



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

⑪ Número de publicación: **2 183 006**

⑤① Int. Cl.<sup>7</sup>: C12Q 1/68  
G06F 19/00

⑫

TRADUCCION DE PATENTE EUROPEA

T3

⑧⑥ Número de solicitud europea: **96928127.8**  
⑧⑥ Fecha de presentación: **12.08.1996**  
⑧⑦ Número de publicación de la solicitud: **0 862 649**  
⑧⑦ Fecha de publicación de la solicitud: **09.09.1998**

⑤④ Título: **Sistemas para generar y analizar matrices de señal de salida estímulo-respuesta.**

④⑤ Fecha de la publicación de la mención BOPI:  
**16.03.2003**

④⑤ Fecha de la publicación del folleto de patente:  
**16.03.2003**

⑦③ Titular/es:  
**The Regents of The University of California  
300 Lakeside Drive, 22nd Floor  
Oakland, California 94612-3350, US**

⑦② Inventor/es: **Rine, Jasper y  
Ashby, Matthew**

⑦④ Agente: **Carpintero López, Francisco**

**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).

## DESCRIPCION

Sistemas para generar y analizar matrices de señales de salida estímulo-respuesta.

### Ámbito de la invención

El ámbito de la invención es la generación y el análisis de perfiles de señales estímulo-respuesta adaptados a sistemas de inteligencia artificial basados en ordenadores, tales como las redes neurales y los sistemas expertos, y su uso como modelos para respuestas sistémicas.

### Antecedentes

Los sistemas de Inteligencia Artificial (AI) pueden integrar las funciones de acumulación, reconocimiento y almacenamiento de datos con protocolos de análisis y decisión de orden superior. Los sistemas de AI, tales como los sistemas expertos y las redes neurales, hallan una amplia aplicación en el análisis cuantitativo. Los sistemas expertos, típicamente, generan una estructura de datos individuales que se analiza según una base de conocimientos, trabajando en conjunto con una base de datos residente; véase, p. ej., Holloway *et al.* (1993) patente de EE.UU. N° 5.253.164, que fue sometida a una reciente revisión judicial; GMIS Inc., 34 USP Q2d 1389 (1995) "MYCIN", otro ejemplo, es un protocolo de ordenadores que utiliza evaluaciones clínicas individuales para generar una estructura de datos personales que se analiza según una base de conocimientos para predecir o diagnosticar el infarto de miocardio, y para determinar la admisión hospitalaria (Goldman *et al.* (1988) *New England Journal of Medicine* 318, 797-803).

Los sistemas de redes neurales son redes de elementos de procesamiento interconectados, cada uno de los cuales puede tener múltiples señales de entrada, pero genera sólo una señal de salida. Una red neural se entrena introduciendo un conjunto de señales de entrenamiento, y correlacionando las respuestas. La red entrenada se utiliza entonces para analizar señales nuevas. Por ejemplo, las redes neurales han sido ampliamente usadas en aplicaciones de reconocimiento de caracteres ópticos y de voz (p. ej., Colley *et al.* (1993) patente de EE.UU. N° 5.251.268).

Los análisis de sistemas complejos tales como los organismos biológicos están especialmente bien adaptados a los sistemas de AI. Complejos patrones de estímulo-respuesta, inmanejables de otra manera, pueden ser eficazmente analizados utilizando protocolos de deducción aplicados a través de sistemas de AI. El desarrollo farmacéutico, por ejemplo, requiere estudios a gran escala de respuestas sistémicas a las modificaciones de la estructura, la forma o la administración de un medicamento. Actualmente, tal información sistémica se suministra, habitualmente, por medio de modelos animales vivos que son costosos y que ofrecen señales de salida limitadas relativamente nada informativas (muerte, pérdida de peso, etc.), y que a menudo enmascaran las numerosas inhibiciones y activaciones de las rutas bioquímicas que subyacen en la respuesta medida de los organismos. Se han descrito múltiples procedimientos basados en cultivos celulares o *in vitro* para identificar los compuestos con un efecto biológico particular mediante la activación de un

informador unido (p. ej., Gadski *et al.* (1992) documento EP 92304902.7 describe procedimientos para sustancias que regulan la síntesis de una apolipoproteína; Evans *et al.* (1991), en la patente de EE.UU. N° 4.981.784, describe procedimientos para identificar el "ligando" (una molécula pequeña que se une a una molécula grande) para un receptor, y Farr *et al.* (1994) WO 94/17208 describe procedimientos y equipos que utilizan potenciadores del estrés para determinar la toxicidad de un compuesto.

La presente invención combina estos enfoques para ofrecer un análisis *in vitro*, o basado en cultivo celular, de patrones de respuesta sistémicos. En particular, la invención incluye procedimientos sofisticados para generar y analizar patrones de respuesta sumamente informativos de la inhibición y la activación estímulo-sistémicos.

### Resumen de la invención

La invención proporciona sistemas y procedimientos para generar una base de datos de matrices de señales de salida y para analizar una matriz de señales de salida por comparación con una base de datos de matrices de señales de salida, para correlacionar candidatos con estímulos y respuestas.

La generación de una base de datos de matrices de señales de salida según la invención incluye: (i) construir una matriz física estimulada; (ii) detectar una señal física en cada unidad de la matriz física; (iii) transducir cada señal física para generar una correspondiente señal de salida eléctrica; (iv) almacenar cada señal de salida en una estructura de datos matriciales de señales de salida, asociando cada señal de salida con las coordenadas X e Y de la correspondiente unidad matricial física y el estímulo; y (v) repetir las etapas (i) - (iv) para almacenar iterativamente las estructuras de datos matriciales de señales de salida para múltiples estímulos, a fin de formar una base de datos matriciales de señales de salida, indexando por estímulos las estructuras de datos matriciales de señales de salida.

El análisis de una matriz de señales de salida por comparación con una base de datos de matrices de señales de salida, según la presente invención, incluye: (a) construir una matriz física estimulada; (b) detectar una señal física en cada unidad de la matriz física; (c) transducir cada señal física para generar una correspondiente señal de salida eléctrica; (d) almacenar cada señal de salida en una estructura de datos matriciales de señales de salida, asociando cada señal de salida con las coordenadas X e Y de la unidad correspondiente de la matriz física y el estímulo; y (e) comparar la estructura de datos matriciales de señales de salida de la etapa (d) con una base de datos de matrices de señales de salida producida mediante el procedimiento precedente de generar una base de datos de matrices de señales de salida.

Las matrices físicas estimuladas comprenden una formación ordenada de unidades que tienen coordenadas X e Y. Cada unidad encierra (1) un respondedor distinto de un ser viviente o bien una sonda que corresponde a un tal respondedor distinto y (2) un identificador para el respondedor o sonda. Al ser viviente se le proporciona

un estímulo capaz de reprimir los respondedores de múltiples unidades, y el identificador proporciona una señal física correspondiente a la inhibición de tal respondedor distinto. Las formaciones pueden comprender la gama completa de respondedores del organismo, que pueden ser genes, elementos reguladores de genes, transcripciones de genes o traducciones de genes, o una clase o subconjunto funcional predeterminado de la gama completa del organismo. En una realización preferida, la formación comprende un conjunto suficiente de respondedores para deducir la acción de un estímulo, no importa cuál sea su mecanismo de acción.

### Breve descripción de las figuras

La Figura 1 es una vista general de un sistema de análisis matricial.

La Figura 2 es un diagrama de flujo que representa las etapas llevadas a cabo al generar una base de datos matriciales de señales de salida.

La Figura 3 es un diagrama que representa una matriz física estimulada.

La Figura 4 es un diagrama de flujo que representa las etapas llevadas a cabo al generar una estructura de datos matriciales de señales de salida y comparar dicha estructura de datos con una base de datos de matrices de señales de salida.

La Figura 5 es un esquema de un sistema experto de diseño, que puede utilizarse para analizar un perfil de señal de salida.

La Figura 6 es un diagrama de flujo que representa las etapas llevadas a cabo al generar un perfil de respuesta de salida de matriz informadora genética para un estímulo desconocido, las tablas de regulación, los perfiles de respuesta de referencia basal, los perfiles de respuesta química conocidos, y los perfiles de respuesta genética conocidos.

La Figura 7 es un diagrama de flujo que representa las etapas llevadas a cabo al analizar un perfil de respuesta de salida de matriz informadora genética para un estímulo desconocido.

### Descripción detallada de la invención

La presente invención se refiere a procedimientos para generar y analizar patrones biológicos de estímulo-respuesta utilizando protocolos de deducción aplicados a través de sistemas de AI, tales como los sistemas expertos y las redes neurales. Los sistemas pueden usarse, por ejemplo, como sustitutos de sistemas *in vitro* o de cultivo celular para estudios de animales vivos sobre la eficacia de fármacos. La Figura 1 muestra una visión general de un sistema según la invención. El sistema 100 incluye una unidad central de procesamiento 110, una memoria de ordenador 122, una interfaz de usuario 114, un bus de comunicaciones del sistema 112 y el subsistema 116 de transducción de señales de salida de la matriz física. El subsistema 116 incluye una matriz física estimulada 120

y un detector y transductor de señales de salida de la matriz física 118. La memoria del ordenador 122 contiene estructuras reunidas de datos matriciales de señales de salida en forma de una base de datos matriciales de señales de salida 124, una función de comparación 126 y, a menudo, una base de conocimientos 128.

La Figura 2 proporciona una representación esquemática de las etapas llevadas a cabo al generar una base de datos matriciales de señales de salida que indexa N estructuras de datos matriciales de señales de salida según los estímulos correspondientes. La primera etapa 205 incluye la asignación a N del valor entero 1 para la generación de la primera estructura de datos matriciales de señales de salida, correspondiente al primer estímulo.

La segunda etapa 210 del procedimiento mostrado en la Figura 2 incluye la construcción de una matriz física estimulada. La Figura 3 proporciona una representación esquemática de una matriz física estimulada. La matriz física estimulada 310 comprende una formación ordenada de unidades que tienen coordenadas X e Y. Mientras que la Figura 3 representa cuatro unidades ilustrativas 312, en la práctica, la matriz tendrá, típicamente, alrededor de cien o más unidades. Las unidades son generalmente una región de un sustrato sólido, tal como una porción bidimensional de la superficie de una oblea con base de silicio, una celda de una placa de microtitulación, etc. Cada unidad encierra un respondedor distinto 314 de un ser viviente, o bien una sonda 316 que corresponde a tal respondedor distinto y a un identificador 318 para el respondedor o sonda. Generalmente, todas las unidades de una matriz dada emplearán un respondedor, o todas tendrán una sonda. Además, para una detección y procesamiento de datos más convenientes, todas las unidades de una matriz dada, generalmente, usan el mismo identificador.

Al ser viviente (u organismo) se le proporciona un estímulo capaz de inhibir los respondedores 314 de múltiples unidades 312, y el identificador 318 proporciona una señal física que corresponde a la inhibición de tal respondedor distinto 314. Las respuestas, habitualmente celulares, ante una amplia variedad de estímulos, pueden controlarse. Los ejemplos de estímulos incluyen agentes farmacológicos candidatos, agentes patógenos sospechosos, ácidos nucleicos transfectados, energía radiante, etc. El estímulo induce la inhibición de múltiples respondedores de la matriz, con respecto a su estado de inducción previo al estímulo, según se mide por la señal de salida previa al estímulo en la unidad correspondiente. Típicamente, el estímulo proporciona un complejo patrón de respuesta de inhibición, silencio e inducción sobre la matriz. El perfil de respuesta refleja los ajustes transcripcionales de las células para mantener la homeostasis en presencia del fármaco. Por lo tanto, mientras que una amplia variedad de estímulos pueden ser evaluados, es importante ajustar las condiciones de incubación (p. ej., la intensidad del estímulo, el tiempo de exposición, etc.) para impedir el estrés celular, y, por tanto, asegurar las mediciones de perfiles de respuesta farmacéuticamente relevan-

tes. Las formaciones pueden comprender la gama completa de respondedores del organismo, que pueden ser genes, elementos reguladores de genes, transcripciones de genes o traducciones de genes (proteínas), o una clase funcional o subconjunto predeterminados de la gama completa del organismo. Al incorporar al menos el 0,5%, preferiblemente al menos el 5%, más preferiblemente al menos la mitad, muy preferiblemente, esencialmente todos los respondedores (p. ej., regiones reguladoras de genes) del organismo, puede obtenerse un modelo *in vitro* o de cultivo celular del organismo (p. ej., un animal). En una realización preferida, la formación comprende un conjunto suficiente de respondedores a fin de modelizar la respuesta sistémica del organismo, y de deducir la acción de un estímulo, no importa cuál sea su mecanismo de acción.

La naturaleza del enlace 320 entre el respondedor y el identificador variará con la aplicación de la matriz. Como ejemplos: cada unidad de una matriz que informa de la expresión genética podría encerrar una célula que tiene una estructura de un gen informador operativamente unido a un promotor transcripcional distinto; como alternativa, cada unidad de una matriz que informa de la expresión genética podría encerrar una sonda de oligonucleótidos distinta, capaz de hibridizar con un correspondiente transcripto distinto de informador; cada unidad de una matriz que informa de la interacción proteína-ADN podría encerrar una célula que tiene una primera estructura de un gen informador, unido operativamente con una posición ligada al factor de transcripción dirigido, y una segunda estructura híbrida que codifica un dominio de activación de la transcripción fusionado con un gen estructural distinto (una matriz de sistema unidimensional monohíbrido); cada unidad de una matriz que informa de interacciones proteína-proteína podría encerrar una célula que tiene una primera estructura de un gen informador operativamente unido a una posición ligada al factor de transcripción diana, una segunda estructura híbrida que codifica un dominio de activación de la transcripción fusionado con un gen de expresión constitutiva distinta, y una tercera estructura que codifica un dominio ligado al ADN fusionado con otro gen de expresión constitutiva distinta (una matriz de sistema bidimensional bihíbrido).

La tercera etapa 212 del procedimiento mostrado en la Figura 2 incluye la detección de una señal física en cada unidad de la matriz física. La señal física dependerá de la naturaleza del respondedor y/o identificador. Típicamente, la señal es un cambio en una o más propiedades electromagnéticas, en particular, propiedades ópticas en la unidad. Como ejemplos, un gen informador puede codificar una enzima que cataliza una reacción en la unidad que altera las propiedades de absorción de la luz en la unidad, los nucleótidos marcados con marcajes fluorescentes o radioactivos pueden incorporarse a los transcritos nacientes, que se identifican entonces cuando se ligan a sondas de oligonucleótidos, etc. Los detectores electrónicos para señales ópticas, radiactivas, etc., están comercialmente disponibles. Por ejemplo, la Figura 1 muestra un detector colorimétrico

multicelular automatizado 118, similar a los lectores automatizados ELISA, que leen una matriz física de tipo placa de microtitulación residente 120. Generalmente, el procedimiento incluye la detección de una señal del informador en cada unidad de la matriz en respuesta al estímulo o estímulos candidatos en una primera intensidad, y nuevamente en una segunda intensidad. Frecuentemente, una de las intensidades será cero, o al menos estará por debajo del umbral necesario para un efecto en cualquiera de las células. Por ejemplo, una señal puede detectarse antes y después de que las células de la matriz sean incubadas con un agente farmacológico candidato durante un tiempo y en condiciones suficientes para que las células respondan a la presencia del agente. La señal también puede controlarse en función de otras variables tales como la intensidad o la duración del estímulo, el tiempo (para análisis dinámicos de respuesta), etc.

La cuarta etapa 214 del procedimiento mostrado en la Figura 2 incluye la transducción de la señal física en cada unidad para generar una correspondiente señal de salida eléctrica. Generalmente, la transducción de señal es una conversión electrónica lineal en una señal digital. Los convertidores electrónicos capaces de transducción lineal de señales están comercialmente disponibles. El detector y transductor de señales de salida de matriz física 118 mostrado en la Figura 1 aloja esta función transductora.

La quinta etapa 216 del procedimiento mostrado en la Figura 2 incluye el almacenamiento en memoria de cada señal de salida en una estructura de datos matriciales de señales de salida, que asocia cada señal de salida con las coordenadas X e Y de la unidad de matriz física correspondiente y el estímulo. Según se muestra en la Figura 1, la estructura de datos puede almacenarse en forma de una base de datos matriciales de señales de salida 124 en la memoria 122 de un ordenador 110.

Después de la quinta etapa del procedimiento procedimiento en la Figura 2, la rutina avanza hasta un bloque de decisión 218. Si no se han completado todas las estructuras de datos requeridas, la rutina avanza hasta la etapa 220, donde se incrementa n y se repiten las etapas 210-216 de generación de estructuras de datos para la n+1-ésima matriz física estimulada. Si, después de la quinta etapa del procedimiento mostrado en la Figura 2, se han completado todas las estructuras de datos requeridas, la rutina avanza hasta la etapa 222. La etapa 222 incluye la formación de una base de datos matriciales de señales de salida que indexa estructuras de datos matriciales de señales de salida según estímulos. Esta etapa de compilación generalmente incluye la digitalización de una señal analógica correspondiente a la señal informadora en cada unidad, el apareo de cada señal digitalizada con un identificador de unidad matricial y el almacenamiento de dichos pares como una base de datos de señales de salida en la memoria de un ordenador.

La Figura 4 proporciona una representación esquemática de las etapas llevadas a cabo al generar una estructura de datos matriciales de señales de salida y comparar dicha estructura de datos con una base de datos matriciales de señales de

salida, indexando N estructuras de datos matriciales de señales de salida según los estímulos correspondientes.

Las primeras cuatro etapas 410, 412, 414 y 416 son según se describe para las etapas 210, 212, 214 y 216 de la Figura 2. En particular, la etapa 410 del procedimiento incluye la construcción de una matriz física estimulada; la segunda etapa 412 incluye la detección de una señal física en cada unidad de la matriz física; la tercera etapa 414 incluye la transducción de cada señal física para generar una correspondiente señal de salida eléctrica; y la cuarta etapa 416 del procedimiento incluye el almacenamiento en memoria de cada señal de salida en una estructura de datos matriciales de señales de salida. La quinta etapa 418 del procedimiento incluye la comparación de la estructura de datos matriciales de señales de salida de la etapa (d) con una base de datos de matrices de señales de salida producida por el procedimiento precedente de generar una base de datos de matrices de señales de salida. La etapa de comparación, generalmente, se lleva a cabo por medio de un sistema informático que emplea tecnología de AI. Por ejemplo, el sistema 100 de la Figura 1 muestra una unidad central de procesamiento y una interfaz de usuario 114 que trabajan conjuntamente con una memoria 122 que almacena la base de datos de matrices de señales de salida 124 y la función de comparación 126. Aquí, la función de comparación 126 proporciona la tecnología de AI para comparar, por ejemplo, una matriz de señales de salida desconocidas con una base de datos 124, para deducir el mecanismo de acción y las características del estímulo responsable. Por ejemplo, puede utilizarse un sistema experto que utiliza una base de conocimientos de entrada, de reglas de comparación, o una red neural adiestrada sobre una población de pares conocidos estímulo-respuesta.

En particular, la Figura 5 proporciona una representación esquemática de un sistema experto 500 que puede usarse para comparar una estructura de datos matriciales de señales de salida de la etapa con una base de datos de matrices de señales de salida. Tal sistema puede implementarse en hardware o en software. La base de conocimientos 510 comprende una base de datos de matrices de señales de salida y una serie de reglas de comparación. La interfaz del sistema 514 permite la entrada de estructuras de datos de biblioteca para la base de datos, y consultar estructuras de datos o estímulos. El motor de inferencia 512 es un programa de ordenador que procesa estructuras de datos para su comparación con la base de datos de conocimientos residente, según las reglas de la base de conocimientos para generar correlaciones y análisis de deducción cualitativos y/o cuantitativos. Tales análisis son convenientemente producidos como un conjunto priorizado de coincidencias, incluyendo cada coincidencia un identificador y un puntaje de afinidad, como en los informes de búsqueda del Transformador de Cadenas de Alineación Local Básica (Basic Local Alignment String Transformer - BLAST), Altschul *et al.* (1990) Basic Alignment Search Tool (Herramienta Básica de Búsqueda de Alineación), *J. Mol. Biol.* 215, 403-410.

Una realización particular de la invención es una matriz de informadores genéticos, de alcance sustancialmente amplio, o matriz informadora del genoma. Tal matriz, de alcance sustancialmente amplio, incluye al menos una mayoría de los genes del organismo y, en una realización preferida, todos los genes distintos, esencialmente, del organismo de destino. Debido a que la levadura, tal como la *Saccharomyces cerevisiae*, es un eucariota fiable, hay una conservación sustancial de la función bioquímica entre la levadura y las células humanas en la mayoría de las rutas, desde la ruta biosintética del esterol hasta el oncogén Ras. En efecto, la ausencia de muchos compuestos antifúngicos eficaces ilustra lo difícil que ha sido hallar objetivos terapéuticos que mataran, preferentemente, células de hongos, pero no células humanas.

Un ejemplo de una ruta de respuesta compartida es la biosíntesis de esterol. En las células humanas, el fármaco Mevacor (lovastatina) inhibe la HMG-CoA reductasa, la enzima reguladora clave de la ruta biosintética del esterol. Como resultado, el nivel de un esterol regulador particular disminuye, y las células responden con la transcripción aumentada del gen que codifica el receptor de LDL. En la levadura, Mevacor también inhibe la HMG-CoA reductasa y disminuye el nivel de un esterol regulador clave. Las células de levadura responden de manera análoga a la de las células humanas. Sin embargo, la levadura no tiene un gen para el receptor de LDL. En cambio, el mismo efecto se mide mediante la transcripción aumentada del gen ERG10, que codifica la acetacetil CoA tiolasa, una enzima también implicada en la síntesis del esterol. De esta manera, la respuesta reguladora se conserva entre la levadura y los humanos, aún cuando la identidad del gen que responde es distinta.

La matriz informadora del genoma se ilustra a continuación con un sistema informador de productos genéticos basado en células; sin embargo, pueden utilizarse los demás tipos de matrices físicas estimuladas descritos en la presente memoria. Una matriz informadora del genoma, que comprende una amplia colección de genes informadores para los alrededor de 6.000 genes de la levadura, emite señales de respuesta reguladora de un único gen a esencialmente cualquier cosa que influya sobre la función de la célula. Por ejemplo, una producción aumentada del informador ERG10 emitiría una señal de presencia de un inhibidor de la síntesis de esterol. En la práctica, una matriz informadora del genoma de la levadura puede hacerse con una colección de varios miles de estirpes de levadura, cada una de las cuales contiene una única fusión genética de un gen de levadura con un gen informador. En una realización, las fusiones informadoras serán genes híbridos en los cuales el gen lacZ está fusionado con un gen de levadura, por lo que la actividad del promotor del gen de levadura dirige la síntesis de la  $\beta$ -galactosidasa. Estas estirpes de fusión que contienen genes se disponen convencionalmente en placas de microtitulación en cultivo líquido, y una colección permanente se mantiene a  $-80^{\circ}\text{C}$ . Pueden hacerse copias de esta colección, y propagarse mediante mecánicas simples, y auto-

matizarse con robótica comercial. La adición de un inhibidor a la colección completa de estirpes da como resultado ningún cambio en la expresión del informador en algunas estirpes, una expresión aumentada en algunas estirpes y una expresión disminuida en otras. Conociendo la identidad del gen informador en cada linaje, se interpreta la respuesta a nivel de genoma.

Los puntos fuertes de este procedimiento se ilustran mejor con ejemplos. Se considera la diferencia entre un ensayo *in vitro* para inhibidores de la HMG-CoA reductasa, según se practica actualmente en la industria farmacéutica, y un ensayo para inhibidores de la biosíntesis de esteroles, según describe el informador ERG10. En el caso de aquél, la información se obtiene sólo para aquellos raros compuestos que inhiben esta enzima. Por el contrario, en el caso del informador ERG10, cualquier compuesto que inhiba casi cualquiera de las aproximadamente 35 etapas de la ruta biosintética del esteroles inducirá la síntesis del informador. De esta manera, el informador puede detectar una gama mucho más amplia de objetivos que pueden ser la enzima purificada; en este caso, 35 veces más que el ensayo *in vitro*.

Los fármacos a menudo tienen efectos secundarios que se deben, en parte, a la falta de especificidad del objetivo. Sin embargo, el ensayo *in vitro* de la HMG-CoA reductasa no proporciona ninguna información sobre la especificidad de un compuesto. Al contrario, una matriz informadora del genoma revela el espectro de otros genes en el genoma, también afectados por el compuesto. Al considerar dos compuestos distintos, cada uno de los cuales induce el informador ERG10, si un compuesto afecta la expresión de otros 5 informadores y un segundo compuesto afecta la expresión de otros 50 informadores, es más probable que el primer compuesto, *a priori*, tenga menos efectos secundarios. Debido a que las identidades de los informadores son conocidas o determinables, la información sobre otros informadores afectados también informa en cuanto a la naturaleza del efecto secundario. Puede utilizarse un panel de informadores para ensayar derivados del compuesto principal, a fin de determinar cuál de los derivados tiene mayor especificidad que el primer compuesto.

Como otro ejemplo, se considera el caso de un compuesto que no afecta al ensayo *in vitro* para la HMG-CoA reductasa, ni induce la expresión del informador ERG10. En el enfoque tradicional del descubrimiento de fármacos, un compuesto que no inhibe el objetivo que se está ensayando no ofrece ninguna información útil. Sin embargo, un compuesto que tiene cualquier efecto significativo sobre un proceso biológico generalmente tiene alguna consecuencia sobre la expresión del gen. Una matriz informadora del genoma puede, por tanto, proporcionar dos clases distintas de información para la mayoría de los compuestos. En algunos casos, la identidad de los genes informadores afectados por el inhibidor da pruebas de cómo funciona el inhibidor. Por ejemplo, un compuesto que induce un promotor dependiente de AMPc en levadura puede afectar a la actividad de la ruta Ras. Incluso cuando el compuesto afecta a la expresión de un conjunto de genes que no muestran

evidencias de la acción del compuesto, la matriz proporciona una evaluación de amplio alcance de la acción del compuesto que puede almacenarse en una base de datos para análisis posteriores. Una biblioteca de tales perfiles de respuesta puede investigarse continuamente, igual que los "Compendios Espectrales" de química son continuamente citados en las ciencias químicas. Por ejemplo, si la base de datos revela que el compuesto X altera la expresión del gen Y, y se publica un artículo que informa de que la expresión del gen Y es sensible a, por ejemplo, la ruta de señalización del fosfato de inositol, el compuesto X es un candidato para modular la ruta de señalización del fosfato de inositol. En efecto, la matriz informadora del genoma es un traductor de información, que lleva información sobre un gen directamente a un compuesto del que es posible que ya se haya descubierto que afecta a la expresión de ese gen. Esta herramienta debería acortar radicalmente la fase de investigación y descubrimiento del desarrollo de fármacos, e influir eficazmente sobre el valor de los documentos de investigación públicamente disponibles sobre todos los genes. Al contrario que los enfoques existentes, que requieren el desarrollo continuo de distintos ensayos, la matriz informadora del genoma mide el efecto de todos los inhibidores con el mismo ensayo.

Una matriz informadora del genoma puede consistir en tres clases de informadores: cien fusiones de genes informadores construidos con tecnología PCR para ese subconjunto de genes que se conocen, y cuya expresión es sumamente reveladora de los cambios fisiológicos en toda la célula; una biblioteca dispuesta aleatoriamente lo suficientemente compleja para incluir una fusión para la mayoría de los genes de levadura; e informadores para los genes clave en una colección de estirpes genéticamente sensibilizadas que contienen mutaciones que sensibilizan esa estirpe ante los efectos sobre una cierta ruta. Por ejemplo, la primera clase de informadores se diseña como perfectas fusiones traductoras de lacZ, en las cuales la secuencia de codificación de la  $\beta$ -galactosidasa se fusiona con el codón de iniciación del gen de interés. Tales fusiones son capaces de detectar cambios en el nivel de expresión, debidos ya sea a los cambios en la transcripción o a los cambios en la traducción. Los genes detectados mediante este procedimiento incluyen el gen HO, para controlar efectos sobre la función del ciclo celular; el gen FUS 1, para detectar cambios en una de las cascadas de MAP quinasa; ERG10 y HMG-CoA reductasa para informar de cambios en la ruta biosintética del esteroles; GCN4, para informar de niveles de aminoacil-ARNt cargados; Ty1, para informar de los niveles de expresión de retroposón; expresión de HIS3 para informar de efectos sobre el oncogén homólogo de JUN de la levadura; HMG2 para informar de la inhibición de la biosíntesis de hemo; genes que informan de la actividad de la ruta RAS, otras cascadas de MAP quinasa, el nivel de actividad de fosfolípidos, etc.

En muchos casos, un fármaco de interés operaría sobre objetivos de proteínas cuyo impacto sobre la expresión del gen no se conocería *a priori*. Por ejemplo, se piensa que el taxol, un reciente avance en las terapias de cáncer de mama po-

tencial, funciona interfiriendo con elementos citoesqueléticos basados en tubulina. Incluso sin conocimientos de qué genes en la matriz informadora del genoma son inducidos o inhibidos por los inhibidores del citoesqueleto microtubular basado en tubulina, la matriz informadora del genoma puede utilizarse para determinarlo. Específicamente, una forma mutante dominante de la tubulina se introduce en todos las estirpes de la matriz informadora del genoma y el efecto del mutante dominante, que interfiere con el citoesqueleto microtubular, se evalúa para cada informador. Este ensayo genético nos informa de qué genes serían afectados por un fármaco que tuviera un mecanismo de acción similar. En el caso del taxol, el mismo fármaco podría utilizarse para obtener la misma información. Sin embargo, el ejemplo demuestra que, incluso si el mismo taxol no estuviera disponible, puede usarse la genética para predeterminar cuál sería su perfil de respuesta en la matriz informadora del genoma. Además, no es necesario saber la identidad de ninguno de los genes que responden. En cambio, el control genético con la tubulina mutante clasifica el genoma entre aquellos genes que responden y aquellos que no lo hacen. Por tanto, si se desearan fármacos que perturban el citoesqueleto de actina, los mutantes de actina dominantes introducidos en la matriz informadora del genoma revelan qué perfil de respuesta cabe esperar para tal agente.

Entre los avances más importantes en el desarrollo de fármacos, ha habido avances en la síntesis combinatoria de las bibliotecas químicas. En la verificación convencional de fármacos con objetivos de enzimas purificadas, las químicas combinatorias pueden ayudar frecuentemente a crear nuevos derivados de un compuesto principal que también inhibirán la enzima objetivo, pero con alguna propiedad distinta y deseable. Sin embargo, los procedimientos convencionales fracasarían en reconocer una molécula que tenga una especificidad sustancialmente divergente. La matriz informadora del genoma ofrece una solución sencilla para reconocer nuevas especificidades en las bibliotecas combinatorias. Específicamente, se prueban colecciones de nuevos compuestos como mezclas por toda la matriz. Si la colección tiene alguna nueva actividad no presente en el compuesto principal original, los nuevos genes son afectados entre los informadores. La identidad de ese gen ofrece una guía hacia el objetivo del nuevo compuesto. Además, la matriz ofrece un añadido extra que compensa una debilidad común en la mayoría de las síntesis químicas. Específicamente, la mayoría de las síntesis producen el producto deseado en suma abundancia, y una colección de otros productos asociados como contaminantes, debido a las reacciones secundarias en la síntesis. Tradicionalmente, la solución para los contaminantes es purificar separándolos. Sin embargo, la matriz informadora del genoma explota la presencia de estos contaminantes. Las síntesis pueden ajustarse para hacerlas menos específicas, con un mayor número de reacciones secundarias y más contaminantes, para determinar si algo en la síntesis total afecta a la expresión de los genes objetivo que interesan. Si hay un componente de la mezcla con la actividad de-

seada sobre un informador particular, ese informador puede usarse para ensayar la purificación del componente deseado de la mezcla. En efecto, la matriz informadora permite una investigación focalizada del efecto sobre genes individuales para compensar la impureza de la mezcla que se está ensayando.

Los isoprenoides son una clase especialmente atractiva para la matriz informadora del genoma. En la naturaleza, los isoprenoides son las campeonas entre las moléculas señalizadoras. Los isoprenoides son derivados del compuesto isopreno de cinco carbonos, que se forma como un compuesto intermedio en la biosíntesis del esterol. Los isoprenoides incluyen muchas de las más famosas fragancias, pigmentos, y otros compuestos biológicamente activos, tales como los sesquiterpenoides antifúngicos. Hay aproximadamente 10.000 derivados caracterizados del isopreno y muchos más potenciales. Debido a que estos compuestos se utilizan en la naturaleza para indicar procesos biológicos, es probable que incluyan algunas de las mejores moléculas permeadoras de membranas.

Los isoprenos poseen otra característica que se presta bien al descubrimiento de fármacos a través de la matriz informadora del genoma. Los compuestos isoprenoides puros pueden ser tratados químicamente para crear una amplia mezcla de distintos compuestos de manera rápida y fácil, debido a la disposición particular de los dobles enlaces en las cadenas de hidrocarburos. En efecto, los isoprenoides pueden ser mutagenizados desde una forma a muchas formas distintas, igual que un gen de tipo silvestre puede ser mutagenizado a muchos mutantes distintos. Por ejemplo, la vitamina D usada para fortalecer la leche se produce por irradiación ultravioleta del derivado del isopreno conocido como ergosterol. Los nuevos isoprenoides biológicamente activos se generan y analizan con una matriz informadora del genoma de la siguiente manera. En primer lugar, se ensaya un isoprenoide puro como el limoneno para determinar su perfil de respuesta en toda la matriz. Luego, el isoprenoide (p. ej., el limoneno) se altera químicamente para crear una mezcla de distintos compuestos. Esta mezcla se ensaya luego por toda la matriz. Si se observa alguna nueva respuesta, entonces la mezcla tiene nuevas especies biológicamente activas. Además, la identidad de los genes informadores ofrece información con respecto a lo que hace la nueva especie activa, una actividad a utilizar para controlar su purificación, etc. Este enfoque puede practicarse en otras familias químicas mutables distintas, además de los isoprenoides.

Los hongos son patógenos importantes en las plantas y animales, y tienen un impacto importante sobre la producción de muchos cultivos alimenticios y sobre la salud animal, incluyendo la humana. En el desarrollo de compuestos antifúngicos, uno de los objetivos principales ha sido determinar qué objetivos son específicos del hongo infeccioso. La matriz informadora del genoma ofrece una nueva herramienta para resolver este problema. Específicamente, se crea una biblioteca informadora a partir del patógeno diana, como *Cryptococcus*, *Candida*, *Aspergillus*, *Pneu-*

*mocystis*, etc. No es necesario conocer la identidad de cada gen informador, ni que se secuencie el genoma de la especie diana. Los compuestos se ensayan para determinar aquellos que tienen un efecto sobre la expresión de cualquier gen en el hongo. Estos compuestos positivos se ensayan luego para determinar aquellos que tienen muy poco, o ningún efecto sobre *Saccharomyces*. Secuenciando los genes informadores afectados específicamente en el hongo de destino, y comparando la secuencia con otras en Genbank, se pueden identificar rutas bioquímicas que son únicas para la especie diana. Los productos identificados útiles incluyen no sólo los agentes que matan el hongo diana, sino también la identificación de objetivos específicos en el hongo para otros ensayos de revisión farmacéutica.

La identificación de compuestos que matan bacterias ha sido exitosamente perseguida por la industria farmacéutica durante décadas. Es bastante sencillo identificar un compuesto que mata bacterias en un ensayo puntual sobre una placa Petri. Sin embargo, hay mucha complejidad en la fisiología y ecología bacterianas, que podría oponer resistencia al desarrollo de terapias de combinación para las bacterias, incluso para compuestos que no matan eficazmente la célula bacteriana. Se consideran, por ejemplo, las bacterias que invaden la uretra y que persisten allí mediante la elaboración de anexos superficiales conocidos como fimbriae. Los antibióticos en el flujo urinario tienen acceso limitado a las bacterias, debido a que el flujo urinario es de corta vida e infrecuente. Sin embargo, si se pudiera bloquear la síntesis de los fimbriae para desligar las bacterias, las terapias existentes se harían más efectivas. Similarmente, si el mecanismo de quimiotactismo de las bacterias se dañara, la capacidad de las bacterias para establecer una infección efectiva quedaría comprometida en algunas especies. Una matriz informadora del genoma para un patógeno bacteriano que contiene informadores para la expresión de genes involucrados en el quimiotactismo o en la síntesis de fimbriae, como ejemplos, identifica no sólo los compuestos que matan las bacterias en un ensayo puntual, sino también aquellos que interfieren con las etapas clave en la biología del patógeno. Estos compuestos serían sumamente difíciles de descubrir por los medios convencionales.

Una matriz informadora del genoma basada en las células humanas proporciona muchas aplicaciones importantes. Por ejemplo, una aplicación interesante es el desarrollo de compuestos antivirales. Cuando las células humanas son infectadas por una amplia gama de virus, las células responden de manera compleja, en la cual sólo unos pocos de los componentes han sido identificados. Por ejemplo, se inducen ciertos interferones así como una ARNs de doble cadena. Ambas respuestas ofrecen individualmente cierta protección. Una matriz que informa de la inducción de genes de interferón y de ARNs de doble cadena es capaz de detectar compuestos que podrían proteger profilácticamente a las células antes de la llegada del virus. Otros efectos protectores pueden inducirse en paralelo. La incorporación de un panel de otros genes informadores

en la matriz se utiliza para identificar aquellos compuestos con el más alto grado de especificidad.

Los procedimientos implicados al utilizar una matriz informadora del genoma, según lo actualmente descrito, son, esencialmente, como se ha descrito anteriormente. Específicamente, las etapas llevadas a cabo al generar y analizar tal base de datos de señales de salida de matrices informadoras del genoma que indexa N estructuras de datos matriciales de señales de salida según los correspondientes estímulos, repiten las etapas del procedimiento general esbozadas en las Figuras 1 y 4. Las etapas primera y segunda incluyen la asignación a n del valor entero 1 para la generación de la primera estructura de datos matriciales de señales de salida, y la construcción de una matriz, informadora del genoma, física estimulada, respectivamente. En esta realización, la matriz física estimulada, típicamente, comprende una placa de microtitulación que tiene 96 pocillos ordenados en coordenadas X e Y. Cada pocillo encierra una célula o una colonia de células que tiene la estructura de un gen informador unido operativamente a un promotor transcripcional distinto. A las células se les proporciona un estímulo capaz de inhibir a los promotores y, de esa manera, reducir la expresión del informador en múltiples pocillos. La tercera etapa del procedimiento de la matriz informadora del gen incluye la detección de una señal física que resulta de la expresión del informador en cada pocillo. Generalmente, el gen informador codifica una enzima, como la lacZ, que proporciona un producto de reacción convenientemente detectado mediante espectroscopía. La cuarta etapa incluye la transducción lineal de la señal informadora óptica en cada pocillo para generar una correspondiente señal de salida eléctrica digital, y la quinta etapa incluye el almacenamiento de cada señal de salida eléctrica en la memoria del ordenador, como una estructura de datos matriciales de señales de salida del informador del gen, que asocia cada señal de salida con las coordenadas del correspondiente pocillo de la placa de microtitulación y el estímulo. Después de la quinta etapa, la rutina avanza hasta un bloque de decisión: si no se han completado todas las estructuras de datos requeridas, N se incrementa y se repiten las etapas uno a cinco de generación de la estructura de datos para la N+1-ésima matriz informadora del genoma estimulada. Si, después de la quinta etapa, se han completado todas las estructuras de datos requeridas, la rutina avanza hasta la etapa de formar una base de datos de matrices de señales de salida, indexando las estructuras de datos matriciales de señales de salida según los estímulos. Para el análisis de un estímulo desconocido (p. ej., un fármaco candidato), se usa un sistema de AI como el previamente descrito para comparar, habitualmente, una serie de dilución de respuestas al estímulo desconocido con la base de datos.

La Figura 6 muestra las etapas llevadas a cabo al generar un perfil de respuesta de salida de la matriz informadora del gen para un estímulo desconocido, las tablas de regulación, los perfiles de respuesta de referencia basal, los perfiles de respuesta química conocidos y los perfiles de res-

puesta genética conocidos.

Para generar una matriz informadora del genoma 610, se construye un conjunto de fusiones de lacZ para un conjunto de amplio alcance de genes de levadura. Las fusiones se construyen, generalmente, en una célula diploide de tipo de apareamiento A/A, para permitir la introducción de mutaciones dominantes por apareamiento, aunque las estirpes haploides también pueden utilizarse con informadores particularmente sensibles para ciertas funciones. Las fusiones se disponen sobre una rejilla que separa fusiones distintas en unidades que tienen coordenadas X-Y definidas. La función de identificación del gen 612 se lleva a cabo determinando, para cada gen marcado por el informador, una secuencia corta adyacente a la posición de la fusión. Esa secuencia se compara luego con la base de datos genómicos de la levadura a fin de establecer la identidad del gen. Se establece una tabla de índices 614 para ligar cada gen en la matriz con la coordenada X-Y de la estructura de fusión para ese gen.

La función de determinación de respuesta basal 616 se lleva a cabo midiendo la respuesta basal de cada célula de la matriz en diversas condiciones físicas, tales como la temperatura y el pH, el medio y la osmolaridad. Esta información se indexa con respecto a la matriz para formar el conjunto de perfiles de respuesta de referencia 618, que se utilizará para determinar la respuesta de cada informador para cada medio en el cual pueda proporcionarse un estímulo.

La función de tratamiento de compuestos 629 se lleva a cabo poniendo en contacto cada unidad de la matriz con un compuesto de ensayo. Generalmente, se transfiere una copia de la matriz completa a un nuevo medio que contiene el primer compuesto de interés, y se obtiene la respuesta para la matriz completa. En una función de retirada de referencia 622, se retira el perfil de respuesta de referencia adecuado del perfil de respuesta, y la diferencia se almacena en la base de conocimientos como el primer perfil de respuesta química 624. Como alternativa, el perfil de respuesta se divide entre el perfil de referencia adecuado para dar como resultado una razón de inducción. El proceso se repite para los compuestos, o mezclas de compuestos, 2 a N.

La función de mutación de genes 626 determina la respuesta de la matriz a la pérdida de función de cada proteína o gen o ARN en la célula, introduciendo un alelo dominante de un gen a cada célula informadora, y determinando la respuesta del informador como una función de la mutación. Con este fin, se prefieren las mutaciones dominantes, pero pueden utilizarse otros tipos de mutaciones. Las mutaciones dominantes se crean por mutagénesis *in vitro* de genes clonados, seguida por una comprobación en las células diploides de los alelos mutantes dominantes. Como alternativa, la matriz informadora puede desarrollarse en una estirpe deficiente para las funciones genéticas UPF, en las cuales la mayoría de las mutaciones sin sentido causan un fenotipo dominante, permitiendo que se construyan mutaciones dominantes para cualquier gen. Los datos obtenidos identifican los perfiles 628 de respuesta genética 1-N. Se observa que N puede ser mayor

que el número total de genes en una especie.

Estos datos están sujetos a una función de ordenamiento 630, que ordena por respuesta genética individual a fin de determinar la especificidad de cada gen para un estímulo particular. Se establece una matriz de ponderación 632, que pondera las señales proporcionalmente a la especificidad de los correspondientes informadores. La matriz de ponderación se revisa dinámicamente, incorporando datos de cada revisión a la matriz de ponderación N+1. Una función de regulación de genes 634 se utiliza luego para construir tablas de regulación 634, 636, que identifican qué celdas de la matriz responden a qué mutación en un gen indexado, y qué mutaciones afectan a qué células de la matriz.

La función de revisión de la matriz de estímulo desconocido 636 prueba secuencialmente nuevas sustancias químicas, o compuestos desconocidos, o mezclas desconocidas, para identificar perfiles de respuesta de salida 642.

La Figura 7 muestra las etapas llevadas a cabo al analizar un perfil de respuesta de salida de la matriz informadora del gen para un estímulo desconocido. Un perfil de respuesta de salida 710, creado para un nuevo estímulo, se muestra sujeto a dos rutas de análisis alternativas. En la ruta mostrada a la izquierda, de los elementos 712-732, el nuevo perfil de respuesta a estímulos 710 está sujeto a una función de comparación 712 que compara los nuevos perfiles de respuesta a estímulos con los perfiles de respuesta a estímulos químicos conocidos. La función de comparación 712, típicamente, proporciona un análisis de comparación en forma de un informe indexado de las coincidencias con los perfiles de respuesta química de referencia, o los perfiles de respuesta genética, clasificados según el valor ponderado de cada informador coincidente, la calificación de evaluación de la respuesta de salida (CALIFICACIÓN ERS). La comparación puede continuar a lo largo de un árbol de decisión lineal: la primera consulta 714 es si hay una coincidencia; si la hay (es decir, una CALIFICACIÓN ERS perfecta), el perfil de respuesta de salida identifica un estímulo con el mismo objetivo 716, o una ruta señalada como objetivo, como uno de los compuestos conocidos sobre los cuales se construye la base de datos de perfiles de respuesta. En caso contrario, la segunda consulta 718 es si el perfil de respuesta de salida es un subconjunto de células en la matriz estimulado por un compuesto conocido. En ese caso, entonces el nuevo compuesto es un candidato 720 para una molécula con mayor especificidad que el compuesto de referencia. En particular, si los informadores que responden unívocamente a la sustancia química de referencia tienen un valor bajo de respuesta ponderada, se concluye que el nuevo compuesto es de mayor especificidad. Como alternativa, si los informadores que responden unívocamente al compuesto de referencia tienen un alto valor de respuesta ponderada, se concluye que el nuevo compuesto es activo aguas abajo en la misma ruta. Si el perfil de respuesta de salida no es un subconjunto de células en la matriz estimulado por un compuesto conocido, la tercera consulta 722 es si la salida se

solapa con el perfil de respuesta de un compuesto de referencia conocido. De ser así, el solapamiento se somete a una función de ordenamiento 724, en la cual se evalúa cuantitativamente con la matriz de ponderación para dar como resultado los informadores común 726 y único 728. Los informadores únicos se ordenan 730 según las tablas de regulación 618, 620, y las mejores coincidencias se usan para deducir el objetivo candidato 732. Si el perfil de respuesta de salida no se solapa, o bien no coincide con un perfil de respuesta química, entonces la base de datos es inadecuada para inferir la función, y el perfil de respuesta de salida puede añadirse a los perfiles de respuesta química de referencia.

En la ruta mostrada a la derecha de la Figura 7 (los elementos 736-750), el perfil de respuesta de salida de un nuevo estímulo químico se compara 736 con el perfil de respuesta genético para el gen de destino. La primera consulta 738 es si hay una coincidencia entre los dos perfiles de respuesta. De ser así, la deducción 740 es que el gen de destino es el presunto objetivo de la sustancia química. En caso contrario, la segunda consulta 742 es si el perfil de respuesta química es un subconjunto de un perfil de respuesta genética. De ser así, la deducción 744 es que el objetivo del fármaco está más abajo del gen mutante, pero en la misma ruta. En caso contrario, la tercera consulta 746 es si el perfil de respuesta de salida incluye como subconjunto un perfil de respuesta genética. De ser así, la deducción 748 es que el objetivo de la sustancia química está en la misma ruta que el gen de destino, pero más arriba. En caso contrario, la deducción 750 es que el perfil de respuesta químico es nuevo y define una ruta huérfana.

Los siguientes ejemplos se ofrecen a modo de ilustración, y no a modo de limitación.

### Ejemplos

#### 1. Matriz de genes promotores-informadores transcripcionales

##### A) Construcción de una matriz física estimulada con el fármaco Mevinolin (lovastatina)

El Mevinolin es un compuesto conocido por inhibir la biosíntesis del colesterol. Inicialmente, se determinó que la concentración máxima no tóxica (según lo medido por el crecimiento y la viabilidad celular) de Mevinolin en las células, mediante dilución en serie, era de 25  $\mu\text{g}/\text{ml}$ . Para producir una matriz estimulada por Mevinolin, cada placa de microtitulación de 60 pocillos se rellena con 100  $\mu\text{l}$  de medio de cultivo que contiene Mevinolin 25  $\mu\text{g}/\text{ml}$  en una solución de etanol al 2%. Se añade una alícuota de cada miembro de la matriz informadora a cada pocillo, admitiendo una dilución de aproximadamente 1:100. Los pocillos se incuban en el medio hasta que la turbidez del informador promedio aumenta 20 veces. Cada celda se cuantifica luego en cuanto a la turbidez como medida de crecimiento, y se trata con una solución de lisis para permitir la medición de la  $\beta$ -galactosidasa de cada fusión.

##### B) Generación de una estructura de datos matriciales de señales de salida

Tanto la turbidez como la  $\beta$ -galactosidasa se leen en contadores de placa de microtitulación comercialmente disponibles (p. ej., BioRad) y los

datos se capturan como un fichero ASCII. De este fichero, se retira el valor de los pocillos individuales en la matriz informadora en una solución de etanol al 2% en el perfil de respuesta de referencia. La diferencia corresponde al perfil de respuesta del Mevinolin. Este fichero se convierte en el ordenador en una tabla indexada por la respuesta de cada pocillo al inhibidor. Por ejemplo, los genes que codifican la acetoacetil-CoA tiolasa y la escualeno sintasa aumentan 10 veces, mientras que SIR3 y LEU2, dos genes no relacionados entre sí, permanecen invariables. La respuesta de la matriz informadora a otros compuestos se determina de manera similar, y se almacena como perfiles de respuesta de salida.

##### C) Comparación de la estructura de datos de la Matriz de Señales de Salida con una base de datos de la Matriz de Señales de Salida

Se construye una matriz física según se describe más arriba, excepto que el Mevinolin se reemplaza por un compuesto de ensayo desconocido. El perfil resultante de respuesta de salida se compara con los perfiles de respuesta de una biblioteca de compuestos bioactivos conocidos, y se analiza según se describe anteriormente. Por ejemplo, si el perfil de salida del compuesto de ensayo muestra tanto la acetoacetil-CoA tiolasa como la escualeno sintasa inducidas por gen, entonces el perfil de salida coincide con el esperado de un inhibidor de la síntesis del colesterol. Si el perfil de respuesta de salida tiene menos células afectadas que el perfil de respuesta al Mevinolin, el compuesto desconocido es un candidato para mayor especificidad. Si el perfil de respuesta de salida de la nueva sustancia química afecta a menos informadores que el perfil de respuesta al Mevinolin, y si los demás informadores afectados por el Mevinolin tienen un menor valor ponderado, entonces el compuesto es un candidato para una mayor especificidad. Si el perfil de respuesta de salida tiene más células distintas afectadas que el perfil de respuesta al Mevinolin, entonces el compuesto es un candidato para una menor especificidad. En el caso en que se ensayen mezclas de compuestos, se evalúan las mayores respuestas ponderadas para determinar si pueden ser desarrolladas hacia el perfil de respuesta de dos compuestos distintos, o de dos perfiles de respuesta genética distintos.

#### 2. Matriz de sondeo de hibridación transcripcional de oligonucleótidos de informador: construcción de la matriz física estimulada y generación de una estructura de datos matriciales de señales de salida

Las sondas de hibridación de oligonucleótidos no marcadas, complementarias con el ARNm de transcripción cada gen de levadura, se disponen sobre un sustrato de sílice grabado con técnicas estándar (p. ej., Fodor *et al.* (1991) Science 252, 767). Las sondas son de una longitud y secuencia tales como para asegurar la especificidad para el correspondiente gen de la levadura, típicamente, de alrededor de 24-240 nucleótidos de longitud. Se trata un cultivo celular HeLa confluyente con Mevinolin 15  $\mu\text{g}/\text{ml}$  en etanol al 2% durante 4 horas, mantenido en una atmósfera de  $\text{CO}_2$  con humedad al 5% a 37°C. Se extrae el ARN mensajero, transcrito en notación inversa y marcado con

un fluoróforo según los procedimientos estándar (Sambrook *et al.*, Molecular Cloning [Clonación Molecular], 3<sup>a</sup> ed.). El ADNc resultante se hibrida para la formación de sondas, la formación se lava hasta liberarla de ADNc marcado no hibridado, se cuantifica la señal de hibridación en cada unidad de la formación usando un escáner microscópico confocal (Dispositivos Moleculares) y los datos de respuesta matricial resultantes se almacenan en forma digital.

### 3. Matriz bi-híbrida bidimensional

#### A) Construcción de una matriz física estimulada

La matriz bi-híbrida bidimensional se diseña para comprobar los compuestos que afecten específicamente la interacción de dos proteínas, p. ej., la interacción de un transductor de señal y activador de transcripción (TSAT) humano con un receptor de interleuquina. Se generan dos fusiones híbridas mediante procedimientos estándar: cada estirpe contiene una porción del gen humano TSAT diana, fusionado con una porción de un gen de levadura o bacteriano que codifica un dominio ligado a ADN (p. ej., GALA:1-147). La secuencia de ADN reconocida por el dominio ligado a ADN (p. ej., UAS<sub>G</sub>) se inserta en lugar de la secuencia potenciadora 5' en el informador seleccionado (p. ej., lacZ). La estirpe también contiene otra fusión que consiste en una porción intracelular del gen receptor diana, cuyo producto proteínico interactúa con el TSAT. Este gen receptor está fusionado con un fragmento de gen que codifica

un dominio de activación transcripcional (p. ej., GAL4:768-881).

#### B) Generación de estructura de datos matriciales de señales de salida

Tanto la turbidez como la galactosidasa se leen en contadores de placa de microtitulación comerciales (BioRad), y los datos se capturan como un fichero ASCII.

#### C) Comparación de la estructura de datos matriciales de señales de salida con la base de datos

Los datos se analizan buscando aquellos compuestos que bloquean la interacción de las dos proteínas humanas, reduciendo la señal producida desde el informador en las diversas estirpes que contienen pares de proteínas humanas. La salida se procesa para identificar compuestos con un gran impacto sobre un informador cuya expresión es dependiente de un único par de proteínas humanas que interactúan. Se usa una matriz de ponderación invertida para evaluar estos datos, ya que los compuestos preferidos no afectan ni siquiera a los informadores menos específicos en la matriz.

Aunque la invención precedente ha sido descrita en bastante detalle por medio de ilustraciones y ejemplos con fines de claridad y comprensión, será inmediatamente evidente para los expertos en la técnica, a la vista de las enseñanzas de esta invención, que pueden hacerse ciertos cambios y modificaciones en ella sin apartarse del alcance de las reivindicaciones adjuntas.

5

10

15

20

25

30

35

40

45

50

55

60

65

## REIVINDICACIONES

1. Un procedimiento para analizar una matriz de señales de salida por comparación con una base de datos matriciales de señales de salida, útil en sistemas expertos y redes neurales para correlacionar estímulos candidatos y respuestas sistémicas de un ser viviente, que comprende las etapas de:

(a) construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad un respondedor distinto de un ser viviente que comprende múltiples respondedores distintos o una sonda que corresponde a dicho respondedor distinto, y un identificador para dicho respondedor distinto o dicha sonda, proporcionando a dicho ser viviente un estímulo que inhibe dicho respondedor distinto de dichas múltiples unidades, suministrando dicho identificador una señal física correspondiente a la inhibición de dicho respondedor distinto, y comprendiendo dicha formación una clase funcional predeterminada de dichos múltiples respondedores distintos;

(b) detección de la señal física en cada citada unidad de la matriz física;

(c) transducción de cada dicha señal física para generar una correspondiente señal de salida eléctrica.

(d) almacenamiento de cada señal de salida eléctrica en una estructura de datos matriciales de señales de salida que asocia a cada señal de salida las coordenadas X e Y de la correspondiente unidad de la matriz física y dicho estímulo;

(e) determinación del efecto de dicho estímulo sobre dicho ser viviente, comparando la estructura de datos matriciales de señales de salida de la etapa (d) con una base de datos matriciales de señales de salida producida por un procedimiento que comprende las etapas:

(i) construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad un respondedor distinto de un ser viviente que comprende múltiples respondedores distintos, o una sonda que corresponde a dicho respondedor distinto, y un identificador para dicho respondedor distinto o dicha sonda, proporcionando a dicho ser viviente un estímulo que inhibe dicho respondedor distinto de dichas múltiples unidades, suministrando dicho identificador una señal física correspondiente a la inhibición de dicho respondedor distinto, y comprendiendo dicha formación una clase funcional predeterminada de los citados múltiples respondedores distintos;

(ii) detección de la señal física en cada dicha unidad de la matriz física;

(iii) transducción de cada dicha señal física para generar una correspondiente señal de salida eléctrica;

(iv) almacenamiento de cada señal de salida eléctrica en una estructura de datos matriciales de señales de salida que asocia cada señal de salida con las coordenadas X e Y de la correspondiente unidad de matriz física y dicho estímulo;

(v) repetición de las etapas (i) - (iv) para almacenar iterativamente las estructuras de datos

matriciales de señales de salida para múltiples estímulos, a fin de formar una base de datos matriciales de señales de salida que indexe las estructuras de datos matriciales de señales de salida según los estímulos.

2. Un procedimiento según la reivindicación 1, comprendiendo dicha etapa de comparación la comparación de la estructura de datos matriciales de señales de salida de la etapa (d) con dicha base de datos matriciales de señales de salida, según las reglas de comparación contenidas en una base de conocimientos.

3. Un procedimiento según la reivindicación 1, comprendiendo dicha etapa de comparación el uso de una red neural entrenada sobre dicha base de datos de señales de salida para comparar la estructura de datos matriciales de señales de salida con dicha base de datos de señales de salida.

4. Un procedimiento según la reivindicación 1, siendo dicha señal física una señal óptica.

5. Un procedimiento según la reivindicación 1, comprendiendo dicha clase funcional un conjunto suficiente de todos los respondedores distintos de dicho ser viviente para deducir en dicha etapa de comparación (e) la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

6. Un procedimiento según la reivindicación 1, comprendiendo dicha clase funcional una mayoría de todos los respondedores distintos de dicho ser viviente.

7. Un procedimiento según la reivindicación 1, comprendiendo dicha etapa de construcción una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad una célula que contiene una estructura recombinante que comprende un gen informador, estando la expresión de dicho gen informador operativamente ligada a la expresión de un gen endógeno distinto de un único organismo que comprende múltiples genes endógenos distintos, proporcionando a cada una de las células citadas un estímulo que inhibe la expresión de dicho gen endógeno distinto en dichas múltiples unidades, suministrando la expresión de dicho gen informador una señal física correspondiente a la inhibición de la expresión de dicho gen endógeno distinto, y comprendiendo dicha formación una clase funcional predeterminada de dichos múltiples genes endógenos distintos.

8. Un procedimiento según la reivindicación 7, comprendiendo dicha clase funcional un conjunto suficiente de todos los dichos genes endógenos distintos de dicho organismo para deducir en dicha etapa de comparación (e) la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

9. Un procedimiento según la reivindicación 7, comprendiendo dicha clase funcional una mayoría de todos los citados genes endógenos distintos de dicho organismo.

10. Un procedimiento según la reivindicación 1, comprendiendo dicha etapa de construcción la construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad una célula que contiene una estructura re-

combinante que comprende un gen informador, estando la expresión de dicho gen informador operativamente ligada a la función de una proteína endógena distinta de un único organismo que comprende múltiples proteínas endógenas distintas, proporcionando a cada una de las citadas células un estímulo que inhibe la función de dicha proteína endógena distinta en dichas múltiples unidades, suministrando la expresión de dicho gen informador una señal física correspondiente a la inhibición de la función de dicha proteína endógena distinta, y comprendiendo dicha formación una clase funcional predeterminada de dichas múltiples proteínas endógenas distintas.

11. Un procedimiento según la reivindicación 10, comprendiendo dicha clase funcional un conjunto suficiente de todas las citadas proteínas endógenas distintas de dicho organismo para deducir en dicha etapa de comparación (e) la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

12. Un procedimiento según la reivindicación 10, comprendiendo dicha clase funcional una mayoría de todas las citadas proteínas endógenas distintas de dicho organismo.

13. Un procedimiento según la reivindicación 1, comprendiendo dicha etapa de construcción una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad un hibridante específico para una transcripción endógena distinta de un ser viviente que comprende múltiples transcripciones endógenas distintas, o para una sonda correspondiente a dicha transcripción endógena distinta, proporcionando al citado ser viviente un estímulo que inhibe la expresión de dicha transcripción endógena distinta de dichas múltiples unidades, suministrando dicho hibridante una señal física correspondiente a la inhibición de la expresión de dicha transcripción endógena distinta, y comprendiendo dicha formación una clase funcional predeterminada de dichas múltiples transcripciones endógenas distintas.

14. Un procedimiento según la reivindicación 13, comprendiendo dicha clase funcional un conjunto suficiente de todas las citadas transcripciones endógenas distintas de dicho ser viviente para deducir en dicha etapa de comparación (e) la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

15. Un procedimiento según la reivindicación 13, comprendiendo dicha clase funcional una mayoría de todas las dichas transcripciones endógenas distintas de dicho ser viviente.

16. Un procedimiento para generar una base de datos de matrices de señales de salida, útil para correlacionar estímulos candidatos y respuestas sistémicas, que comprende las etapas de:

(a) construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad un respondedor distinto de un ser viviente que comprende múltiples respondedores distintos o una sonda que corresponde a dicho respondedor distinto, y un identificador para dicho respondedor distinto o dicha sonda, proporcionando el citado ser viviente un estímulo que

inhibe dicho respondedor distinto de dichas múltiples unidades, suministrando dicho identificador una señal física correspondiente a la inhibición de dicho respondedor distinto, y comprendiendo dicha formación una clase funcional predeterminada de dichos múltiples respondedores distintos.

(b) detección de la señal física en cada dicha unidad de la matriz física;

(c) transducción de cada dicha señal física para generar una correspondiente señal de salida eléctrica;

(d) almacenamiento de cada señal de salida eléctrica en una estructura de datos matriciales de señales de salida que asocia cada señal de salida con las coordenadas X e Y de la correspondiente unidad de la matriz física y dicho estímulo;

(e) repetición de las etapas (a) - (d) para almacenar iterativamente las estructuras de datos matriciales de señales de datos para múltiples estímulos, a fin de formar una base de datos matriciales de señales de salida que indexe las estructuras de datos matriciales de señales de salida según los estímulos.

17. Un procedimiento según la reivindicación 16, en el que dicha señal física es una señal óptica.

18. Un procedimiento según la reivindicación 16, comprendiendo dicha clase funcional un conjunto suficiente de todos los respondedores distintos de dicho ser viviente para deducir la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

19. Un procedimiento según la reivindicación 16, comprendiendo dicha clase funcional una mayoría de todos los respondedores distintos de dicho ser viviente.

20. Un procedimiento según la reivindicación 16, comprendiendo dicha etapa de construcción la construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad una célula que contiene una estructura recombinante que comprende un gen informador, estando la expresión de dicho gen informador operativamente ligada a la expresión de un gen endógeno distinto de un único organismo que comprende múltiples genes endógenos distintos, proporcionando a la citada célula un estímulo que inhibe la expresión de dicho gen endógeno distinto en dichas múltiples unidades, suministrando la expresión de dicho gen informador una señal física correspondiente a la inhibición de la expresión de dicho gen endógeno distinto, y comprendiendo dicha formación una clase funcional predeterminada de dichos múltiples genes endógenos distintos.

21. Un procedimiento según la reivindicación 20, comprendiendo dicha clase funcional un conjunto suficiente de todos los genes endógenos distintos de dicho organismo, a fin de deducir la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

22. Un procedimiento según la reivindicación 20, comprendiendo dicha clase funcional una mayoría de todos los genes endógenos distintos de dicho organismo.

23. Un procedimiento según la reivindicación 16, comprendiendo dicha etapa de construcción la construcción de una matriz física estimulada

que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad una célula que contiene una estructura recombinante que comprende un gen informador, estando la expresión de dicho gen informador operativamente ligada a la función de una proteína endógena distinta de un único organismo que comprende múltiples proteínas endógenas distintas, proporcionando a la citada células un estímulo que inhibe la función de dicha proteína endógena distinta en dichas múltiples unidades, suministrando la expresión de dicho gen informador una señal física correspondiente a la inhibición de la función de dicha proteína endógena distinta, y comprendiendo dicha formación una clase funcional predeterminada de dichas múltiples proteínas endógenas distintas.

24. Un procedimiento según la reivindicación 23, comprendiendo dicha clase funcional un conjunto suficiente de todas las proteínas endógenas distintas de dicho organismo para deducir la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

25. Un procedimiento según la reivindicación 23, comprendiendo dicha clase funcional una mayoría de todas las proteínas endógenas distintas de dicho organismo.

26. Un procedimiento según la reivindicación 16, comprendiendo dicha etapa de construcción la

construcción de una matriz física estimulada que comprende una formación ordenada de unidades que tienen coordenadas X e Y, encerrando cada unidad un hibridante específico para una transcripción endógena distinta de un ser viviente que comprende múltiples transcripciones endógenas distintas, o para una sonda correspondiente a dicha transcripción endógena distinta, proporcionando el citado ser viviente un estímulo que inhibe la expresión de dicha transcripción endógena distinta de dichas múltiples unidades, suministrando dicho hibridante una señal física correspondiente a la inhibición de la expresión de dicha transcripción endógena distinta, y comprendiendo dicha formación una clase funcional predeterminada de dichas múltiples transcripciones endógenas distintas.

27. Un procedimiento según la reivindicación 26, comprendiendo dicha clase funcional un conjunto suficiente de todas las transcripciones endógenas distintas de dicho ser viviente para deducir la ruta por la cual dicho estímulo deduce dicha estructura de datos matriciales de señales de salida.

28. Un procedimiento según la reivindicación 26, comprendiendo dicha clase funcional una mayoría de todas las transcripciones endógenas distintas de dicho ser viviente.

---

**NOTA INFORMATIVA:** Conforme a la reserva del art. 167.2 del Convenio de Patentes Europeas (CPE) y a la Disposición Transitoria del RD 2424/1986, de 10 de octubre, relativo a la aplicación del Convenio de Patente Europea, las patentes europeas que designen a España y solicitadas antes del 7-10-1992, no producirán ningún efecto en España en la medida en que confieran protección a productos químicos y farmacéuticos como tales.

---

Esta información no prejuzga que la patente esté o no incluida en la mencionada reserva.

---

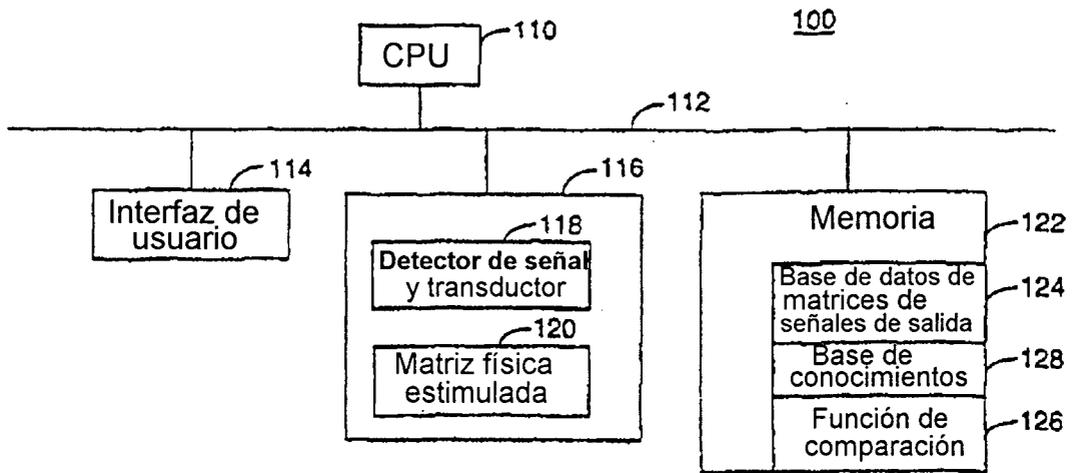


FIGURA 1

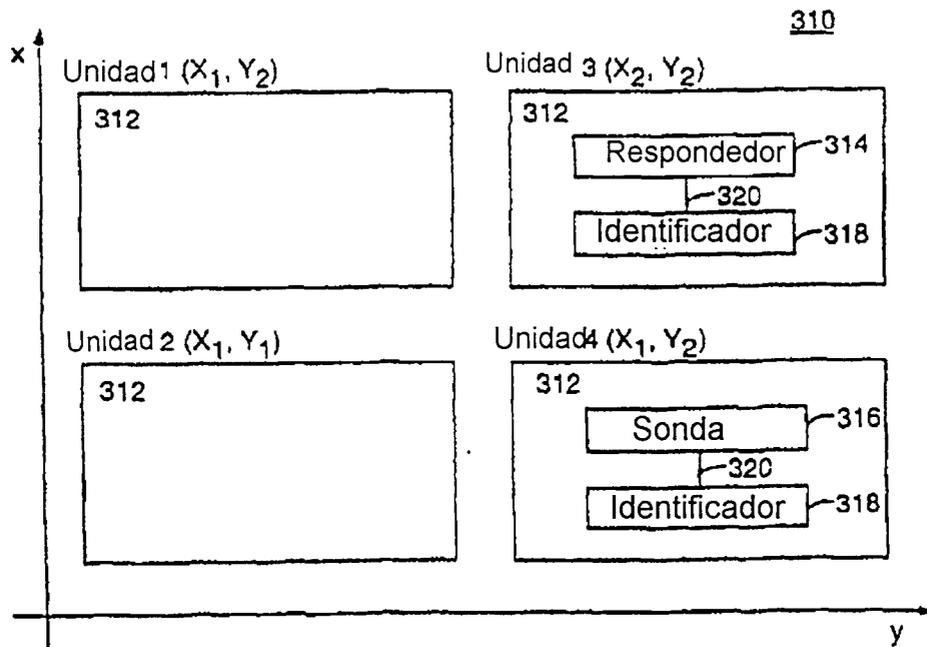


FIGURA 3

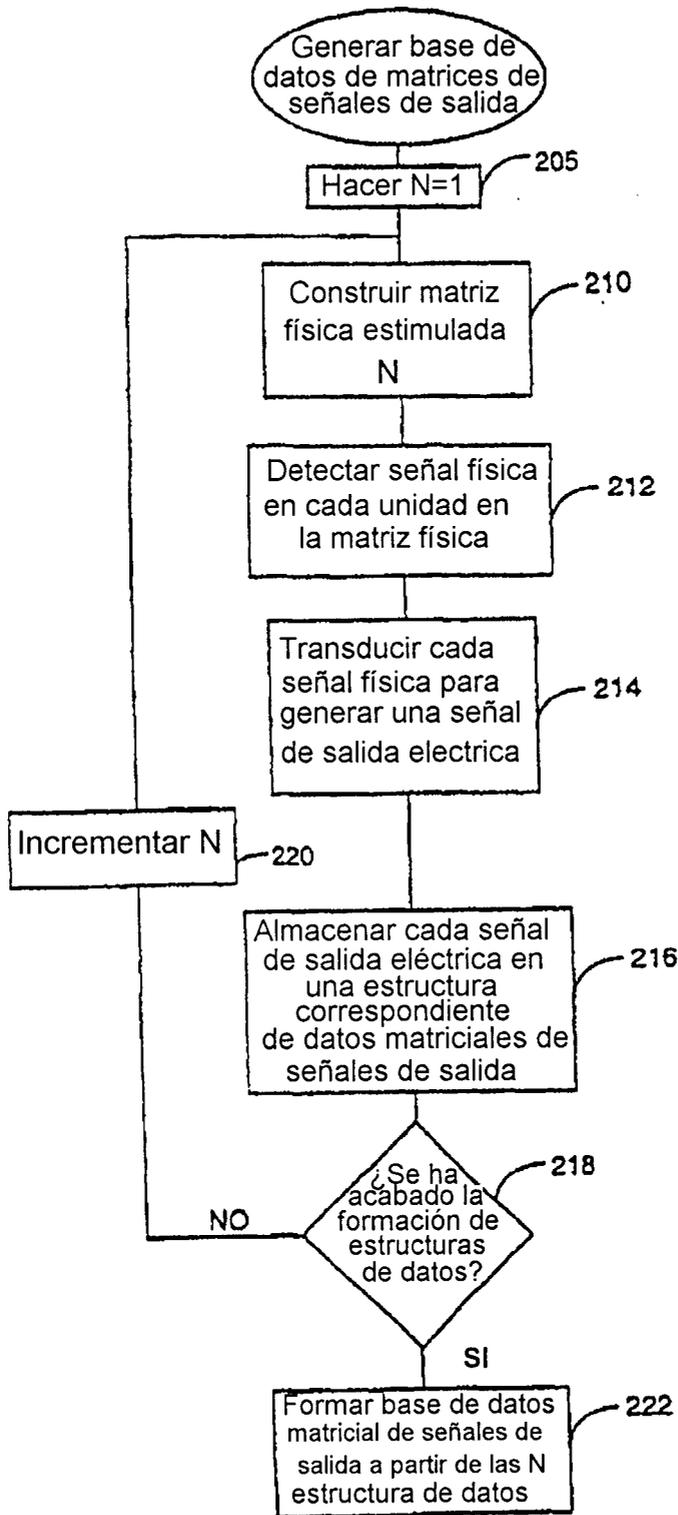


FIGURA 2

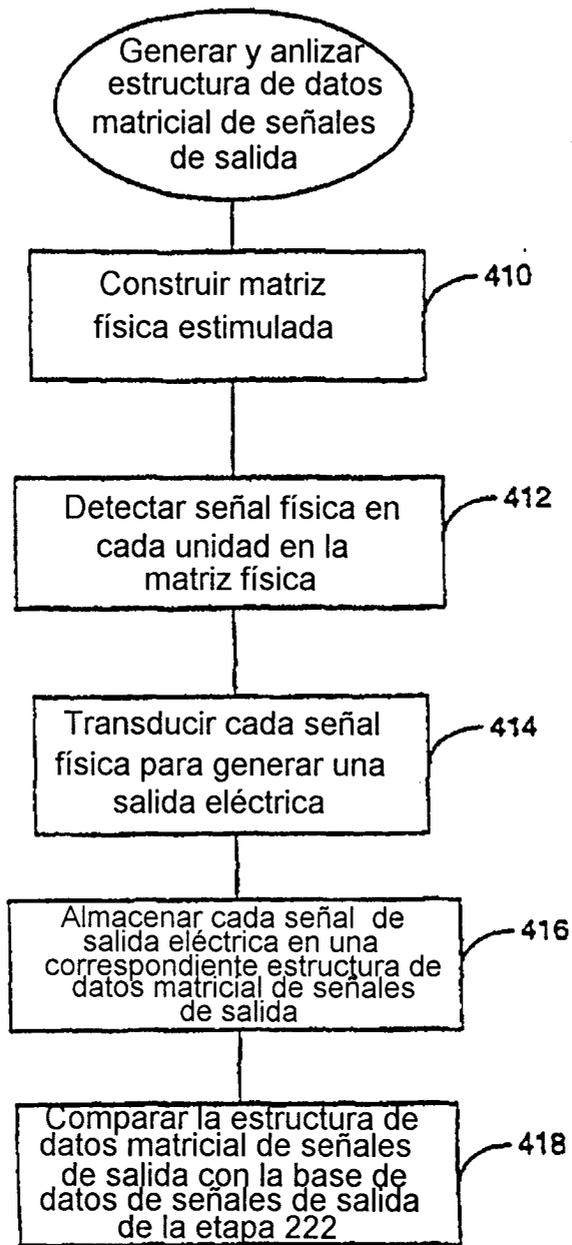


FIGURA 4

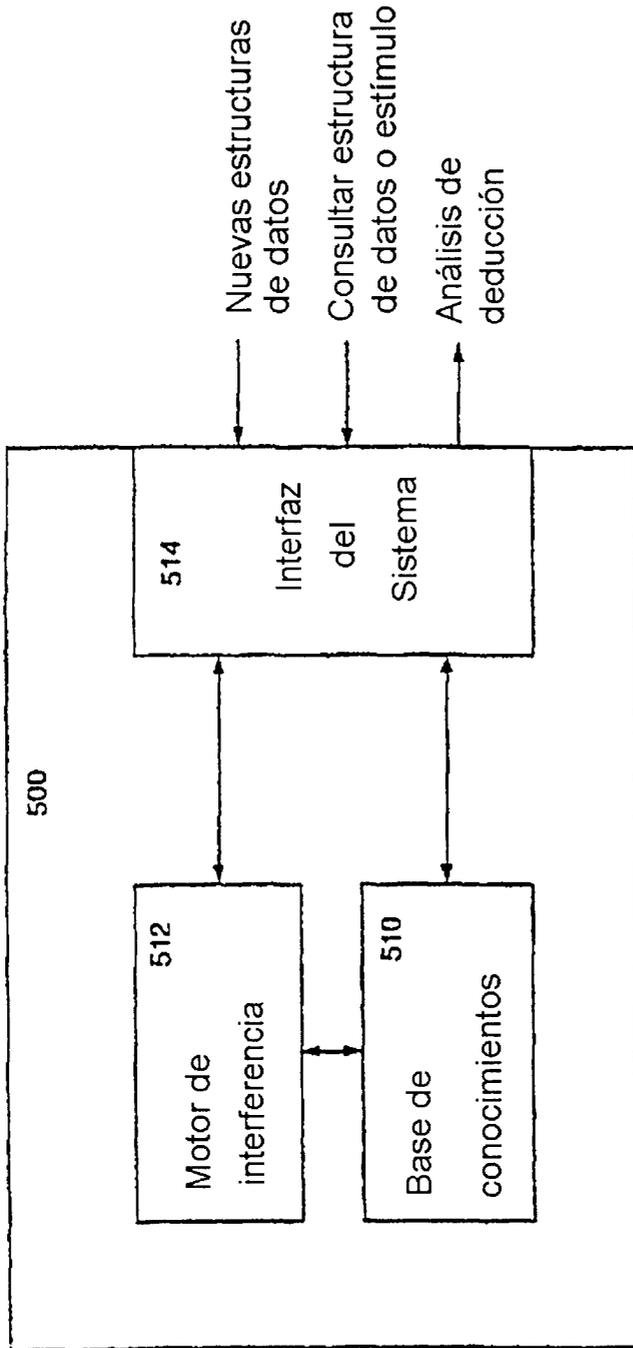


FIGURA 5

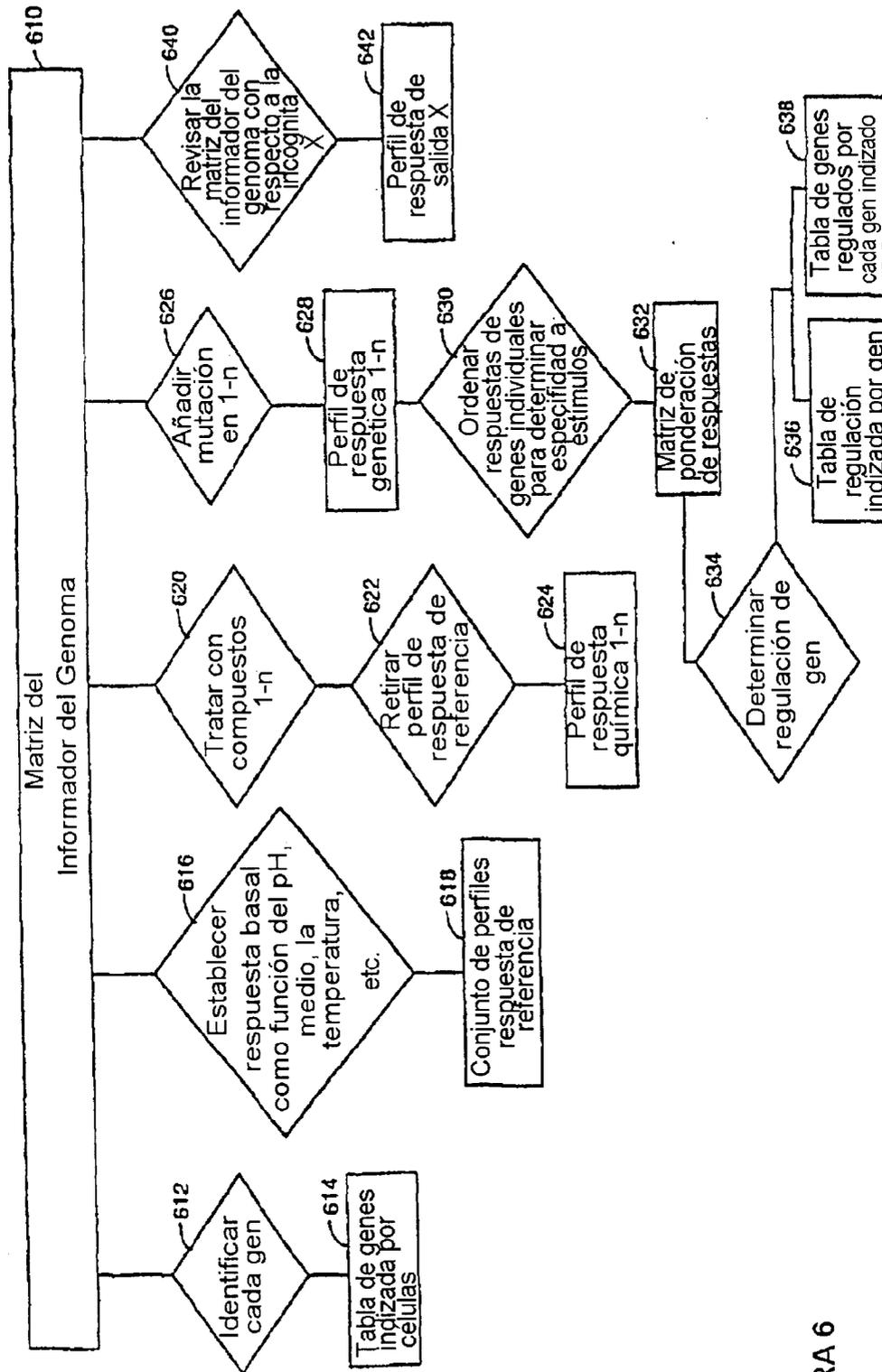


FIGURA 6

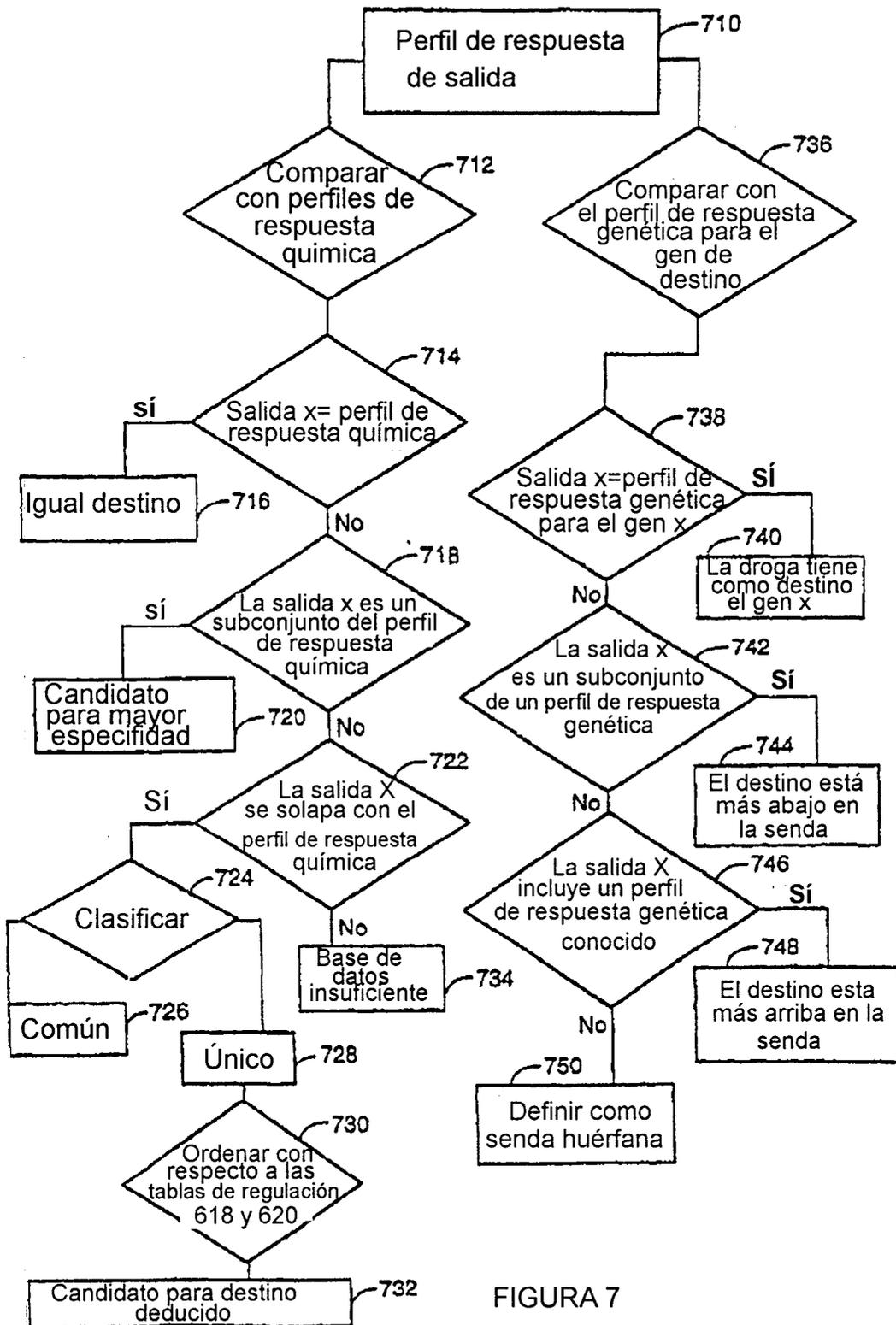


FIGURA 7