para la comparación pueden realizarse, además de manualmente, por medio del algoritmo de homología local de Smith y Waterman (49), por medio del algoritmo de homología local de Neddleman y Wunsch (50), por medio del método de búsqueda de similitud de Pearson y Lipman (51), por medio de programas informáticos que utilizan estos algoritmos (GAP, BESTFIT, BLAST P, BLAST N, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI). Con el fin de obtener el alineamiento óptimo, se utiliza preferentemente el programa BLAST, con la matriz BLOSUM 62. Se pueden utilizar igualmente las matrices PAM o PAM250.

La presente invención permite obtener así, con un rendimiento alto, bancos recombinatorios a partir de ácidos nucleicos que presentan una identidad muy inferior a la identidad actualmente requerida en el estado de la técnica (generalmente superior a 70%).

El banco de ácidos nucleicos introducido en la levadura en la etapa a. del procedimiento según la invención es, preferentemente, el mismo un banco combinatorio de ácidos nucleicos.

Este banco de ácidos nucleicos es, preferentemente, una mezcla de productos de PCR obtenida por amplificación de un banco combinatorio de fases abiertas de lectura, utilizando una pareja de iniciadores situados en las regiones que flanquean dichas fases abiertas de lectura. Este banco combinatorio de fases abiertas de lectura se obtiene a partir de ADN variantes de secuencias, difiriendo en una o varias mutaciones, y que pertenece a una misma familia génica en el sentido de la invención.

Se utiliza preferentemente una sola pareja de iniciadores para efectuar la reacción de PCR tal como se describe en el párrafo anterior, pero el experto en la técnica podría utilizar igualmente parejas de iniciadores diferentes. Sin embargo es más práctico utilizar una única pareja de iniciadores.

En particular, se utiliza una pareja de iniciadores situada en las regiones promotoras y terminadoras de la traducción en la levadura, regiones que permiten la expresión de fases abiertas de lecturas en este organismo. Así, es probable que estas regiones, que estarán presentes en todos los fragmentos de ADN del banco de ácidos nucleicos introducido en la levadura, serán las secuencias nucleicas implicadas en la recombinación con las secuencias homólogas del vector de expresión co-introducido, que permitirán la clonación de fases abiertas de lectura en dicho vector, y la formación del banco de expresión funcional.

Así como se ha precisado anteriormente, en banco de ácidos nucleicos introducido en la levadura es, preferentemente, el mismo un banco combinatorio de ácidos nucleicos, que pertenece a una misma familia génica en el sentido de la invención. Se puede obtener este banco combinatorio mediante métodos clásicos de fragmentación de ADN y de reensamblaje por extensión con iniciador (“extensión de primer”).

La etapa de fragmentación del ADN se efectúa mediante métodos conocidos por el experto en la técnica, como por ejemplo la digestión por enzimas de restricción, o la nebulización. Sin embargo se prefiere fragmentar el ADN por digestión parcial con una ADNasa, preferentemente la ADNasaI, que permite obtener de manera más controlada fragmentos de un tamaño deseado. Además, esto permite obtener efectivamente fragmentos aleatorios, lo que no siempre es el caso con las otras técnicas de fragmentación enzimática. En la práctica, y con el fin de obtener un banco combinatorio que presenta una gran variedad de combinación y un gran número de proteínas mosaicas diferentes, se busca obtener fragmentos de tamaño comprendido entre 15 y 700 pares de bases (pb), preferentemente de 40 a 500 pb, de manera más preferida de 100 a 300 pb.

Los fragmentos se reensamblan entre ellos mediante una técnica de extensión por iniciadores (“extensión de primer”). En el principio, los fragmentos obtenidos pueden hibridarse, y la adición de una ADN polimerasa permite obtener una extensión de fragmentos híbridos, y la reconstitución de genes funcionales, mediante varios ciclos de extensión.

ES 2 301 553 T3

Así, la presente invención tiene igualmente por objetivo un procedimiento de construcción de un banco combinatorio de expresión funcional a partir de un banco combinatorio de ácidos nucleicos que pertenecen a una misma familia génica, que comprende las etapas que consisten en:

- 5 a. introducir dicho banco combinatorio de ácidos nucleicos en una levadura, simultáneamente con un vector de expresión,
- b. obtener dicho banco de expresión funcional por recombinación de dicho banco combinatorio de ácidos nucleicos con dicho vector de expresión en dicha levadura.

10 siendo dicho banco combinatorio de ácidos nucleicos una mezcla de productos de PCR obtenida por amplificación de un banco combinatorio de fases abiertas de lectura, utilizando una pareja de iniciadores situados en las regiones que flanquean dichas fases abiertas de lectura, obteniéndose dicho banco combinatorio a partir de ADN homólogos o
15 variantes de secuencia difiriendo en una o varias mutaciones, y dicho banco combinatorio de fases abiertas de lectura se obtiene por reensamblaje por “extensión de primer” de productos de fragmentación de al menos dos fases abiertas de lecturas que codifican proteínas funcionales, presentando dichas fases abiertas de lecturas entre ellas una identidad de secuencia superior a 40%.

El experto en la técnica conoce otras técnicas que permiten la recombinación entre fragmentos de ADN y su mezcla (DNA shuffling). Así, un método alternativo es el método de oligo-ligadura, que puede realizarse opcionalmente con
20 ligasas termoestables. Pueden elegirse otros métodos adecuados por el experto en la técnica para la mezcla de ácidos nucleicos.

Con el fin de realizar el ensamblaje de fragmentos, se utiliza preferentemente una reacción de amplificación por polimerasa (PCR). Las diferentes etapas de esta reacción deben controlarse, con el fin de poder obtener una tasa
25 importante de genes mosaicos. Así, la etapa de hibridación es una etapa importante para asegurar la posibilidad de obtener una recombinación entre fragmentos que presentan una identidad de secuencia relativamente baja, en particular para los valores bajos de genes que pertenecen a una misma familia génica (35%, o 40%). Así, la reacción de PCR realizada de manera preferida durante la etapa de reensamblaje se caracteriza porque cada uno de sus ciclos
30 presenta al menos dos escalones de hibridación, preferentemente al menos dos escalones, por temperaturas decrecientes regularmente espaciadas. Es igualmente importante que el conjunto de las etapas de hibridación tenga una duración total de más de cuatro minutos. Un modo particular de realizar la reacción de PCR es tal que cada ciclo presente al menos cuatro escalones de hibridación de más de 60 segundos, por temperaturas decrecientes regularmente
35 espaciadas.

Los inventores han mostrado en efecto que estas condiciones de reensamblaje permite la obtención de fragmentos de un tamaño superior a los ácidos nucleicos de partida. En particular, cuando los ácidos nucleicos de partida son
40 vectores de expresión que llevan los genes de la misma familia génica, las etapas de fragmentación y de reensamblaje pueden permitir obtener fragmentos de ADN transformantes en la levadura, es decir que llevan a la vez genes mosaicos, y los elementos del vector que permiten su replicación y su mantenimiento en la levadura. Esto asegura que el método de reensamblaje según el presente procedimiento es extremadamente eficaz (véanse también los ejemplos).

Con el fin de obtener un banco de expresión funcional en la levadura, el procedimiento según la presente invención propone la co-introducción de un vector de expresión y de un banco de ácidos nucleicos que pertenecen a una misma
45 familia génica, obtenida por “family shuffling” así como se ha descrito en los párrafos anteriores.

Con el fin de obtener dicho banco de ácidos nucleicos, es interesante partir de ácidos nucleicos que pertenecen a una misma familia génica ya clonada en un vector de expresión preferentemente, estos ácidos nucleicos se clonan
50 todos en el mismo vector de expresión, y se utiliza dicho vector para la co-introducción en la levadura.

Así, después de la etapa de reensamblaje descrita anteriormente, y en la medida en que las condiciones utilizadas permiten obtener largos fragmentos, en particular de tamaño igual o superior al tamaño del vector de partida (es decir
55 más largo que los ácidos nucleicos que pertenecen a la misma familia génica que se busca mezclar), se efectúa una reacción de PCR utilizando una pareja de iniciadores en las regiones flanqueadas de las fases abiertas de lectura. Se trata preferentemente de iniciadores situados en el vector de expresión, y se elige en particular en las regiones promotora y terminadora de transcripción de dicho vector, así como se ha precisado anteriormente.

Como ADN de partida, se puede así utilizar todo vector que contiene los ácidos nucleicos que pertenecen a la misma familia génica que se desea recombinar. Se puede elegir un vector multicopia en la levadura, o un vector mono-
60 copia en la levadura, o un vector cuyo carácter multi- o mono- copia sea inducible. También se puede elegir un vector de expresión para una levadura o un vector de expresión para una célula eucariota, que sea lanzadera de la levadura. Se puede elegir igualmente un vector que contiene los elementos necesarios para la réplica de manera autónoma en *Escherichia coli*. Se puede coger por supuesto un vector que no posee ninguna de las propiedades desarrolladas anteriormente, o que presenta una combinación de dichas propiedades.

Preferentemente, se realiza el procedimiento según la invención eligiendo, como vector de partida, el vector de
65 expresión co-introducido en la levadura con el banco de ácidos nucleicos.

Este vector de expresión posee los elementos para replicarse de manera autónoma en la levadura, como vector multicopia, o vector monocopia, o vector condicional. Puede poseer igualmente genes que permiten su selección en medios adecuados, en particular genes de resistencia a antibióticos o de complementación de auxotrofia si la levadura utilizada presenta esta propiedad.

El vector de expresión puede ser un vector de expresión para la levadura. En este caso, posee elementos que permiten la transcripción y la traducción de manera eficaz en la levadura. Puede ser alternativamente un vector de expresión en otro huésped, procariota o eucariota, es decir poseer los elementos (orígenes de replicación) que le permiten replicarse de manera autónoma en este otro huésped. Se elige preferentemente un vector que permite la expresión en un huésped eucariota superior, en particular una célula de mamífero. Tal vector asocia, a una casete de expresión de eucariota superior, un origen de replicación y un marcador de selección para la levadura.

El vector comprende preferentemente un promotor, señales de iniciación y de terminación de la traducción, así como regiones apropiadas de regulación de la transcripción. Puede poseer igualmente señales particulares que especifiquen la secreción de la proteína traducida. Los vectores que pueden utilizarse son bien conocidos por el experto en la técnica.

Se utiliza preferentemente, como vector que lleva los ácidos nucleicos que pertenecen a una misma familia génica que se desea fragmentar, un vector que presenta un tamaño, que incluye las fases abiertas de lectura, superior a 7 kilobases (kb). Se puede utilizar el mismo vector para la co-introducción en la levadura, para la etapa de recombinación en la levadura.

La recombinación se efectúa en la levadura, preferentemente una levadura del género *Saccharomyces*, de manera más preferida *S. cerevisiae*. Sin embargo se pueden utilizar otros tipos de levaduras, entre las cuales *Candida*, *Yarrowia*, *Kluyveromyces*, *Schizosaccharomyces*, *Torulopsis*, *Pichia*, *Hansenula*. El experto en la técnica elegirá la levadura apropiada en función de sus competencias y sus conocimientos y el objetivo buscado. Esta levadura puede modificarse a nivel genómico para expresar proteínas exógenas, que permiten complementar las proteínas mosaicas que se busca generar.

El procedimiento según la presente invención posee varias ventajas, que serán en particular visibles a la luz de los ejemplos. Sin embargo, se pueden resumir algunas:

- el procedimiento no necesita paso a un huésped procariota para la obtención del banco, lo que simplifica las manipulaciones a efectuar;
- el procedimiento según la invención permite, en una etapa, proceder a la clonación en el vector de expresión del banco de ácidos nucleicos introducido en la levadura, y aumentar la diversidad por recombinación homóloga u homeóloga entre los diferentes ácidos nucleicos del banco combinatorio introducido en la levadura;
- cuando el vector de expresión es multicopia, se obtiene una mezcla de productos en la levadura, que consiste en varias copias del vector, que posee cada una un gen mosaico diferente.. Cada clon de levadura obtenido contiene por lo tanto individualmente un banco de genes mosaicos, y esto permite ensayar las actividades de las diferentes proteínas de manera más rápida y eficaz;
- cuando el vector de expresión puede replicarse igualmente en *E. coli*, se puede entonces efectuar la segregación de los diferentes plásmidos al preparar el ADN plasmídico de al menos un clon de levadura obtenido, al transformar *E. coli* con dicho ADN plasmídico extraído, y al seleccionar los clones transformados en el medio adecuado para obtener una discriminación de los elementos del banco combinatorio de expresión funcional.

Así, el experto en la técnica que desee mejorar las propiedades funcionales de una proteína podrá preparar, por el procedimiento según la invención, un banco combinatorio de expresión funcional en la levadura a partir de ácidos nucleicos de interés que pertenecen a una misma familia génica. Podrá luego ensayar los clones de levaduras para seleccionar aquellos para los cuales la propiedad buscada es aparente, y obtener las secuencias realmente interesantes al efectuar la discriminación por paso a un huésped procariota.

El procedimiento según la invención permite así producir proteínas funcionales mosaicas activas, siendo ellas mismas objetos de la invención. Así, un procedimiento de producción de proteínas funcionales mosaicas activas, caracterizado porque se construye un banco combinatorio de expresión funcional por un procedimiento según la invención, que expresa proteínas mosaicas, y que se seleccionan las proteínas funcionales mosaicas activas por estudio de su actividad, es igualmente un objetivo de la invención.

Preferentemente, las proteínas mosaicas que se busca generar son enzimas que poseen actividades mejoradas (termoestabilidad, función nueva, modificación de función, aumento de actividad, modificación de la especificidad del sustrato, modificación de la actividad en un medio preciso tal como un disolvente, un pH...). La utilización del procedimiento según la invención para generar nuevas enzimas presenta muchas ventajas, ya que las actividades de las nuevas proteínas generadas pueden entonces ser ensayadas directamente en la levadura. Se utiliza entonces preferen-

temente como ácidos nucleicos de partida, ácidos nucleicos que pertenecen a una misma familia génica, que codifican enzimas. Las proteínas mosaicas activas obtenidas se dicen entonces derivadas de enzimas.

Los ejemplos de la presente invención muestran la aplicación del procedimiento a la generación de nuevas proteínas derivadas de citocromos P450s. Los citocromos P450s (P450s) pueden reconocer una larga variedad de sustratos y catalizar un número todavía más grande de reacciones. Estas enzimas se muestran en prácticamente todos los organismos vivos (20). En los mamíferos, los P450s están implicados en la formación de hormonas esteroides pero tienen igualmente un papel predominante en el metabolismo de medicamentos y de contaminantes que pueden llevar a veces a procesos de toxicidad y de carcinogénesis químicas (20-22). Los P450s 1A1 y 1A2 humanos tienen del orden de 70% de identidad de secuencia y poseen ciertas especificidades de sustratos diferentes. Están dentro los P450s más activos en el metabolismo de carcinógenos químicos (23) y están implicados, en el ser humano, en el cáncer de pulmón para *CYP1A1* (24-26), en la activación de promutágenos contenidos en la alimentación (27) o en los cánceres de hígado inducidos por la aflatoxina B1 para el *CYP1A2*. El conjunto de las propiedades de los P450s de mamíferos han hecho en efecto excelentes candidatos para la aplicación de estas técnicas de evolución molecular (28).

Un caso particular de la presente invención se refiere por lo tanto al procedimiento según la presente invención, caracterizado porque además el vector de expresión eucariota utilizado para la mezcla recombinatoria contiene una fase abierta de lectura que codifica una enzima membranaria eucariota. Preferentemente, dicha enzima eucariota se elige entre el grupo constituido por los citocromos P450 eucariotas, enzimas de conjugación eucariotas (de fase II), miembros de la familia de transportadores ABC eucariotas.

En este caso, puede ser interesante utilizar una cepa de levadura que presenta una modificación genética que permite la sobreexpresión de al menos proteína elegida entre el grupo constituido por una P450 reductasa endógena o exógena, una adrenodoxina, una adrenodoxina reductasa, un citocromo b5 heterólogo, una enzima de fase II (en particular una epóxido hidrolasa). Tales cepas se describen en la patente EP 595 948. Estas cepas permiten en particular recrear el entorno natural de funcionamiento de los P450s eucariotas (40, 41).

La utilización de cepas de levaduras genéticamente modificadas permite además recrear complejos proteicos, con varios elementos fijos (expresados de manera constitutiva para la levadura), y un elemento variable (el producto de genes mosaicos obtenidos por el procedimiento según la invención).

El procedimiento según la presente invención puede aplicarse también a otras proteínas. Por ejemplo, puede ser interesante generar receptores, lo que permite determinar las secuencias implicadas en el reconocimiento y la asociación del ligando, o proteínas quimeras basadas en las proteínas dianas de los antibióticos, que permite determinar los grados de resistencia en función de mutaciones.

De manera habitual, es necesario efectuar numerosos ciclos de "DNA-shuffling" antes de obtener una proteína que presenta las características y/o propiedades deseadas. En el caso presente, después de la selección de los clones de levadura que expresan proteínas que tienen una actividad próxima a la actividad buscada, es posible efectuar una simple reacción de PCR directamente sobre dichos clones, al utilizar iniciadores apropiados que flanquean las fases abiertas de lectura, y proceder a una nueva mezcla recombinatoria al repetir las etapas del procedimiento según la invención.

Sin embargo es deseable poder mejorar la rapidez de obtención de las propiedades deseadas, al efectuar una relación entre las estructuras secuenciales de las proteínas mosaicas obtenidas, y las estructuras funcionales de dichas proteínas. Esto permite entonces unir fácilmente las secuencias de ADN del gen, o las uniones entre las secuencias, a una función enzimática u otra (fijación de un sustrato, termofilia...).

La presente invención se refiere por lo tanto igualmente a un procedimiento de análisis de un banco combinatorio de expresión funcional, caracterizado porque comprende las etapas siguientes:

- a. transformación de una cepa de *Escherichia coli* con el ADN plasmídico extraído de la cepa de levadura o de un conjunto de levaduras,
- b. hibridación del ADN plasmídico contenido en cada uno de los clones individuales de *Escherichia coli* obtenidos de la etapa a. con una o varias sondas específica(s) de una secuencia parental.

Este procedimiento, mejoradas las etapas que se describirán a continuación, puede utilizarse sobre cualquier banco combinatorio, a partir del momento en que los diferentes ácidos nucleicos que forman el banco han sido discriminados.

La hibridación se efectúa sobre una macro- o una micro-red de ADN, estando constituida dicha red del ADN plasmídico contenido en cada uno de los clones individuales de *Escherichia coli* obtenidos de la etapa a., o de un producto de PCR de este, o bien de dichas sondas específicas, unida(s) a un soporte sólido, estando localizados cada uno de los ácidos nucleicos por su posición en dicha red.

En el primer caso, se fija el ADN plasmídico contenido en cada uno de los clones individuales de *Escherichia coli* obtenidos de la etapa a., o de un producto de PCR de este sobre un soporte sólido (vidrio, silicio, membrana apropiada (Nylon, nitrocelulosa)...). Los métodos de fijación del ADN son conocidos por el experto en la técnica, y

el ADN puede fijarse de manera más o menos sólida sobre el soporte utilizado. No siempre es necesario extraer el ADN plasmídico en los clones de *E. coli* obtenidos, pudiendo estos ser lisados directamente sobre el soporte sólido utilizado, o pudiendo efectuarse la PCR que permite la amplificación de los fragmentos que corresponden a los genes mosaicos directamente sobre los clones bacterianos en la extracción de ADN previa.

En el segundo caso, se fijan las sondas sobre el soporte sólido. Existen varios métodos para preparar un soporte que lleva las sondas. Se pueden sintetizar las sondas luego fijarlas sobre el soporte (pudiendo hacerse el direccionamiento mecánicamente, electrónicamente, por chorro de tinta...) o sintetizar las sondas directamente sobre el soporte (por direccionamiento fotoquímico o por chorro de tinta, por ejemplo). El experto en la técnica elegirá el método más apropiado para el resultado buscado.

En función del número de sondas utilizadas, se obtiene una huella de hibridación más o menos fina, para cada uno de los clones ensayados. Cuanto más elevado es el número de sondas, más fina será la huella obtenida. Se pueden elegir sondas que se sitúan de manera homogénea sobre toda la longitud del gen. Alternativamente, puede ser de provecho utilizar sondas que se determinan en un conjunto de zonas de secuencias que se saben que codifican regiones importantes para la función y/o actividad de la proteína. Así, se puede obtener una huella secuencial determinada.

Además, las condiciones de hibridación de las sondas varían en función del grado de especificidad de dichas sondas para cada estructura parental. Así, cuando dos estructuras parentales diferentes de una simple base sobre el fragmento que corresponde a la sonda, es necesario aplicar condiciones de restricciones más elevadas que si las estructuras parentales son muy diferentes. El experto en la técnica sabe determinar las mejores condiciones de hibridación, en particular siguiendo la enseñanza de Sambrook *et al.* Es igualmente importante notar que ciertos genes mosaicos pueden presentar una intensidad de hibridación con una sonda dada menor que otros genes. En efecto, la eficacia de la transferencia del ADN sobre el soporte sólido puede efectuarse de manera más o menos eficaz, o bien la región del gen sobre la que la sonda debe hibridarse es ella misma mosaica y constituida por fragmentos que provienen de genes "parientes" diferentes.

Se puede entonces proceder a un análisis estadístico de las intensidades de hibridación, con ayuda de un programa informático apropiado. El programa convierte primero las señales de hibridaciones en coordenadas de un tipo parental por un sistema de máscaras con una función booleana XOR antes del análisis estadístico propiamente dicho.

El análisis del banco combinatorio puede efectuarse de la manera siguiente:

- Se atribuye un código a cada secuencia nucleica generada, en función de la capacidad de las sondas utilizadas para hibridar dicha secuencia. Puede ser ventajoso utilizar un código binario (0 si el lugar sondado corresponde a un cierto tipo parental, 1 si corresponde a otro tipo parental), pero se pueden utilizar también otros tipos de codificación. Así, cada secuencia generada en el banco posee una "firma" individual. En el caso en el que se utilizan 6 sondas y se utiliza un código binario, se prevén 2^6 posibilidades (de 000000 a 111111)
- La frecuencia de cada una de las firmas así obtenidas se compara entonces con la frecuencia esperada si la mezcla recombinatoria de ADN se efectuara completamente al azar (en el caso de 6 sondas, la frecuencia teórica de cada dibujo es entonces de $1/2^6$). Este análisis permite definir un "pariente preferencial" para cada una de las posiciones sondadas (a veces se deben aportar ciertas correcciones, en particular cuando las proporciones de ácidos nucleicos parentales de partida no son iguales).
- El estudio de las firmas permite igualmente precisar las relaciones que pueden existir en el seno de una misma mosaica, en particular las asociaciones entre tipos parentales que pueden encontrarse entre cada segmento. Por ejemplo, es importante poder determinar fácilmente la necesidad de una correlación entre dos segmentos nucleicos no son necesariamente adyacentes para la obtención de una función biológica.
- El análisis puede afinarse igualmente con el fin de obtener resultados que pueden proporcionar varia información. Los ejemplos ilustran tal etapa al divulgar un método en el que cada firma del banco se convierte en un número decimal, y en la que se dibuja una curva, que lleva en abscisa dicho número decimal y la frecuencia acumulada en ordenadas. El análisis de dicha curva, y su modelización por simulación permite igualmente obtener información interesante sobre la probabilidad de obtener un cierto tipo de estructura parental en un lugar dado, y las correlaciones que existen entre diferentes fragmentos.

Los análisis estadísticos así descritos se facilitan gracias a la utilización de instrumentos informáticos, cuyo desarrollo no plantea problemas al experto en la técnica.

Las simulaciones de correlaciones entre diversos segmentos puede efectuarse generando rejillas más o menos aleatorias, según las correlaciones deseadas. Por ejemplo, una rejilla puede generarse para la que un segmento tiene una probabilidad superior a 50% de ser del mismo tipo parental que el segmento vecino. El número de rejillas que pueden generarse así es extremadamente importante, y puede permitir así definir una aproximación de los resultados observados.

Cuando se observan las correlaciones entre diferentes segmentos, es probable que la aplicación de una selección funcional sobre la población de clones (que reduce así la población de secuencias que pasan la criba) llevará a un aumento del número de correlaciones, y a una evolución (convergencia) de los resultados estadísticos obtenidos. Se debería obtener por lo tanto la aparición de un dibujo característico de la selección aplicada, lo que da una firma secuencial que depende de la selección funcional aplicada sobre el sistema.

Resumiendo, la presente invención se refiere igualmente a un procedimiento de análisis de huellas de hibridación que pueden obtenerse por el procedimiento de análisis del banco combinatorio descrito anteriormente, caracterizado porque comprende las etapas que consisten en:

- a. calcular la frecuencia de aparición de cada una de las combinaciones posibles,
- b. definir una firma de la repartición estadística de las combinaciones, por un tratamiento matemático y estadístico adecuado.

Así, la presente invención proporciona un medio de producir de manera eficaz bancos combinatorios de expresión funcional a partir de ácidos nucleicos que pertenecen a una misma familia génica en el sentido de la invención, que pueden poseer un grado de identidad relativamente bajo.

Además, la presente invención presenta la ventaja de poder efectuar el ensayo de las actividades de proteínas mosaicas producidas, directamente sobre los clones de levaduras obtenidas, sin etapa previa de purificación.

La presente invención proporciona igualmente un procedimiento de análisis de bancos combinatorios, basado en una hibridación y un análisis estadístico de las huellas de hibridaciones obtenidas.

La presente invención proporciona por lo tanto herramientas que pueden utilizarse para la determinación de las relaciones que pueden existir entre las estructuras secuenciales y las estructuras funcionales de las proteínas. Así, la presente invención se refiere igualmente a un procedimiento de determinación de relaciones entre firmas secuenciales y firmas funcionales de una proteína, caracterizado porque comprende las etapas que consisten en

- a. preparar un banco combinatorio de expresión funcional para un procedimiento según la invención,
- b. producir las proteínas funcionales mosaicas activas,
- c. analizar las diferencias funcionales y/o de actividad entre dichas proteínas mosaicas,
- d. analizar los ácidos nucleicos que corresponden a dichas proteínas mosaicas por un procedimiento de análisis por hibridación según la invención, seguido de manera opcional por un análisis estadístico por un procedimiento según la invención,
- e. relacionar las diferencias de estructura secuencial observadas en la etapa d. con las diferencias funcionales y/o de actividad observadas en la etapa c.

La realización de este procedimiento, para identificar las zonas de secuencias importantes o las relaciones entre zonas de secuencias unidas a una función de interés permite predecir las estructuras que tienen dicha función, por deducción de la estructura buscada, en función de la relación estructura-función obtenida por el procedimiento descrito anteriormente.

Así, es posible obtener proteínas que poseen propiedades mejoradas, tal como se han descrito anteriormente, o proteínas que reconocen un gran número de sustrato (enzimas "genéricas"), al pilotar las mezclas de informaciones genéticas, con el fin de obtener las proteínas de interés más rápidamente, y de manera más eficaz.

Los diferentes procedimientos descritos en el estado de la técnica permitían obtener proteínas de interés para una repetición de las mezclas recombinatorias de ADN, al someter las proteínas obtenidas a las cribas cada vez más finas. La presente invención, que permite relacionar las estructuras y las funciones de las proteínas mosaicas obtenidas, permite proceder a nuevas mezclas recombinatorias al utilizar, como ácidos nucleicos de partida, que los ácidos nucleicos que han sido identificados como que portan las estructuras u organizaciones de estructuras de interés.

Así, la presente invención se refiere un procedimiento de obtención de una proteína que poseen propiedades mejoradas, caracterizado porque comprende las etapas que consisten en:

- a. construir un banco combinatorio de expresión funcional para un procedimiento según la invención,
- b. analizar dicho banco combinatorio de expresión funcional,
- c. analizar las huellas de hibridación obtenidas en la etapa b. por un procedimiento según la invención,

d. determinar las relaciones entre las estructuras secuenciales y las estructuras funcionales de las proteínas por comparación de dichas huellas de hibridación con las propiedades de las proteínas mosaicas correspondientes,

e. predecir las estructuras de interés o las organizaciones de estructuras en las proteínas mosaicas,

f. repetir las etapas a. a e., al utilizar, como ácidos nucleicos de partida para la generación del banco combinatorio de expresión funcional, portando los ácidos nucleicos las estructuras de interés o las organizaciones de estructuras identificadas en la etapa e., un número suficiente de veces para obtener la proteína que posee las propiedades mejoradas buscadas.

La etapa f. consiste en repetir las etapas anteriores hasta que una proteína que presenta las propiedades deseadas haya podido ser identificada. La presente invención, debería permitir disminuir el número de ciclos de fabricación de banco combinatorio - análisis de las proteínas con respecto a los métodos de la técnica anterior.

Las proteínas obtenidas por el procedimiento descrito son igualmente un objeto de la invención.

La invención se refiere igualmente a un procedimiento de determinación de una estructura de una proteína importante en respuesta a una presión de selección a partir de un banco combinatorio de expresión funcional obtenido por un procedimiento según la invención, para los elementos de los que se ha obtenido una firma, que comprende las etapas de:

- normalizar dicho banco, por la homogeneización de las firmas, por ejemplo por cribado con un ayuda de un aparato robótico apropiado. Esta etapa permite asegurar que cada huella se encuentra con la misma probabilidad en el banco normalizado.
- aplicar una presión de selección,
- analizar el banco de expresión resultante al realizar los procedimientos de análisis de firma de secuencia según la invención,
- estudiar los cambios de firmas de secuencia inducidos por la presión de selección sobre el banco normalizado inicial y deducir las estructuras seleccionadas o contra seleccionadas en respuesta a la presión de selección.

Hay que notar que el hecho de normalizar el banco antes de aplicar la presión de selección permite de hecho cribar una diversidad más importante al cribar el mismo número de clones que si no se normalizara. En efecto, se puede notar que ciertas estructuras (tal como se han analizado para las huellas) están presentes en probabilidades superiores a lo que se esperaría en caso de mezcla aleatoria. La normalización permite por lo tanto disminuir la influencia de este problema.

Los ejemplos que siguen se limitan a la generación de nuevos citocromos P450s, para ilustrar la invención. Sin embargo, no deben considerarse como limitantes de la invención, y en particular el tipo de proteínas y de ácidos nucleicos pudiendo utilizarse en los procedimientos descritos en la presente invención. El experto en la técnica puede así realizar fácilmente los procedimientos de la invención, al sustituir otros genes a los genes de los citocromos P450 descritos en los ejemplos.

Descripción de las figuras

Figura 1: *principio de la construcción de los bancos*. A: línea 1, marcador ADN (ADN de λ digerido por *Pst* I); las líneas 2, 3, 4 y 5, 6, 7 corresponden respectivamente a los plásmidos p1A1/V60 y p1A2/V60 digeridos con la ADNasa I. Las líneas 2 y 5 corresponden a la fragmentación con 0,0112 unidades, las líneas 3 y 6 con 0,0056 unidades y las líneas 4 y 7 a 0,0028 unidades de ADNasa I por μ g de ADN. B: reacción de reensamblaje. Línea 1, marcador ADN; las líneas 2, 3 y 4 corresponden a las reacciones de reensamblaje entre fragmentos de p1A1/V60 y p1A2/V60 al mezclar respectivamente las reacciones de las líneas 2 y 5, 3 y 6, 4 y 7. C: reacción de amplificación. Línea 1, marcador ADN; las líneas 2, 3 y 4 corresponden respectivamente a la amplificación con los plásmidos PYP60, p1A1/V60 y p1A2/V60; las líneas 5, 6 y 7 corresponden a la amplificación con el ADN previamente reensamblado utilizado como matriz (líneas B2, B3 y B4). La banda presentada en la línea 6, panel C se purificó y se utilizó como tal para cotransformar *S. cerevisiae* con plásmido pYP60 previamente linealizado. Se observa la existencia de sucesos de recombinación entre los diferentes ácidos nucleicos del banco introducido en la levadura.

Figura 2: posiciones respectivas y secuencias de seis sondas utilizadas para realizar las matrices de caracterización del banco. Los números por arriba o por abajo corresponden a la posición 5' de alineamiento de cada sonda sobre las secuencias. Las sondas por lo alto o lo bajo hibridan respectivamente las secuencias del P450 1A1 o del P450 1A2. Las barras verticales en el rectángulo central representan todas las posiciones de desapareamiento entre la secuencia del P450 1A1 y del P450 1A2.

Figura 3: Los resultados de hibridación han sido tratados con Microsoft Excel generando una rejilla de 384 puntos con el código de color siguiente: Los cuadrados oscuros representan estructuras asimiladas a estructuras de tipos parentales (1A1 o 1A2) para las zonas de secuencias correspondientes a las seis sondas y los cuadrados claros representan estructuras mosaicas.

Figura 4: Frecuencias acumuladas experimentales y teóricas para la observación de los 64 tipos de estructuras mosaicas posibles. El eje horizontal corresponde a una codificación de las estructuras mosaicas al utilizar $N = P1 + 2*P2 + 4*P3 + 8*P4 + 16*P5 + 32*P6$, donde P1 a P6 tienen los valores de 0 ó 1 dependiendo respectivamente de la hibridación con las secuencias 1A1 o 1A2. Los círculos abiertos representan las curvas experimentales deducidas de los estados de hibridación de la rejilla de 384 clones con las seis sondas oligonucleótidas. La curva continua corresponde a curvas teóricas al considerar una proporción homogénea de 0,56:0,44 para las secuencias parentales 1A2 y 1A1 parental secuencias y una mezcla perfecta (ausencia de correlación cruzada). La curva en punteado representa la misma curva para una proporción de 50:50 para las secuencias parentales 1A1 y 1A2. Los círculos negros representan la curva teórica obtenida para simulaciones al considerar una proporción homogénea de 0,56: 0,44 para las secuencias parentales 1A2 y 1A1 pero una probabilidad de relaciones parentales de 0,1: 0,6: 0,85: 0,1: 0,1 entre los segmentos sondados 1-2, 2-3, 3-4, 4-5 y 5-6 respectivamente. La relación se define como sigue: 0 corresponde a la independencia y 1 a una relación total.

Figura 5: Representación de las frecuencias parentales y recombinantes para la asociación entre dos sondas. La frecuencia de cada asociación se ha determinado mediante una de las macros generadas en Microsoft Excel. La suma de las cuatro diferentes frecuencias (parentales y recombinantes) es siempre 1. A: asociación entre dos sondas adyacentes; B: asociación entre sondas separadas por una sonda; C: asociación entre sondas distantes (separadas por dos o tres sondas). Los histogramas negros y gris oscuros representan las asociaciones parentales mientras que el gris claro y el gris semi-oscuro representan las asociaciones recombinantes.

Figura 6: Detección colorimétrica de estructuras mosaicas funcionalmente competentes para la oxidación del naftaleno. La bioconversión se realizó en 1 ml de cultivo de levaduras en presencia de 1,6 mM de naftaleno. La extracción en fase sólida y el desarrollo de la coloración se realizó enteramente en placas de microvaloración como se describe en los Ejemplos. La coloración oscura indica los clones positivos.

Figura 7: Representación esquemática de las secuencias de 10 estructuras mosaicas seleccionadas al azar: A en la población total; B en la sub-población de clones activos. Parra cada estructura se realizó un alineamiento con las dos secuencias parentales. Estos alineamientos se utilizaron como datos de partida para un programa de análisis de secuencias y un programa de visualización que ha generado la figura. Las zonas grises y negras corresponden respectivamente a secuencias que pertenecen a los P450 parentales 1A1 o 1A2. Los trazos verticales finos inferiores o superiores indican las zonas de desapareamiento nucleotídicos con la segunda estructura parental. Las marcas que atraviesan las secuencias indican las posiciones de secuencias que no se aparean con ninguna de las dos secuencias parentales y que deben corresponder por lo tanto a mutaciones. Las partes horizontales transparentes corresponden a segmentos de secuencias para las que la pertenencia a uno u otro de los tipos parentales no ha podido determinarse por análisis de secuencias.

Ejemplos

Ejemplo 1

Métodos

1.A: *Cepas, plásmidos y biología molecular*

Se utilizaron dos cepas de *S. cerevisiae*: W303-1B, igualmente denominada W(N) (Mat a; *ade2-1; his3, leu2, ura3, trp1, can^R, cyr⁺*), y W(R) que deriva de W(N) por la inserción del promotor inducible *GAL10-CYC1* más arriba de la P450 reductasa endógena (*YRED*) de levadura. Esta cepa se describió anteriormente por Truan et col. (40) y en la patente EP 595 948, incorporadas aquí como referencia.

La cepa de *E. coli* utilizada fue DHS-1 (F^- , *recA1, gyrA96, thi-1, hisR17, supE44, λ^-*). Los vectores de expresión utilizados fueron p1A1/V60 (42) y p1A2N60 (43, incorporada aquí como referencia); estos dos vectores se construyeron por inserción de los ORFs de los *CYP1A1* y *CYP1A2* humano entre los sitios de restricciones *BamHI/KpnI* y *BamHI/EcoRI* de pYeDP60 respectivamente. Estos dos vectores de expresión contienen igualmente *URA3* y *ADE2* como marcador de selección y sitúa las fases abiertas de lectura (ORFs) bajo el control del promotor *GAL10-CYC1* y del terminador *PGK* (39, incorporada aquí como referencia). Todos los medios utilizados se describieron anteriormente en los documentos incorporados aquí como referencia (40, 42).

Las bacterias DH5-1 se convirtieron en electrocompetentes según el protocolo descrito por Sambrook *et al.* (44) incorporada aquí como referencia, y las células transformadas siguiendo las recomendaciones del constructor del electroportador (Biorad). Las células se seleccionaron sobre medios LB sólidos que contenían 50 μ g/ml de ampicilina.

ES 2 301 553 T3

Transformación de levadura

Después de un precultivo de 12 horas en 5 ml de medio YPGA (para la cepa W(N)) o YPLA (para la cepa W(R)), las células se diluyeron en 50 ml de medio YPGA para obtener una densidad final de 2.10^6 células/ml. Seis horas más tarde, las células se lavaron dos veces con agua estéril y una vez con tampón TE-acetato de litio (Tris-HC 10 mM pH 7,5, EDTA 1 mM, acetato de litio 100 mM). Las células se resuspendieron luego en 1 ml de tampón TE-acetato de litio.

El ADN transformante se añadió luego a 50 μ l de la solución de células anteriormente obtenida, así como 50 μ g de ADN de esperma de salmón (previamente sonicado y desnaturalizado a 95°C), y 350 μ l de una solución de PEG 4000 al 40% (p/v). Esta solución se incubó luego a 30°C durante 30 minutos y se sometió a un choque térmico a 42°C durante 45 minutos. Después de centrifugación, se eliminó el sobrenadante y las células se resuspendieron en 200 μ l de una solución NaCl 0,1M. Las células se seleccionaron luego sobre un medio SWA6 sólido (39, 42, incorporadas aquí como referencia).

Extracción del ADN plasmídico de levadura

Las colonias se resuspendieron en 1 ml de tampón A que contenía 20% (v/v) de triton X-100, 50 mM de Tris-HCl pH 8,0, 50 mM de EDTA y 200 mM de NaCl. Luego, se añadió 1 volumen de bolas de vidrio (Braun Scientifics, diámetro 0,45 mm) y la solución se agitó vigorosamente con aparato Vortex durante 2 min con 300 μ l de una mezcla fenol/cloroformo/alcohol isoamílico (50:49:1, en vol.). Después de recuperación de la fase acuosa, el ADN se precipitó en etanol y se resuspendió en 50 μ l de agua.

Secuencias

Se seleccionaron previamente cinco clones bacterianos derivados del banco inicial y cinco clones funcionales y se secuenciaron. Las secuencias se realizaron bien por ESGS (ESGS, grupo Cybergene, Evry Francia) o utilizando el kit ABI y el secuenciador ABI siguiendo los protocolos del constructor (Perkin Elmer).

1.B: Mezcla recombinatoria de ADN basado en PCR modificada

La técnica utilizada deriva de la descrita por Stemmer (2, 3, 15), incorporadas aquí como referencia. La fragmentación aleatoria con la ADNasa 1 (Grado II, Sigma-Aldrich) en presencia de Mn^{2+} y realizada con las modificaciones descritas por Lorimer y Pastan (45) y Zhao (46), incorporadas aquí como referencia.

Se resuspendieron 2,5 μ g de cada ADN plasmídico (pIA1/V60 y pIA2/V60) separadamente en un tampón que contenía 50 mM de Tris-HCl pH 7,4, 10 mM de $MnCl_2$ para un volumen final de 40 μ l. La ADNasa 1 se añadió en tres concentraciones diferentes (0,0112 U/ μ g de ADN, 0,0056 U/ μ g de ADN y 0,0028 U/ μ g de ADN). La digestión se realizó a 20°C durante 10 min y la ADNasa 1 se inactivó a 90°C durante 10 min. Los fragmentos obtenidos se purificaron sobre una columna Centriscp (Princeton Separation Inc., Philadelphia, NJ).

Durante la reacción de reensamblaje, los fragmentos purificados (10 μ l de cada plásmido fragmentado) se amplificaron por una reacción de PCR en 40 μ l al utilizar 2,5 U de Taq-polimerasa (Stratagene).

El programa de PCR utilizado consistía en: 1 ciclo de desnaturalización a 96°C, durante 1.5 min; 35 ciclos de (30s de desnaturalización a 94°C, 9 etapas diferentes de hibridación separadas 3°C yendo de 65°C a 41°C y de 1,5 min cada una y una etapa de elongación de 1.5 min a 72°C) y finalmente 7 min a 72°C.

La segunda reacción de amplificación se realizó con iniciador 5' localizado en el promotor *GAL10-CYC1* (SEQ ID N° 1) y un iniciador 3' localizado en el terminador PGK (SEQ ID N° 2).

1.C: Construcción y caracterización del banco

Los productos de amplificación por PCR se separaron por gel de electroforesis luego se purificaron. Los ADNs se insertaron en pYeDP60 utilizando la recombinación *in vivo* (reparación por recombinación) en la levadura (37, 38, 43, 47, 48). Se realizó la cotransformación de la cepa W303-IB con 1/20° del producto de PCR (inserto) y 0,025 μ g de pYeDP60 previamente linealizado utilizando las enzimas de restricción *EcoRI* y *BamHI*.

El ADN extraído de la levadura se utilizó para transformar la cepa de *E. coli* DH5-1 utilizando la resistencia a la ampicilina aportada por el plásmido. Se inocularon 378 pocillos de una placa de microvaloración de 384 pocillos con colonias bacterianas independientes elegidas al azar en el banco, 3 pocillos con bacterias DH5-1 previamente transformada con pIA1/V60 y los 3 pocillos restantes con DH5-1 transformados por pIA2/V60. Después de 24 horas de crecimiento en medio TB (44) que contenía 100 μ g/ml de ampicilina, los 384 pocillos se replicaron luego sobre seis membranas de Nylon N+ (Amersham). Cada filtro se situó en un medio LB sólido que contenía 100 μ g/ml de ampicilina. Después de 12 horas de crecimiento, la lisis de las colonias bacterianas, la fijación y la desnaturalización del ADN, la prehibridación de los filtros se realizaron según el protocolo preconizado por el fabricante (Amersham).

ES 2 301 553 T3

Se añadieron 11 pmoles de oligonucleótidos a 3,3 pmoles de γ -ATP marcadas con ^{32}P , a 2 μl de polinucleótido quinasa y 18 μl de tampón (New England Biolabs). El conjunto se incubó 2 h a temperatura ambiente. La prehibridación de los filtros se realizó según el protocolo preconizado por el fabricante. La sonda marcada se ajustó en un tubo de hibridación que contenía uno de los filtros y el conjunto se incubó durante 12 h a 42°C. Los filtros se lavaron luego en una solución de SSPE 2x/0,1% de SDS durante 10 min. Los filtros se analizaron por autoradiografía, según un protocolo conocido.

Cada sonda se marcó una segunda vez y se hibridó con un filtro diferente para asegurar la reproducibilidad de los resultados.

1.D: Selección de los clones que contenían P450s funcionales

Las colonias bacterianas crecieron durante 24 horas en placas de microvaloración de 96 pocillos. La extracción del ADN se realizó utilizando el protocolo del aparato Multiscreen de minipreparación de ADN por filtración en microplacas de 96-pocillos (Millipore). Cada ADN purificado se utilizó para transformar la cepa de levadura W(R) en placa de microvaloración de 96 pocillos y las células se seleccionaron sobre medios SWA6 sólidos.

Después de 3 días de crecimiento a 30°C, se sembró 1 ml de medio líquido SWA5 con una alícuota de cada colonia en una microplaca de 96 pocillos Deepwell (ABGene) durante 15 horas. El medio se eliminó luego y se reemplazó por 1 ml de medio YPLA que contenía 1,6 mM de naftaleno (Merck).

Para cada cultivo, el medio de cultivo se situó luego en los pocillos correspondientes de una microplaca de 96 pocillos Multiscreen (MABV N12, Millipore) que contenía 90 μl de resina de gel sílice C18 octadecilo funcionalizada (Aldrich). Después de una filtración a vacío del medio de cultivo, el sustrato y los productos de la reacción se unieron a la sílice. La resina se lavó luego 2 veces con agua y los metabolitos se eluyeron con 50 μl de isopropanol. Después de adición de 20 μl de una solución de Diazo-Blue-B a 2 mg/ml (Fluka) se observó la reacción coloreada generada por el acoplamiento entre los precursores de diazo y los fenoles extraídos del medio de cultivo.

1.E: Análisis estadísticos

Para cada sonda, se construyó una rejilla que representaba las intensidades de hibridaciones de los 384 clones. Las intensidades de las hibridaciones se analizaron visualmente teniendo en cuenta el ruido de fondo próximo. Las manchas que excedían muy ampliamente el ruido de fondo local de las manchas negativas se consideraron positivas, incluso si eran menos intensas que las manchas más positivas. Estas respuestas intermedias que se podían deber a un desapareamiento parcial de la sonda (consecutivo de etapas de PCR) o también a una eficacia de transferencia menos importante de ciertas manchas sobre el filtro. Las ambigüedades se solucionaron mediante la realización de hibridación de otro filtro por la misma sonda.

Las seis rejillas de 384 pocillos se introdujeron en las tablas de Microsoft Excel y se realizó un análisis estadístico con macros Excel escritas en Microsoft Visual Basic, y retomando las etapas de análisis tal como se han descrito en la descripción. El programa convierte primero las señales de hibridaciones en datos de un tipo parental por un sistema de máscaras con una función booleana XOR antes del análisis estadístico. Los análisis estadísticos se realizaron siguiendo las etapas enumeradas en la descripción.

Se realizaron simulaciones numéricas utilizando un generador de números aleatorios y rutinas de cálculos de probabilidad. El programa puede ajustarse para simular todos los errores posibles en la probabilidad de encontrar uno u otro de los tipos parentales para las zonas de secuencias correspondientes a cada una de las sondas, así como todas las “relaciones” posibles entre segmentos adyacentes o distantes. Un primer juego de parámetros permitió modular las probabilidades relativas de encontrar uno u otro de los tipos parentales para cada zona de secuencia sondeada. Un segundo juego de parámetros permitió introducir de una (o varias) relación genética entre dos fragmentos (o más) de secuencia (correspondiente a dos o más sondas).

Los programas de simulaciones y de análisis estadísticos se utilizaron para generar rejillas correspondientes a diferentes situaciones de entre fragmentos. En todos los ensayos, los resultados de los análisis estadísticos estuvieron en concordancia con los parámetros introducidos en el programa de simulaciones. El método de asociación de estas técnicas de simulación y de análisis se utilizó igualmente para determinar las fluctuaciones estadísticas sobre los datos al realizar análisis de 10 ciclos repetidos de simulaciones y de análisis para cada juego de parámetros. El generador de números aleatorios se reinició entre cada simulación para hacer sucesos independientes.

Ejemplo 2

Construcción de un banco de expresión para mezcla recombinatoria de ADN de una misma familia

El principio de la estrategia utilizada está descrita en la Figura 1: asocia una etapa de mezcla recombinatoria de ADN *in vitro* por PCR modificada en una segunda etapa de mezcla recombinatoria *in vivo* por recombinación en la levadura. Esta última etapa se utilizó igualmente como un instrumento de clonación eficaz. Esto constituye una estrategia de shuffling completa que permite la expresión en una célula eucariota y la selección funcional sin tener necesidad de etapa de clonación intermedia en *E. coli*.

ES 2 301 553 T3

La primera etapa (Figura 1) consiste en una fragmentación doble hélice del plásmido entero con la ADNasa 1 que conduce a fragmentos de ADN de pequeños tamaños (Figura 1A).

Los resultados de la fragmentación de los plásmidos p1A1/V60 y p1A2/V60 (figura 1A, líneas 2 y 5; 3 y 6 4 y 7) se mezclaron en proporción equimolar y se sometió a un programa de PCR original “de hibridaciones progresivas” (véase Ejemplo 1) que implica 9 etapas de hibridación que van de 61°C a 41°C para forzar la recombinación entre fragmentos que tienen pocas homologías. Como se muestra en la figura 1B, en tales condiciones, se formó una larga traza (smear en inglés) de ADN de alto peso molecular cualquiera que fueran los fragmentos tomados al principio.

Aunque se haya encontrado que este material tenía propiedades de transformación directa de la levadura debido a una recombinación entre fragmentos *in vivo* y la reconstitución de vectores de levadura completos y funcionales (11 kb) (resultados no mostrados), una nueva etapa de PCR, utilizando iniciadores situados sobre las secuencias de ADNc flanqueantes de iniciación de la transcripción CYC1 y las secuencias de terminación de la transcripción PGK, fue necesaria para obtener un banco de tamaño razonable (Figura 1C, líneas 5, 6 y 7). Esta última etapa dio como resultado una amplificación de banda de ADN bien definida de aproximadamente 1,9 kb que comprendía el ADNc “mezclado-recombinado” y las regiones flanqueantes del vector.

El producto de PCR mostrado en la figura 1C, línea 6 se utilizó para cotransformar la levadura con pYeDP60 linealizado en el sitio de expresión para utilizar las propiedades de recombinación homólogas (reparación por recombinación) de la levadura.

La cotransformación del banco del ADNc de buen tamaño en la levadura y del vector linealizado ha llevado a una serie de sucesos de recombinación ya observados durante experiencias precedentes de recombinación homeólogas o reparación por recombinación (37, 38, 43). La selección se basó solamente en la recirculación del vector después de uno o varios sucesos de recombinación. Las experiencias dieron aproximadamente 10,000 clones.

La mayoría de los clones de levadura están transformados por varios plásmidos. En efecto, se observó una población heterogénea de plásmidos después de extracción de ADN de una sola colonia de levadura, la transformación de *E. coli* y la segregación de los clones.

Esto permitió evaluar la complejidad del banco inicial entre 25,000 y 100,000 estructuras mosaicas para una sola experiencia de transformación de levadura. El banco es utilizable como tal para la selección funcional.

Experiencias similares utilizando fragmentos ADN de más bajo peso molecular (menos de 100 pb) como se describe en la Figura 1A, líneas 1 y 5 han conducido también a un banco explotable pero con una eficacia inferior. Los ADNs de más alto peso molecular (Figura 1A, líneas 4 y 7) no han sido utilizadas para la construcción de un banco a causa de las posibilidades de una fuerte tasa de contaminaciones por estructuras parentales.

Ejemplo 3

Análisis estadístico de una sub-población del banco

El ADN plasmídico se preparó a partir del banco de levadura y se utilizó para transformar *E. coli* utilizando el marcador de resistencia a la ampicilina presente sobre el plásmido de levadura. Esta etapa permitió la segregación de plásmidos individuales que estaban inicialmente presentes como una población heterogénea en cada colonia de levadura. Se construyó una matriz a partir de una placa de microvaloración de 384-pocillos que contenía 378 clones de *E. coli* elegidos al azar para análisis de estructuras utilizando 6 sondas repartidas a lo largo de la secuencia de los P450s parentales descritos sobre la figura 2 (SEQ m N° 3 a SEQ ID N° 8). Los pocillos restantes se sembraron con bacterias previamente transformadas con plásmidos controles que contenían una u otra de las secuencias parentales (P450 1A1 o 1A2).

Las seis sondas (22-36 bases) se eligieron para hibridar alternativamente sobre las dos secuencias parentales en regiones de bajas similitudes de secuencias entre los dos P450s parentales: 3 sondas pertenecían al p1A1/V60 y 3 al p1A2/V60. Cada sonda se marcó con ^{32}P y se utilizó para hibridar las réplicas sobre filtros (en condiciones que favorecerían las hibridaciones específicas). Las experiencias se repitieron utilizando diferentes asociaciones de filtros y de sondas para eliminar eventuales artefactos. El análisis de las intensidades de hibridación se realizó manualmente. Los niveles intermedios de intensidad de hibridación (del orden de 15% de las manchas) se consideraron como respuestas positivas. Estas respuestas deben corresponder a desapareamientos de un par de bases debido a mutaciones inducidas por las diferentes etapas de PCR (confirmándose esto por los datos de secuenciación (véase más adelante)) o también a diferencias de eficacia de transferencia de ADN.

La figura 3 presenta el dibujo global de hibridación para las seis sondas. La frecuencia de estructuras que presentan un dibujo de hibridación similar a uno de los parientes (de ahora en adelante denominados “parentales”) para todas las sondas calculadas en el banco (figura 3A, cuadrados oscuros) es de 11,4% para las estructuras que corresponden a P450 1A2 y de 2,4% para las estructuras que corresponden a P450 1A1. La suma de estas dos frecuencias (13,8%) es superior al valor teórico de 3,1% $((0,5)^6 + (0,5)^6)$ que corresponde a una reasociación totalmente aleatoria de los fragmentos de secuencias parentales. Una ilustración en “falsos colores” de diferentes estructuras mosaicas (no mostrada) ilustra bien

ES 2 301 553 T3

el exceso de clones parentales de tipo 1A2 o de tipo 1A1 pero sugiere una repartición general bastante homogénea de diferentes tipos de estructuras mosaicas.

Con el fin de ir más lejos en la caracterización de la población, se realizó un análisis estadístico utilizando un programa basado en tablas Excell y rutinas en Visual Basic. Se calculó la probabilidad de presencia de cada secuencia parental en cada una de las 6 posiciones sondadas (Tabla 1). Esta frecuencia fue bastante homogénea ($0,56 \pm 0,02$ para los fragmentos de tipo 1A2) para el conjunto de los segmentos analizados. El pequeño exceso en la frecuencia para los segmentos de tipo 1A2 refleja probablemente la evaluación de las tasas de ADN parentales durante la mezcla de los fragmentos de ADN parentales. La proporción teórica de las secuencias parentales se volvió a calcular con nuevos valores de frecuencia: $3,7\%$ ($0,58^6 + 0,42^6$). Este último valor no corresponde siempre a la proporción de parentales observados ($13,8\%$).

TABLA 1

Frecuencia de las partes de secuencias mosaicas que pertenecen a cada tipo parental, en las posiciones sondadas. Las sondas P1 a P6 comienzan en las posiciones respectivas de las secuencias de P450 1A1 o 1A2, según la sonda considerada: 3, 612, 683, 879, 1377 y 1513 (véase la figura 2). Para cada sonda, el número de señales de hibridación relativo a 1A1 o a 1A2 se calculó y se dividió por el número total de clones ensayados (378)

Sonda	Frecuencia del tipo 1A1	Frecuencia del tipo 1A2
P1	0,48	0,52
P2	0,43	0,57
P3	0,45	0,55
P4	0,45	0,55
P5	0,44	0,56
P6	0,41	0,59
Media $\pm$ D. E.	$0,43 \pm 0,02$	$0,56 \pm 0,02$

Para caracterizar más en detalle la población se calculó la curva de las frecuencias acumuladas para la probabilidad de observación de las 64 clases de quimeras (figura 5). Se utilizó un código binario que asocia arbitrariamente un valor de 0 ó 1 según la naturaleza de cada segmento (1A1 o 1A2), para los segmentos 1 a 6, para cada estructura mosaica. Las secuencias parentales 1A1 y 1A2 corresponden respectivamente a los códigos 0 y 63. La curva experimental (figura 5, círculos vacíos) tiene un aspecto irregular que comprende cinco escalones. La aparición de estos escalones era completamente inesperada, e imprevisible, ya que no correspondían a lo que se hubiera esperado en el caso de una independencia de recombinación entre los diferentes fragmentos.

Se calcularon entonces tres curvas teóricas como se describe en el Ejemplo 1 utilizando aproximaciones de tipo Monte Carlo (simulaciones numéricas) utilizando diferentes hipótesis:

- (i) una probabilidad igual de encontrar los diferentes tipos parentales en las zonas de secuencias que corresponden a las diferentes sondas y una independencia total de la naturaleza de cada segmento de secuencia,
- (ii) hipótesis (i) pero con una probabilidad de encontrar fragmentos de tipo 1A2 en las zonas de secuencias que corresponden a diferentes sondas de $55,8\%$;
- (iii) hipótesis (ii) pero la probabilidad de mezcla recombinatoria entre los diferentes segmentos de secuencia ya no es infinita (mezcla imperfecta), pero con las relaciones variables entre la naturaleza de segmentos consecutivos.

La curva de frecuencia acumulada (figura 5) que corresponde a la hipótesis (i) es lineal mientras que en el caso que corresponde a la hipótesis (ii) la curva se redondea pero sigue siendo irregular. Esta curva, (que refleja el porcentaje real de fragmentos parentales) reproduce efectivamente bien la tendencia general de la curva calculada a partir de los resultados experimentales, pero no presenta los escalones observados.

Numerosas curvas que corresponden a la hipótesis (iii) se generaron con diferentes tipos de relaciones entre segmentos y se encontró una curva que correspondía a la curva experimental (círculos llenos). La adición de relaciones genéticas apropiadas entre las secuencias sondadas permite determinar una curva correspondiente que sigue la curva

experimental. Por supuesto, varias soluciones deben ser posibles, pero una probabilidad de relación entre fragmentos parentales de 0,1; 0,6 0,85 0,1 0,1 entre los segmentos sondados 1-2, 2-3, 3-4, 4-5 y 5-6 respectivamente, da un resultado satisfactorio. Estos resultados sugieren que, incluso si la proporción de cada tipo parental a lo largo de la secuencia es homogénea, la probabilidad de mezcla recombinatoria depende del segmento de secuencia considerada. Así, los escalones de la curva de resultados obtenida corresponden a una correlación entre diferentes segmentos de secuencia.

El cálculo de las frecuencias de cada tipo parental en la población se simuló después de la incorporación de las probabilidades de relaciones en el modelo. Los resultados medios que resultan de 10 simulaciones informáticas dan una frecuencia de estructuras de tipo parental de $13,9 \pm 1,3\%$ (del cual $9,8 \pm 1,4\%$ para el 1A2 y $4,1 \pm 1,09\%$ para 1A1), lo que corresponde bastante bien a los valores experimentales de 13,8%; (11,4% para el 1A2 y 2,4% para el 1A1). La heterogeneidad de la probabilidad de mezcla recombinatoria a lo largo de la secuencia puede ser perfectamente responsable por lo tanto del exceso aparente de estructuras de tipo parental en la población.

Con el fin de verificar la existencia de relaciones entre fragmentos, se analizaron las asociaciones entre las diferentes sondas. La figura 6 presenta la frecuencia de las asociaciones de zonas de secuencia de mismo tipo parental y de tipo parental diferentes para cada una de las asociaciones de sondas posibles.

En la figura 6.A, es posible ver las probabilidades de asociaciones próximas (entre zonas adyacentes). Esto pone claramente en evidencia que las asociaciones P1-P2, P4-P5 y P5-P6 muestran una independencia completa a la diferencia de asociaciones P2-P3 y P3-P4 que muestran una disminución de la frecuencia de asociación entre fragmentos de tipo parental diferentes.

La figura 6.B muestra la asociación entre dos sondas espaciadas por una sonda. De nuevo se puede observar una asociación que muestra una relación casi completa entre P2 y P4. Las otras asociaciones presentan independencias completas entre sondas.

Esto es igualmente verdadero para asociaciones entre sondas más alejadas (figura 6.C). Se calcularon otras asociaciones a largas distancias (P1-P5; P2-P6 y P1-P6), revelan las mismas características que las de la figura 6.C y no se muestran aquí.

Estos resultados confirman bien el modelo predictivo incluso si el número de relaciones en el modelo es de solamente 2. De la manera sorprendente los valores obtenidos para estos datos no corresponden a un modelo genético. En efecto, la distancia (entre los segmentos unidos) parece más importante en el caso de P2-P4 comparado con P2-P3 o P3-P4. Una explicación posible de este fenómeno puede estar ligado al número posible de sobrecruzamientos en este intervalo (P2-P4).

La existencia de escalones correspondientes a una correlación entre fragmentos, cuando se utiliza el análisis descrito más arriba permite sacar una conclusión importante. En efecto, cuando una presión de selección funcional se ejerce sobre los clones, es probable que introducirá un error más importante de correlaciones entre diferentes regiones de los genes estudiados. Así, puede ser posible definir dibujos de asociación entre varias regiones del gen, que están ligadas a actividades y/o propiedades funcionales. Esto debe permitir acelerar el proceso de definición de proteínas que presentan funciones y/o propiedades mejoradas, eligiendo las secuencias a asociar.

Ejemplo 4

Selección de clones funcionales

Una ventaja mayor de la estrategia de mezcla recombinatoria (shuffling) desarrollada en la presente invención es que el banco está, por primera vez, directamente construido dentro un microorganismo eucariota (la levadura). Es además posible utilizar cepas de levadura cuyo genoma se ha modificado con el fin de permitir reconstituir sistemas proteicos (enzimáticos) complejos.

En las experiencias de la presente invención, se utilizaron cepas de levaduras que poseían un genoma modificado, para permitir la reconstitución de un sistema membranario con acoplamiento de diferentes parejas. Los clones de levadura transformados que resultan de las etapas de shuffling pueden utilizarse entonces como tal para un cribado funcional de la actividad de las proteínas mosaicas construidas.

La utilización del banco primario ofrece además la ventaja de estar constituido por clones que contienen múltiples plásmidos mosaicos lo que mejora considerablemente la complejidad del banco, y permite cribar las actividades de varias proteínas mosaicas ensayando la actividad justo sobre un clon de levadura.

Sin embargo, está claro que los clones seleccionados para su funcionalidad necesitan una etapa de segregación suplementaria para un estudio bioquímico más detallado. Esta segregación puede realizarse mediante sub-clonaciones repetidas o mediante extracciones de ADN de los clones positivos, seguido de una transferencia en *E. coli* y de una retransformación de la levadura.

Las experiencias siguientes demuestran la factibilidad de una selección funcional directa *in vivo* en placas de microvaloración.

El método está basado en una técnica universal de detección por coloración de fenoles aromáticos formados por bioconversión *in vivo* directa de los hidrocarburos policíclicos aromáticos en los cultivos en microplacas de 96-pocillos (véase Ejemplo 1).

Los derivados fenólicos se extraen luego por fijaciones hidrófobas (sobre resinas de C18) directamente sobre microplacas y se revelan por colorimetría consecutiva por acoplamiento con precursores de colorantes diazo fast (Figura 7).

El cribado del banco mosaico 1A1/1A2 se realizó utilizando naftaleno que es un buen sustrato de las dos enzimas parentales. Con el fin de determinar la proporción real de estructuras funcionales del banco primario en la levadura se transfirió en *E. coli* y se utilizaron 96 clones independientes (y por lo tanto conteniendo más que un tipo de plásmido) para retransformar la levadura en placas de microvaloración. La frecuencia de clones funcionales en tales condiciones (12% para la librería construida con la Taq DNA polimerasa) se reconfirmó por métodos clásicos que utilizan análisis de productos extraídos por HPLC.

Estos controles han permitido ver que la detección colorimétrica es fiable y tiene una sensibilidad suficiente para detectar clones con una actividad naftaleno hidroxilasa que representa solamente 10% de la actividad parental (pudiendo deberse estas diferencias de cantidades de metabolitos producidos a diferencias de actividades pero igualmente de expresión de las enzimas mosaicas).

El método de detección utilizado se reveló igualmente muy eficaz para la detección de metabolitos derivados de su metabolismo del fenantreno o de otros hidrocarburos policíclicos aromáticos.

Ejemplo 5

Análisis de secuencias del banco

Se secuenciaron cinco clones seleccionados al azar independientemente de criterios funcionales y los cinco clones elegidos en la sub-población de clones funcionales (véase más adelante para la selección). Estas estructuras resultaron ser mosaicas que contenían, además, mutaciones adicionales.

Las estructuras mosaicas se describen en la figura 7. La figura está basada en un alineamiento entre las estructuras mosaicas y las dos secuencias parentales y se realizó con ayuda de un programa informático apropiado: para cada estructura se realizó un alineamiento nucleotídico con las dos secuencias parentales. Estos alineamientos se utilizaron como datos de partida para un programa de visualización que generó la figura, dibujando en gris o en negro las partes de secuencias que pertenecen respectivamente a los P450 parentales 1A1 o 1A2, y añadiendo trazos verticales finos inferiores o superiores para indicar las zonas de desapareamientos nucleotídicos con la segunda estructura parental. Además, los trazos que atraviesan las secuencias indican las posiciones de secuencias que no se aparean con ninguna de las dos secuencias parentales y que deben corresponder por lo tanto a mutaciones. El programa informático dibuja igualmente partes horizontales transparentes que corresponden a segmentos de secuencias para las que la pertenencia a uno u otro de los tipos parentales no ha podido determinarse por análisis de secuencias.

El análisis de estas 10 secuencias seleccionadas al azar confirma la presencia de estructuras mosaicas para cada secuencia. Analizando el conjunto de estas estructuras se puede notar un número medio de fragmentos diferentes de $5,4 \pm 2,2$. La distribución de tamaño de estos fragmentos es homogénea. Para los 54 fragmentos considerados, 32 tienen tamaños comprendidos entre 0 y 200 bp, 12 entre 200 y 500 bp y 10 entre 500 y 1000 bp. Además, aproximadamente 60% de los fragmentos son de tamaño inferior a 200 bp, siendo el tamaño del fragmento más pequeño intercambiado aproximadamente de 20 bp. Estos resultados están en concordancia con el tamaño medio de los fragmentos de partida derivados de la fragmentación con la ADNasa I (200-300 bp, véase figura 1A).

El análisis de la actividad naftaleno hidroxilasa de los 5 clones tomados al azar mostró que solo uno era activo (clon A1). A continuación se consideró como un clon activo, al igual que los 5 elegidos sobre criterios de actividad. La tasa media de mutaciones por secuencia se calculó para los clones activos e inactivos. Para los clones inactivos (A2, A3, A4, A5), el número medio de mutaciones es $14,0 (\pm 4,2)$. Para los clones activos es inferior ($8,3 \pm 3,2$). Esto no es sorprendente a causa del modo de selección (actividad). En efecto, las secuencias de los clones inactivos pueden contener codones de terminación precoces.

Finalmente, los diferentes resultados observados durante los análisis estadísticos se confirmaron para los datos de secuencias. Además, incluso si el número de clones secuenciados es bajo (10), los datos obtenidos proporcionan una vista detallada de algunas estructuras mosaicas. La relación entre fragmentos observada (entre las sondas 2, 3 y 4) durante los análisis estadísticos se observa igualmente en estas secuencias. En efecto, no se observa intercambio de fragmentos en la parte central que corresponde a dichas sondas.

La tasa elevada de mutaciones está en concordancia con una relativamente baja proporción de estructuras funcionales (15%) en la población. Sin embargo, experimentos similares de mezcla recombinatoria realizados al utilizar enzimas más fieles que la ADN polimerasa Taq tal como la Pfu y la ADN polimerasa Dynazyme EXT han dado una proporción más alta (80-90%) de estructuras funcionales. La tasa de mutaciones puede adaptarse así según lo que se desee.

Los ejemplos anteriores ilustran un aspecto de la invención y el experto en la técnica sabe aportar los ajustes necesarios para generalizar las enseñanzas, sin alejarse del espíritu de la invención.

Referencias

1. Van der Meer *et al.* (1992) *Microbiological Reviews*, 56(4), 677-94.
2. Stemmer, W. P. (1994) *Nature*, 370(6488), 389-91.
3. Stemmer, W. P. (1994) *Proc. Natl. Acad. Sci. USA*, 91 (22) 10747-51.
4. Cramer *et al.* (1997) *Nature Biotechnology*, 15(5), 436-8.
5. Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA*, 94 (9) 4504-9.
6. Cramer *et al.* (1996) *Nature Biotechnology*, 14(3), 315-9.
7. Cramer *et al.* (1996) *Nature Medicine*, 2(1), 100-2.
8. Giver *et al.* (1998) *Current Opinion in Chemical Biology*, 2(3), 335-8.
9. Giver *et al.* (1998) *Proc. Natl. Acad. Sci. USA*, 95(22), 12809-13.
10. Kumamaru *et al.* (1998) *Nature Biotechnology*, 16(7), 663-6.
11. Moore *et al.* (1997) *J. Mol. Biol.*, 272(3), 336-47.
12. Moore *et al.* (1996) *Nature Biotechnology*, 14(4), 45 8-67.
13. Yano *et al.* (1998) *Proc. Natl. Acad. Sci. USA*, 95(10), 5511-5.
14. Harayama, S. (1998) *Trends In Biotechnology*, 16(2), 76-82.
15. Cramer *et al.* (1998) *Nature*, 391(6664), 288-91.
16. Nixon *et al.* (1998) *Trends In Biotechnology*, 16(6), 25 8-64.
17. Kimura *et al.* (1997) *Journal of bacteriology*, 179(12), 3936-43.
18. Back, K. *et al.* (1996) *Proc. Natl. Acad. Sci. USA*, 93, 6841-5.
19. Campbell *et al.* (1997) *Nat Biotechnol*, 15(5), 43 9-43.
20. Nelson *et al.* (1987) In Guengerich, F. P. (ed.), *Mammalian cytochrome P-450*. CRC Press, Boca Raton and Florida.s, pp. 19-79.
21. Harris, C. C. (1989) *Carcinogenesis*, 10(9), 1563-6.
22. Kadlubar *et al.* In Guengerich, F. P. (ed.), *Mammalian cytochrome P-450*. CRC Press: Boca Raton and Florida.s., 1987, pp. 81-130.
23. Buters *et al.* (1999) *Drug Metab Rev*, 31(2), 437-47.
24. Kawajiri *et al.* (1990) *Princess Takamatsu Symposia*, 21, 55-61.
25. Kawajiri *et al.* (1990) *FEBS Letters*, 263(1), 131-3.
26. Kawajiri *et al.* (1993) *Critical Reviews in Oncology-Hematology*, 14, 77-87.
27. Mace *et al.* (1994) *Molecular Carcinogenesis*, 11(2), 65-73.
28. Joo *et al.* (1999) *Chemistry & Biology*, 6(10), 699-706.

ES 2 301 553 T3

29. **Shao et Arnold** (1996) *Current Opinion in Structural Biology*, 6(4), 513-8.
30. **Arnold**, F. H. (1998) *Nature Biotechnology*, 16(7), 617-8.
- 5 31. **Michnick**, S. W. et **Arnold**, F. H. (1999) *Nat Biotechnol*, 17(12), 1159-60.
32. **Kikuchi et al.** (1999) *Gene*, 236(1), 159-67.
33. **Kikuchi et al.** (2000) *Gene*, 243(1-2), 133-7.
- 10 34. **Ostermeier et al.** (1999) *Nat Biotechnol*, 17(12), 1205-9.
35. **Volkov et al.** (1999) *Nucleic Acids Res*, 27(18), e18.
- 15 36. **Okuta et al** (1998) *Gene*, 212(2), 221-8.
37. **Pompon**, D. et **Nicolas**, A. (1989) *Gene*, 83(1), 15-24.
38. **Mezard**, C., **Pompon**, D. et **Nicolas**, A. (1992) *Cell*, 70(4), 659-70.
- 20 39. **Cullin**, C. et **Pompon**, D. (1988) *Gene*, 65(2), 203-17.
40. **Truan et al** (1993) *Gene*, 125(1), 49-55.
- 25 41. **Pompon et al.** (1997) *J Hepatol*, 26 Suppl 2, 81-5.
42. **Urban et al.** (1990) *Biochimie*, 72(6-7), 463-72.
43. **Bellamine et al.** (1994) *Eur J Biochem*, 225(3), 1005-13.
- 30 44. **Sambrook et al.** (1989) *Molecular cloning: a laboratory manual*. 2nd Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
45. **Lorimer**, I. A. et **Pastan**, I. (1995) *Nucleic Acids Res*, 23(15), 3067-8.
- 35 46. **Zhao**, H. et **Arnold**, F. H. (1997) *Nucleic Acids Research*, 25(6), 1307-8.
47. **Pompon et al.** (1996) *Methods Enzymol*, 272, 51-64.
- 40 48. **Pompon**, D. (1988) *Eur J Biochem*, 177(2), 285-93.
49. **Smith et Waterman** (1981) *Ad. App. Math.* 2: 482
50. **Neddleman et Wunsch** (1970) *J Mol. Biol.* 48: 443
- 45 51. **Pearson et Lipman** (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444

REIVINDICACIONES

1. Procedimiento de construcción de un banco combinatorio de expresión funcional a partir de un banco combinatorio de ácidos nucleicos que pertenecen a una misma familia génica, que comprende las etapas que consisten en:
 - a. introducir dicho banco combinatorio de ácidos nucleicos en una levadura, simultáneamente con un vector de expresión
 - b. obtener dicho banco de expresión funcional por
 - recombinación homóloga de dicho banco combinatorio de ácidos nucleicos con dicho vector de expresión en dicha levadura, y
 - recombinación homóloga u homeóloga (entre las secuencias similares pero no idénticas), entre los diferentes ácidos nucleicos del banco combinatorio introducido en dicha levadura, con el fin de aumentar la complejidad y la diversidad del banco combinatorio de expresión funcional obtenido.en el que dicho banco combinatorio de ácidos nucleicos es una mezcla de productos de PCR obtenida por amplificación de un banco combinatorio de fases abiertas de lectura, presentando dichas fases abiertas de lecturas entre ellas una identidad de secuencia superior a 40%, utilizando una pareja de iniciadores situados en regiones que flanquean dichas fases abiertas de lectura, obteniéndose dicho banco combinatorio a partir de ADN homólogos o variantes de secuencia difiriendo en una o varias mutaciones y dicho banco combinatorio de fases abiertas de lectura se obtiene por reensamblaje con “extensión de primer” de productos de fragmentación de al menos dos fases abiertas de lecturas que codifican proteínas funcionales y dicha etapa de reensamblaje por “extensión de primer” se efectúa por PCR,**caracterizado** porque cada ciclo de dicha etapa de reensamblaje por PCR presenta al menos dos escalones de hibridación.
2. Procedimiento según la reivindicación 1, **caracterizado** porque dichas regiones que flanquean las fases abiertas de lectura son regiones promotoras y terminadoras que permiten la expresión en la levadura.
3. Procedimiento según una de las reivindicaciones 1 ó 2, **caracterizado** porque cada ciclo de dicha etapa de reensamblaje por PCR presenta al menos cuatro escalones de hibridación de más de 60 segundos, por temperaturas decrecientes regularmente espaciadas.
4. Procedimiento según una de las reivindicaciones 1 a 3, **caracterizado** porque dichos productos de fragmentación se obtienen a partir de un vector de expresión autónomo de levadura de tamaño total superior a 7 kb, incluyendo las fases abiertas de lectura.
5. Procedimiento según la reivindicación 4, **caracterizado** porque dicho vector de expresión que incluye las fases abiertas de lectura es un vector de expresión para una célula eucariota, y lanzadera para la levadura.
6. Procedimiento según una de las reivindicaciones 4 ó 5, **caracterizado** porque dicho vector de expresión que incluye las fases abiertas de lectura contiene igualmente los elementos necesarios para replicarse de manera autónoma en *Escherichia coli*.
7. Procedimiento según una de las reivindicaciones 4 a 6, **caracterizado** porque dicho vector de expresión que incluye las fases abiertas de lectura contiene una fase abierta de lectura que codifica una enzima membranaria eucariota.
8. Procedimiento según la reivindicación 7, **caracterizado** porque dicha enzima eucariota se elige entre el grupo constituido por los citocromos P450 eucariotas, enzimas de conjugación eucariotas (de fase II), miembros de la familia de transportadores ABC eucariotas.
9. Procedimiento según una de las reivindicaciones 1 a 8, **caracterizado** porque dicho vector de expresión con el que se efectúa la recombinación en la levadura se linealiza en el sitio de clonación normal del ADNc y posee secuencias promotoras y terminadoras de transcripción, efectuándose la recombinación a nivel de dichas secuencias.
10. Procedimiento según una de las reivindicaciones 1 a 9, **caracterizado** porque dicho vector de expresión posee igualmente la capacidad de replicarse de manera autónoma en células eucariotas y/o en *Escherichia coli*.
11. Procedimiento según una de las reivindicaciones 1 a 10, **caracterizado** porque la cepa de levadura utilizada presenta una modificación genética que permite la sobreexpresión de al menos proteína elegida entre el grupo constituido por una P450 reductasa endógena o exógena, una adrenodoxina, una adrenodoxina reductasa, un citocromo b5 heterólogo, una enzima de fase II (en particular una epóxido hidrolasa).

ES 2 301 553 T3

12. Procedimiento según una de las reivindicaciones 1 a 11, **caracterizado** porque la etapa b. está seguida de etapas que consisten en:

c. extraer el ADN plasmídico de al menos un clon de levadura,

d. transformar una cepa de *Escherichia coli* con dicho ADN plasmídico extraído, y seleccionar los clones transformados sobre un medio apropiado para obtener una discriminación de los elementos del banco combinatorio de expresión funcional.

13. Procedimiento de producción de proteínas funcionales mosaicas activas, **caracterizado** porque se construye un banco combinatorio de expresión funcional por un procedimiento según una de las reivindicaciones 1 a 12, se expresan proteínas mosaicas, y se seleccionan las proteínas funcionales mosaicas activas por el estudio de su actividad.

14. Procedimiento según la reivindicación 13, **caracterizado** porque dichas proteínas funcionales mosaicas activas derivan de enzimas.

15. Procedimiento según la reivindicación 14, **caracterizado** porque dichas proteínas funcionales mosaicas activas derivan de citocromos P450.

16. Procedimiento de análisis de un banco combinatorio de expresión funcional según una de las reivindicaciones 1 a 12, **caracterizado** porque comprende las etapas siguientes:

a. preparación de un banco combinatorio de expresión funcional por un procedimiento según una de las reivindicaciones 1-12,

b. transformación de una cepa de *Escherichia coli* con el ADN plasmídico extraído de la cepa de levadura o de un conjunto de levaduras,

c. hibridación del ADN plasmídico contenido en cada uno de los clones individuales de *Escherichia coli* obtenidos de la etapa a. con una o varias sondas específica(s) de una secuencia parental.

17. Procedimiento según la reivindicación 16, **caracterizado** porque dicha hibridación se efectúa sobre una macro- o una micro-red de ADN, estando constituida dicha red bien por ADN plasmídico contenido en cada uno de los clones individuales de *Escherichia coli* obtenidos de la etapa a., o por un producto de PCR de este, o bien por dichas sondas específicas, unida(s) a un soporte sólido, estando localizados cada uno de los ácidos nucleicos por su posición en la red.

18. Procedimiento de determinación de las relaciones entre firmas secuenciales y firmas funcionales de una proteína, **caracterizado** porque comprende las etapas que consisten en

a. preparar un banco combinatorio de expresión funcional por un procedimiento según una de las reivindicaciones 1 a 12,

b. producir las proteínas funcionales mosaicas activas por un procedimiento según una de las reivindicaciones 13 a 15,

c. analizar las diferencias funcionales y/o de actividad entre dichas proteínas mosaicas,

d. analizar los ácidos nucleicos que corresponden a dichas proteínas mosaicas por un procedimiento según la reivindicación 16 ó 17, seguido de manera opcional por un procedimiento de análisis de una huella de hibridación que comprenden las etapas que consisten en:

i. calcular la frecuencia de aparición de cada una de las combinaciones obtenidas en la etapa a,

ii. proceder a un análisis estadístico comparando la frecuencia de aparición calculada en la etapa i. con la frecuencia de aparición teórica para cada combinación, por un tratamiento matemático y estadístico adecuado

e. determinar las relaciones que pueden existir entre las estructuras secuenciales observadas en la etapa d. con las diferencias funcionales y/o de actividad en la etapa c.

19. Procedimiento de predicción de estructuras que tienen una función determinada, **caracterizado** porque se realiza el procedimiento según la reivindicación 18 para identificar las zonas de secuencias, o las relaciones entre las zonas de secuencias, unidas a dicha función, y de la que se deduce la estructura buscada.

20. Procedimiento de obtención de una proteína que posee propiedades mejoradas, **caracterizado** porque comprende las etapas que consisten en:

ES 2 301 553 T3

- a. construir un banco combinatorio de expresión funcional por un procedimiento según una de las reivindicaciones 1 a 12,
- 5 b. analizar dicho banco combinatorio de expresión funcional por un procedimiento según una de las reivindicaciones 16 ó 17,
- c. analizar las huellas de hibridación obtenidas en la etapa b. por un procedimiento de análisis de huella de hibridación que comprende las etapas que consisten en:
 - 10 i. calcular la frecuencia de aparición de cada una de las combinaciones posibles obtenidas en la etapa a,
 - ii. proceder a un análisis estadístico comparando la frecuencia de aparición calculada en la etapa c.i. con la frecuencia de aparición teórica para cada combinación, por un tratamiento matemático y estadístico adecuado
- 15 d. determinar las relaciones entre las estructuras secuenciales y las estructuras funcionales de las proteínas por comparación de dichas huellas de hibridación con las propiedades de las proteínas mosaicas correspondientes, por un procedimiento según la reivindicación 18
- 20 e. predecir las estructuras de interés o las organizaciones de estructuras en las proteínas mosaicas por un procedimiento según la reivindicación 19,
- f. repetir las etapas a. a e., al utilizar, como ácidos nucleicos de partida para la generación del banco combinatorio de expresión funcional, portando los ácidos nucleicos las estructuras de interés o las organizaciones de estructuras identificadas en la etapa e., un número suficiente de veces para obtener la proteína que posee las propiedades mejoradas buscadas.
- 25

21. Procedimiento de determinación de una estructura de una proteína importante en respuesta a una presión de selección a partir de un banco combinatorio de expresión funcional obtenido por un procedimiento según una de las reivindicaciones 1 a 12 y analizado según un procedimiento de una de las reivindicaciones 16 ó 17, para los elementos de las que se ha obtenido una firma, que comprende las etapas de:

- normalizar dicho banco asegurando que cada firma se encuentra con la misma probabilidad en el banco normalizado,
- 35 - aplicar una presión de selección,
- proceder a un análisis estadístico de las frecuencias de firmas de secuencia del nuevo banco así obtenido con respecto a las del banco normalizado de partida;
- 40 - estudiar el cambio de las frecuencias de aparición de las firmas de secuencias del nuevo banco obtenido, con respecto al banco de partida normalizado, deduciendo así las estructuras presentes o ausentes en respuesta a la presión de selección.

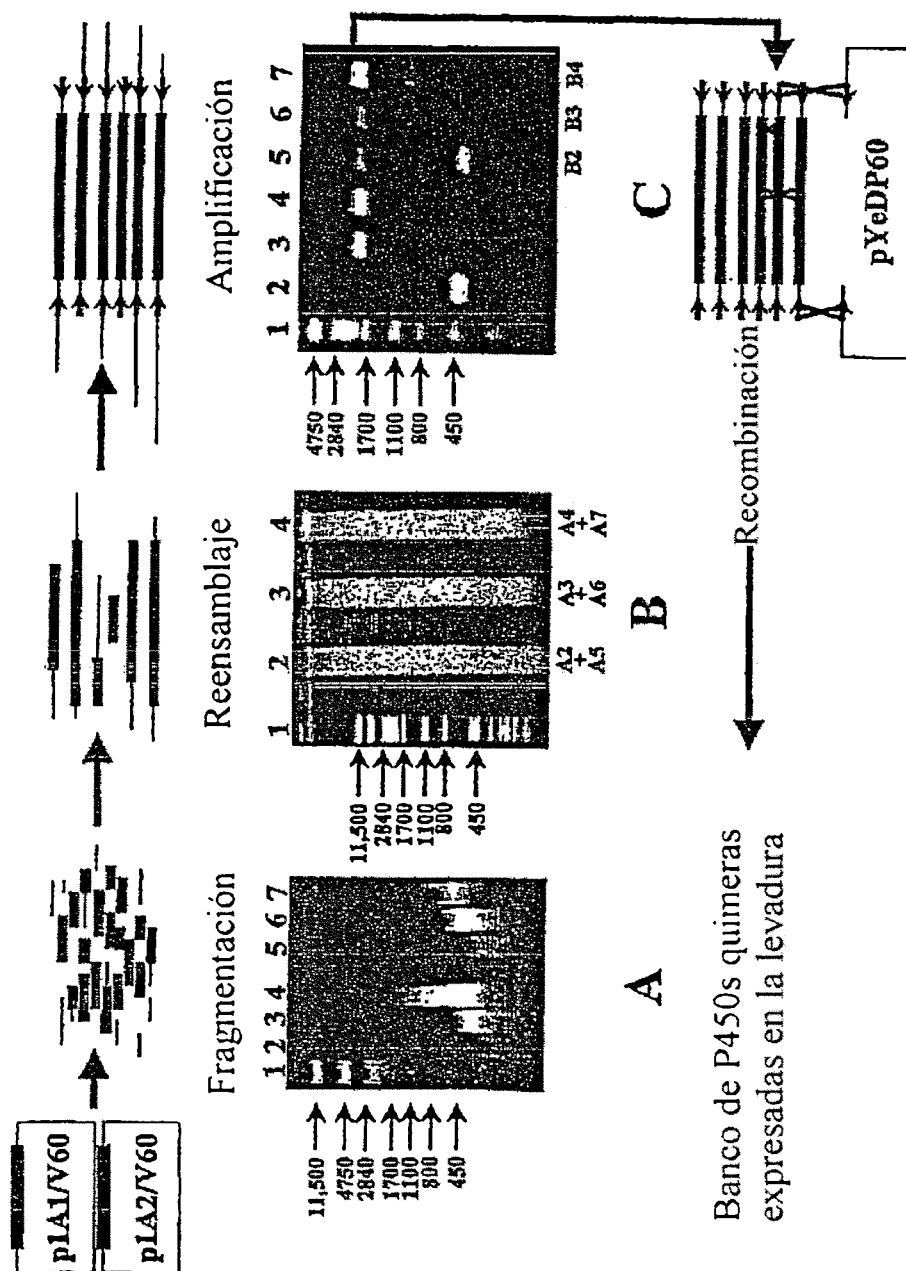


FIG. 1

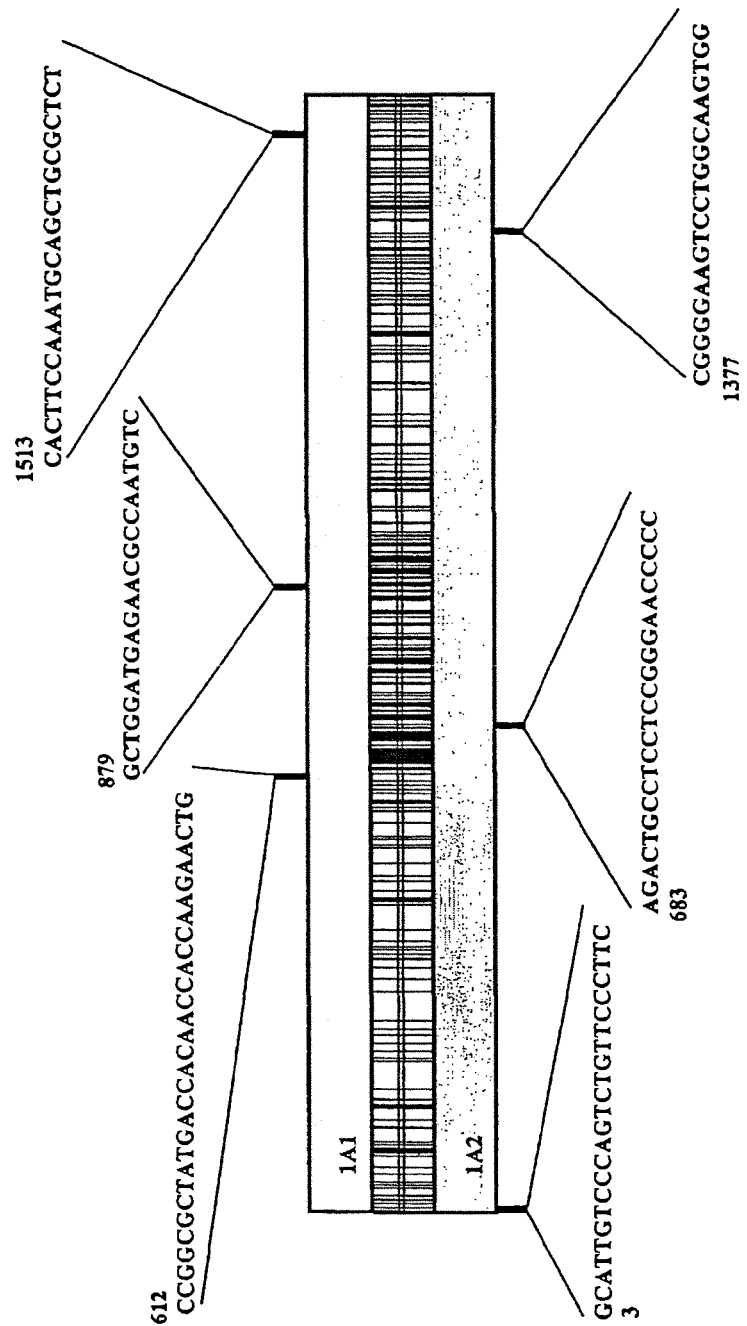


FIG. 2

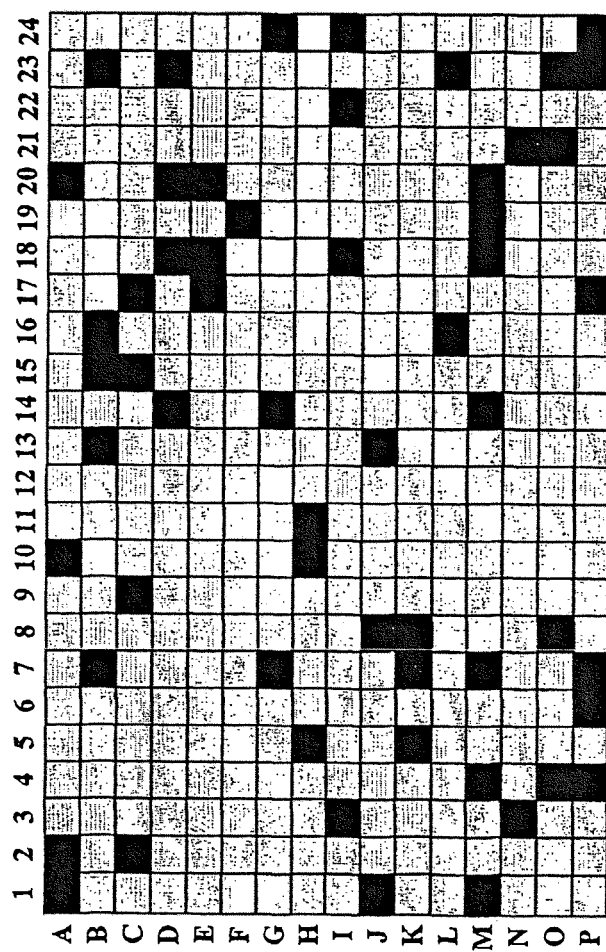


FIG. 3

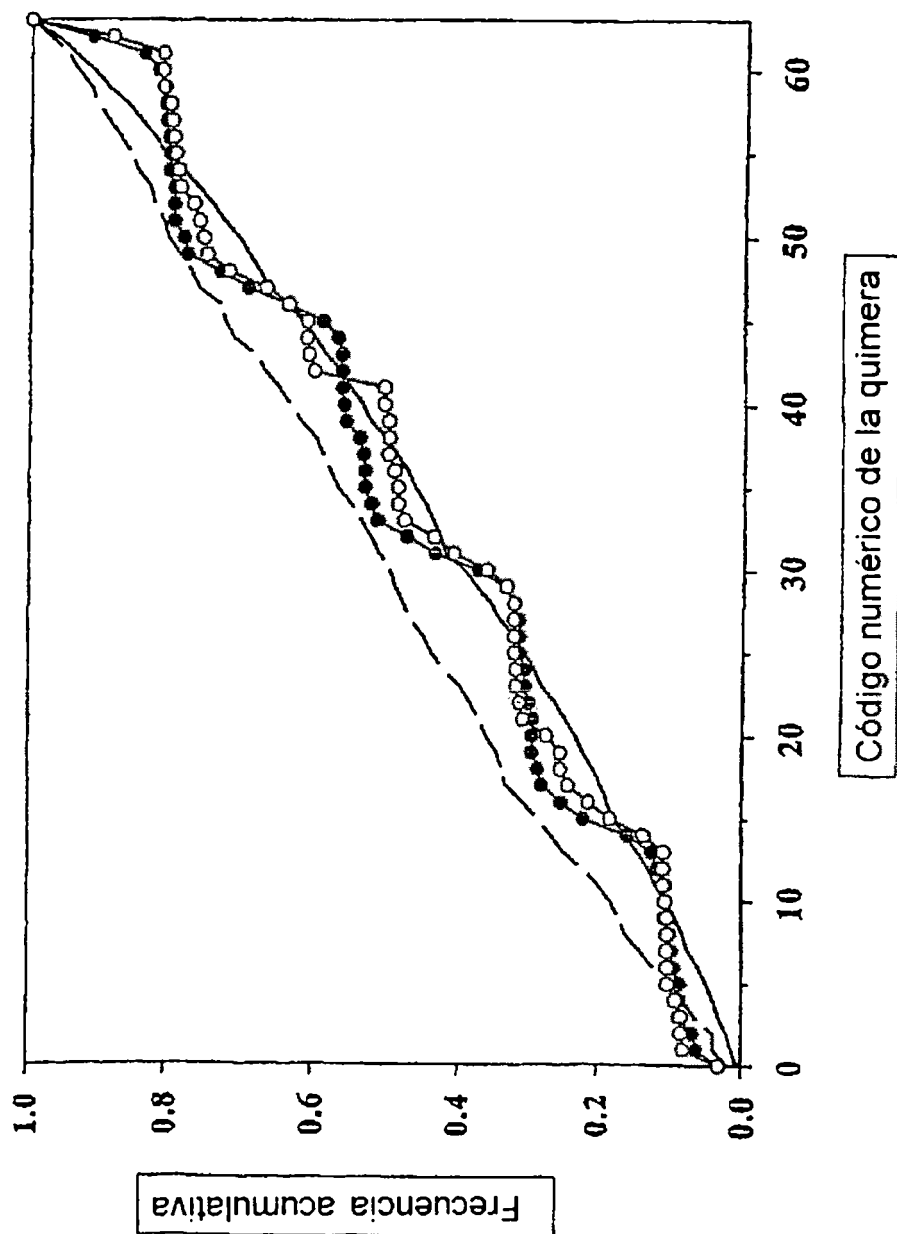


FIG. 4

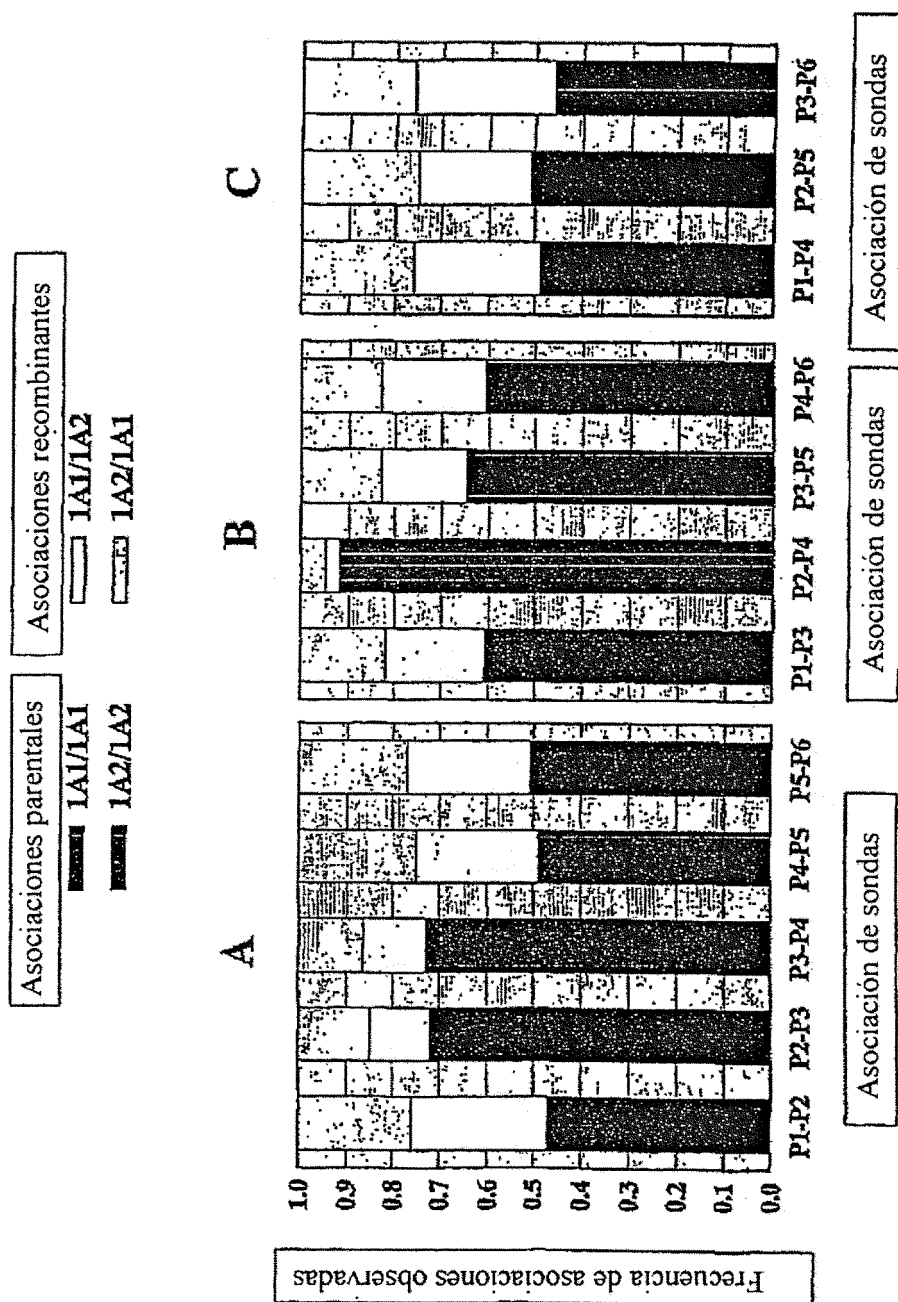


FIG. 5

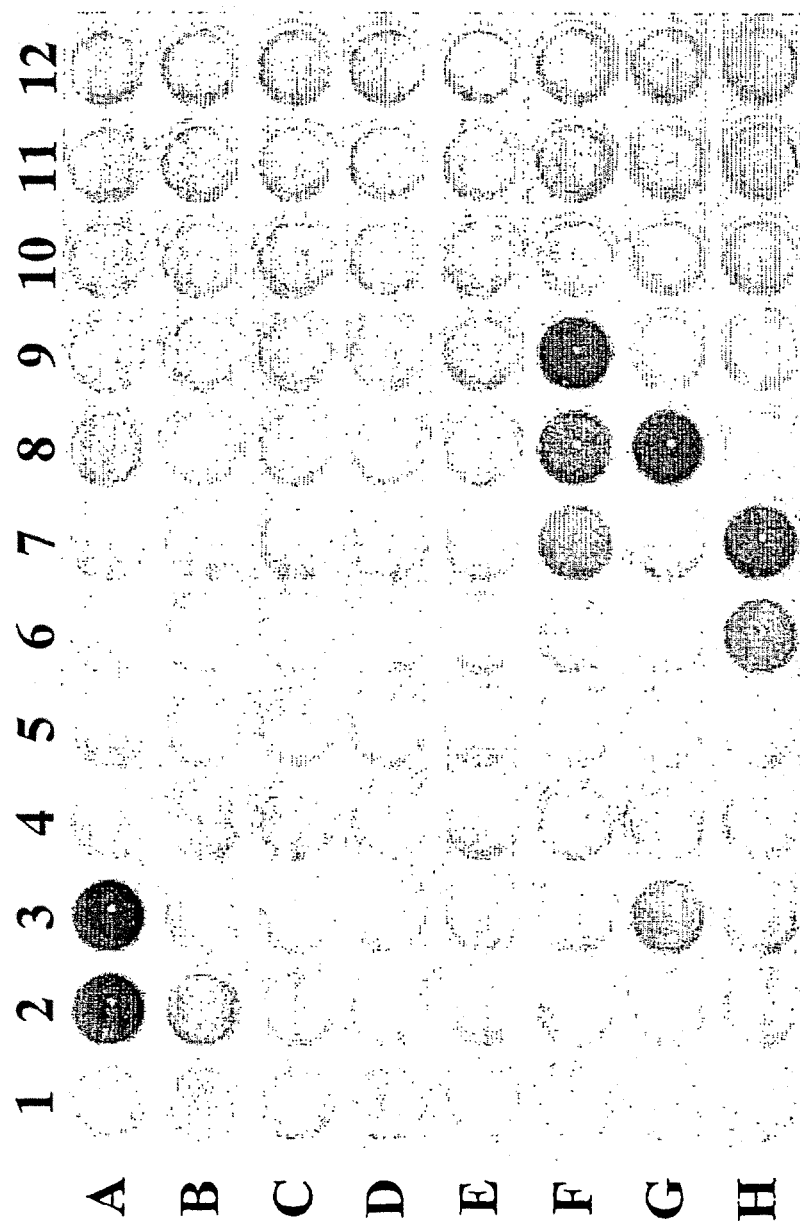


FIG. 6

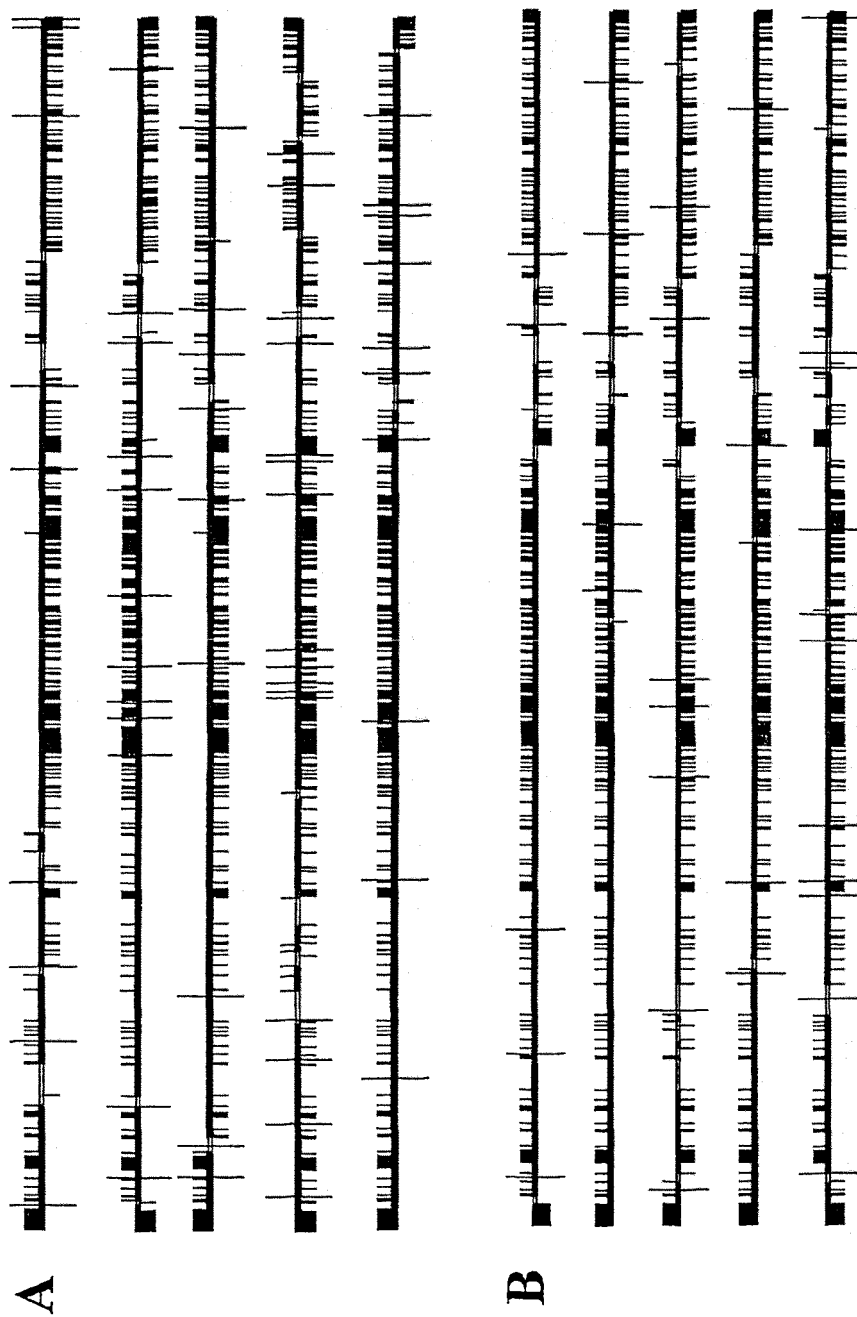


FIG. 7

ES 2 301 553 T3

LISTA DE SECUENCIAS

5	<110> Aventis Pharma SA Centre national de la Recherche Scientifique (CNRS).	
	<120> BANCOS COMBINATORIOS MEJORADOS POR RECOMBINACIÓN EN LA LEVADURA Y PROCEDIMIENTO DE ANÁLISIS	
10	<130> D18972	
	<150> FR 00 07555	
15	<151> 2000-06-14	
	<160> 8	
	<170> PatentIn Ver. 2.1	
20	<210> 1	
	<211> 24	
	<212> ADN	
25	<213> Levadura	
	<400> 1	
30	cggtgatata gcgtggatgg ccag	24
	<210> 2	
	<211> 16	
35	<212> ADN	
	<213> Levadura	
	<400> 2	
40	gcaccaccac cagtag	16
	<210> 3	
45	<211> 24	
	<212> ADN	
	<213> <i>Homo sapiens</i>	
50	<400> 3	
	gcattgtccc agtctgttcc ettc	24
55	<210> 4	
	<211> 31	
	<212> ADN	
60	<213> <i>Homo sapiens</i>	
	<400> 4	
65	ccggcgctat gaccacaacc accaagaact g	31
	<210> 5	

ES 2 301 553 T3

	<211> 24	
	<212> ADN	
	<213> <i>Homo sapiens</i>	
5	<400> 5	
	agactgcctc ctccgggaac cccc	24
10	<210> 6	
	<211> 22	
	<212> ADN	
15	<213> <i>Homo sapiens</i>	
	<400> 6	
20	gctggatgag aacccaatg tc	22
	<210> 7	
	<211> 21	
25	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<400> 7	
30	cggggaagtc ctggcaagtg g	21
	<210> 8	
35	<211> 24	
	<212> ADN	
	<213> <i>Homo sapiens</i>	
40	<400> 8	
	cacttcaaaa tgcagctgcg ctct	24
45		
50		
55		
60		
65		