

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 726 941**

51 Int. Cl.:

**C12N 9/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.03.2002 PCT/FR2002/00951**

87 Fecha y número de publicación internacional: **26.09.2002 WO02074943**

96 Fecha de presentación y número de la solicitud europea: **18.03.2002 E 02722349 (4)**

97 Fecha y número de publicación de la concesión europea: **20.02.2019 EP 1427818**

54 Título: **Moléculas de ácidos nucleicos que codifican una dextrano-sacarasa que cataliza la síntesis de dextrano que porta ramificaciones de tipo alfa-1,2 osídicas**

30 Prioridad:

**16.03.2001 FR 0103631  
19.12.2001 FR 0116495**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**10.10.2019**

73 Titular/es:

**INSTITUT NATIONAL DES SCIENCES  
APPLIQUÉES (INSA) (100.0%)  
135 avenue de Rangueil  
31077 Toulouse Cedex 4 , FR**

72 Inventor/es:

**MONSAN, PIERRE;  
REMAUD, MAGALI;  
BOZONNET, SOPHIE y  
WILLEMOT, RENÉ-MRAC**

74 Agente/Representante:

**ISERN JARA, Jorge**

**Observaciones:**

**Véase nota informativa (Remarks, Remarques o  
Bemerkungen) en el folleto original publicado por  
la Oficina Europea de Patentes**

**ES 2 726 941 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

Moléculas de ácidos nucleicos que codifican una dextrano-sacarasa que cataliza la síntesis de dextrano que porta ramificaciones de tipo alfa-1,2 osídicas

5 La presente solicitud se refiere al campo de la glucotecnología y, más particularmente, de la síntesis de oligosacáridos u oligósidos con efecto prebiótico, terapéutico o diagnóstico.

10 La presente solicitud divulga moléculas de ácidos nucleicos que codifican una enzima que tiene una actividad de glucosiltransferasa que cataliza la síntesis de dextranos o de oligósidos que portan ramificaciones de tipo  $\alpha(1 \rightarrow 2)$  osídicas.

15 La solicitud divulga también las enzimas sintetizadas por los ácidos nucleicos descritos en este contexto, así como sus sistemas de expresión en células procariontas o eucariotas. La solicitud divulga finalmente la utilización de dichas enzimas en la producción de oligosacáridos en la alimentación, o como principio activo de productos terapéuticos y/o cosméticos. La invención se refiere más particularmente a polipéptidos aislados tales como se definen en las reivindicaciones 1 y 2, sobre ácidos nucleicos aislados tales como se definen en las reivindicaciones 3 y 4, vectores tales como se definen en las reivindicaciones 5 y 6, células según las reivindicaciones 7 y 8, procesos de producción de una glucosiltransferasa según las reivindicaciones 9 a 11 y la utilización de la glucosiltransferasa obtenida en la fabricación de una composición farmacéutica o cosmética, según la reivindicación 12.

20 Los oligósidos y heterooligósidos juegan el papel de señales de reconocimiento y de efector tanto en animales como en plantas (se habla entonces de oligosacarinas), al unirse específicamente a lectinas, glucosiltransferasas, glucosidasas, moléculas de adhesión, etc... De este modo, los determinantes antigénicos de los grupos sanguíneos son ósidis, y nuestra defensa contra una serie de bacterias patógenas está dirigida contra las estructuras osídicas de la envuelta bacteriana. Por otro lado, una de las razones principales del rechazo de los xenoinjertos es la existencia de estructuras osídicas propias de cada especie. Estas propiedades, así como los conocimientos adquiridos estos últimos años sobre las glucosiltransferasas y las lectinas, contribuyen a hacer de ciertos oligósidos candidatos preferentes para terapéutica o profilaxis de trastornos asociados al equilibrio microbiológico de diferentes órganos tales como el intestino, o la piel. Por ejemplo, los oligósidos constituyen una alternativa interesante a la utilización de microorganismos y de antibióticos para regular la composición de la flora intestinal (efecto prebiótico). Ciertos oligósidos pueden ser considerados "fibras solubles" cuando no son metabolizados por las enzimas digestivas humanas y animales; al llegar al colon, interactúan con la flora microbiana y afectan específicamente al crecimiento y la adhesión de ciertas especies. Incorporadas a dosis baja (menos del 1 %) en la alimentación, ciertas de estas moléculas osídicas mejoran el estado de salud y estimulan el aumento de peso de los animales.

35 Una revisión de las diferentes glucosiltransferasas, su estructura y su actividad, se describe en Vincent Monchois et al. (1). En resumen:

40 a) Parece que la estructura de las glucosiltransferasas y/o dextrano-sacarosas estudiadas está muy conservada y está constituida, partiendo de la parte aminada de la proteína, por una secuencia señal, por un dominio variable, por un dominio catalítico y por un dominio de unión a glucano.

45 b) Los glucooligósidos (GOS) son sintetizables por glucosiltransferasas tales como dextrano-sacarosas, a partir de sustratos poco costosos tales como sacarosa y en presencia de un azúcar aceptor de glucosa. Otros sustratos, tales como  $\alpha$ -D-fluoro-glucosa, paranitrofenil- $\alpha$ -D-glucopiranosido,  $\alpha$ -D-glucopiranosido- $\alpha$ -D-sorbofuranósido o 4-O- $\alpha$ -D-galactopiranosilsacarosa también pueden utilizarse.

50 Estas enzimas catalizan a partir del sustrato la transferencia de unidades de glucosa en moléculasceptoras. En presencia de un aceptor de glucosa tal como maltosa, o isomaltosa, las glucosiltransferasas catalizan la síntesis de oligosacáridos de bajo peso molecular que comprenden mayoritariamente cadenas de 3 a 7 glucosas, de acuerdo con la reacción:



55 En tal caso, las enzimas presentan generalmente una especificidad de síntesis de enlaces osídicos conformes a la que forma el polímero donador.

60 En cambio, en ausencia de aceptor, la enzima sintetiza glucanos de alto peso molecular de tipo dextrano mediante transferencias sucesivas de unidades  $\alpha$ -D-glucopiranosilos a partir de sacarosa, de acuerdo con la reacción:



65 c) Las estructuras y la función de los glucanos o de los oligósidos sintetizados por las glucosiltransferasas dependen de la cepa bacteriana productora.

En el conjunto de este texto, se denominará de forma genérica glucosiltransferasas a las diferentes enzimas capaces de catalizar la síntesis de polímeros de glucosa a partir de sacarosa. Éstas son producidas generalmente por cepas bacterianas de tipo *Leuconostoc*, *Lactococcus*, *Streptococcus* o *Neisseria*. El tamaño y la estructura de los glucanos producidos dependen de la cepa productora.

Las unidades de glucosa se acoplan mediante enlaces osídicos  $\alpha(1\rightarrow6)$  como en el dextrano, mediante enlaces  $\alpha(1\rightarrow3)$ , como en el caso del mutano, o mediante una alternancia de los dos tipos (alternano).

De la misma forma, la existencia y la naturaleza de las ramificaciones, su longitud y su posición varían de acuerdo con el origen de la cepa productora.

Las glucosiltransferasas que producen glucanos o GOS que contienen al menos un 50 % de enlaces  $\alpha(1\rightarrow6)$  se denominan dextrano-sacarosas. Los GOS sintetizados por estas enzimas portan, llegado el caso, ramificaciones  $\alpha(1\rightarrow2)$ ,  $\alpha(1\rightarrow3)$  y/o  $\alpha(1\rightarrow4)$ . Dichas dextrano-sacarosas son producidas concretamente por bacterias de tipo *Leuconostoc mesenteroides*.

d) La dextrano-sacarasa de *L. mesenteroides* NRRL B-1299 tiene la particularidad de producir, por su parte, un dextrano altamente ramificado cuya mayoría de ramificaciones son de tipo  $\alpha(1\rightarrow2)$ . Utilizada en presencia de sacarosa y de maltosa, molécula aceptora de glucosa, conduce a la formación de GOS que presentan, para algunos, un enlace  $\alpha(1\rightarrow2)$  en su extremo no reductor y, para otros, ramificaciones  $\alpha(1\rightarrow2)$  en los residuos intermedios entre los extremos. Por este motivo, resisten a la degradación por las enzimas (hidrolasas) del tracto digestivo superior, en ser humano y en animales, y solamente son degradados por géneros bacterianos capaces de fermentar tales como *Bacteroides* y *Bifidobacterium*, considerados beneficiosos para el organismo del huésped.

Un fenómeno idéntico se produce a nivel de la piel, permitiendo prever aplicaciones en cosmetología, ya que es el desequilibrio de la flora microbiana cutánea el que está en el origen de numerosos problemas cosméticos y dermatológicos. Es debido a estas características que se designan en este contexto mediante el término GOS de interés.

En el conjunto del texto, los polisacáridos sintetizados por las glucosiltransferasas descritas en este contexto son, bien dextranos de alto peso molecular cuando la reacción se realiza sin aceptor de glucosa, o bien oligósidos cuando la reacción se realiza en presencia de aceptor de glucosa tal como maltosa o isomaltosa sin que esto se especifique necesariamente. En efecto, la funcionalidad de la enzima se caracteriza por la naturaleza de los enlaces glucosa-glucosa [ $\alpha(1\rightarrow6)$ ,  $\alpha(1\rightarrow2)$ ] u otros y no por el peso molecular del polisacárido sintetizado.

Las dextrano-sacarosas de *L. mesenteroides* encuentran ya numerosas aplicaciones en la industria, y en particular las de la cepa NRRL B-1299 para las cuales un procedimiento de síntesis de GOS que presentan ramificaciones  $\alpha(1\rightarrow2)$  se ha descrito en la patente EP 0325 872 B1.

Marguerite Dols et al. (2) han mostrado que los GOS producidos por las dextrano-sacarosas de esta cepa son, de hecho, una mezcla de al menos tres familias de moléculas similares que difieren, de hecho, en el número y el posicionamiento de las ramificaciones de tipo  $\alpha(1\rightarrow2)$ , lo que conlleva la hipótesis de la existencia de diferentes actividades enzimáticas de tipo glucosiltransferasa en esta cepa bacteriana.

Teniendo en cuenta el interés industrial en el campo de los alimentos prebióticos, en cosmetología o en farmacia de los GOS que presentan ramificaciones  $\alpha(1\rightarrow2)$  y recordado anteriormente, se ha previsto aislar y caracterizar una enzima particular entre las producidas por *L. mesenteroides* NRRL B-1299 que estaría más particularmente implicada en la síntesis de oligósidos que presentan las ramificaciones  $\alpha(1\rightarrow2)$ . La identificación y la caracterización de dicha enzima ofrecen la ventaja, por un lado, de proporcionar un procedimiento de producción uniforme y reproducible de los GOS de interés y, por otro lado, de identificar las características esenciales de la enzima productora de estos GOS de interés, para, llegado el caso, mejorar las prestaciones de los productos de la reacción enzimática en función de la utilización prevista.

El problema técnico sustentado por los desarrollos descritos en este contexto era, de este modo, poder disponer de una enzima y por lo tanto, ácidos nucleicos aislados que codifican esta enzima que permiten la producción mejorada de GOS con ramificaciones  $\alpha(1\rightarrow2)$ .

La presente solicitud aporta una solución técnica a las diferentes cuestiones mencionadas anteriormente, proporcionando, según un aspecto de la invención, una nueva dextrano-sacarasa, denominada DSR-E codificada por un gen dotado de una estructura nueva e inesperada (*dsrE*) y capaz de catalizar la síntesis de glucanos o de oligosacáridos que comprenden ramificaciones  $\alpha(1\rightarrow2)$ .

Entre la fecha de depósito de la solicitud de patente francesa N°0103631, cuya prioridad se reivindica y en la que la dextrano-sacarasa se denominaba DSR-D, y la de la presente solicitud, otra dextrano-sacarasa, diferente de la enzima objeto de la solicitud prioritaria, se ha descrito y se ha denominado también DSR-D. Por esto, en el marco de

la presente descripción, la dextrano-sacarasa descrita, reivindicada en un modo particular de realización y representada en la figura 1b) ya no se denomina DSR-D como en el documento de prioridad, sino DSR-E. De hecho, las dextrano-sacarosas DSR-D de acuerdo con dicho documento de prioridad y DSR-E son completamente idénticas.

5 Por estructura nueva e inesperada, se entiende el hecho de que la organización de la proteína difiere de la de todas las demás glucosiltransferasas descritas actualmente (1) y cuyo dominio catalítico está situado cadena arriba de un dominio de unión a glucano, constituyendo este último la parte carboxílica de la proteína.

10 De este modo, la presente solicitud divulga un polipéptido aislado que tiene una actividad enzimática de glucosiltransferasa capaz de formar dextranos que presentan ramificaciones  $\alpha(1\rightarrow2)$ , caracterizado por que comprende al menos un dominio de unión a glucano y un dominio con actividad catalítica situado cadena abajo del dominio de unión a glucano. Por situado cadena abajo, se entiende el hecho de que la parte aminada de la secuencia con actividad catalítica o dominio catalítico es proximal a la parte carboxílica del dominio de unión a glucano. Estos dos dominios pueden ser inmediatamente contiguos o, por el contrario, estar separados por una región variable.

La glucosiltransferasa según la invención consta preferentemente de un péptido señal.

20 Según la invención, la glucosiltransferasa comprende dos dominios catalíticos situados a uno y otro lado del dominio de unión a glucano.

25 La presencia de un dominio con actividad catalítica en la parte carboxílica de la enzima es una característica esencial de esta última en su capacidad para formar enlaces  $\alpha(1\rightarrow2)$  osídicos. En efecto, como muestran los experimentos descritos a continuación, la delección de este dominio en una enzima que tiene al menos dos dominios catalíticos conduce a la producción de glucanos o de oligósidos que tienen esencialmente enlaces osídicos de tipo  $\alpha(1\rightarrow6)$  y desprovistos de enlaces de tipo  $\alpha(1\rightarrow2)$ .

30 Más exactamente, el dominio catalítico, en cuanto está situado cadena abajo de un dominio de unión a glucano, permite la síntesis de oligósidos que contienen enlaces  $\alpha(1\rightarrow2)$ .

Además, los experimentos descritos a continuación demuestran que la especificidad de función de la dextrano-sacarasa DSR-E, a saber su capacidad para catalizar la formación de enlaces osídicos  $\alpha(1\rightarrow2)$  es imputable, no a la presencia concomitante de dos dominios catalíticos, sino más bien a la concatenación de un dominio de unión a glucano y de un dominio catalítico, y más particularmente del dominio catalítico CD2.

El análisis comparativo de las diferentes glucosiltransferasas que incluyen las dextrano-sacarosas puso de manifiesto un muy fuerte grado de conservación de su dominio catalítico.

40 El dominio catalítico situado en la parte carboxi-terminal de la glucosiltransferasa descrita en este contexto tiene una secuencia que presenta al menos un 44 % de identidad y un 55 % de similitud con los dominios catalíticos de las otras glucosiltransferasas analizadas. En particular, el dominio catalítico en la parte carboxílica de la glucosiltransferasa según la invención tiene al menos un 65 % de identidad y al menos un 80 % de similitud con la secuencia ID No. 1, estando la tríada catalítica Asp/Glu/Asp en posiciones respectivas 230/268/342 conservada.

45 En el conjunto del texto, se entiende por X % de similitud con respecto a una secuencia de referencia, el hecho de que un X % de los aminoácidos son idénticos o están modificados por sustitución conservativa tal como se define en el software de alineamiento de las secuencias de aminoácidos ClustalW (<http://bioweb.pasteur.fr/docs/doc-gensoft/clustalw/>) y que (100-X) % pueden deleccionarse, sustituirse por otros aminoácidos, o incluso que (100-X) % pueden añadirse a la secuencia de referencia. Una estructura primaria particular de la enzima de acuerdo con la presente descripción se representa en la secuencia ID No. 2 que representa una secuencia de 2835 aminoácidos de una dextrano-sacarasa de *L. mesenteroides* NRRL B-1299.

55 Esta dextrano-sacarasa, denominada DSR-E, posee como la mayor parte de las glucosiltransferasas y de las dextrano-sacarosas una secuencia señal, una región variable poco conservada, un dominio catalítico altamente conservado (CD1), un dominio de unión a glucano (GBD) y un segundo dominio catalítico (CD2) en la parte carboxílica de la proteína. DSR-E es la primera glucosiltransferasa analizada y que presenta dos dominios catalíticos, en la configuración presentada en la figura 1 b). Es también la primera glucosiltransferasa de la que un dominio catalítico está situado en la parte carboxílica de la proteína.

60 La figura 1b) hace que se manifieste también que el dominio de unión a glucano es sensiblemente más largo que el descrito anteriormente para las dextrano sacarosas conocidas; de este modo, otra característica de las enzimas según la invención es el tamaño de este dominio que es superior a 500 aminoácidos.

65 La comparación y el análisis de la secuencia de DSR-E con las secuencias de las glucosiltransferasas o de las dextrano-sacarosas ya descritas (1), así como los medios utilizados para este fin se indican en el ejemplo 2 detallado

a continuación. Aparece claramente que si la existencia de dos dominios catalíticos diferencia sustancialmente DSR-E de las otras enzimas, en cambio las secuencias de dichos dominios están sustancialmente conservadas. En particular, los aminoácidos necesarios para la actividad catalítica se conservan en el segundo dominio catalítico, a saber la tríada Asp/Glu/Asp situada en las posiciones respectivas 2210/2248/2322 de la secuencia ID No. 2.

5 La presente invención se refiere más particularmente a un polipéptido aislado tal como se define en la reivindicación 1.

10 Según otro aspecto, la presente solicitud divulga también cualquier polipéptido aislado que tiene una actividad catalítica de glucosiltransferasa capaz de formar dextranos u oligosacáridos que tienen ramificaciones  $\alpha(1\rightarrow2)$  tal como obtenido por modificación, sustitución, inserción o delección de secuencias de aminoácidos pero que consta de secuencias que presentan al menos un 80 % y preferentemente al menos un 90 % de similitud con las secuencias siguientes de la secuencia ID No. 2:

423 - 439	2120 - 2138
478 - 501	2161 - 2184
519 - 539	2202 - 2214
560 - 571	2243 - 2250
631 - 645	2315 - 2322
1014 - 1021	2689 - 2696

15 De acuerdo con un aspecto particular, finalmente, un polipéptido con actividad catalítica contiene los siguientes aminoácidos:

20 W en posiciones 425 y 2122,  
E en posiciones 430, 565 y 2127, 2248,  
D en posiciones 487, 489, 527, 638, 2170, 2172, 2210 y 2322,  
H en posición 637 y 2321,  
Q en posición 1019 y 2694.

25 Según un modo particular de realización, la invención se refiere también a un polipéptido aislado tal como se define en la reivindicación 2.

30 Los polipéptidos con actividad de glucosiltransferasas capaces de formar enlaces osídicos  $\alpha(1\rightarrow2)$  pueden presentarse en forma aislada, o por el contrario integrarse en una proteína más amplia, como por ejemplo una proteína de fusión. En efecto, puede ser ventajoso incluir secuencias que presentan otra función, como por ejemplo una secuencia etiqueta específico de un ligando que permite facilitar su purificación. Estas secuencias etiqueta pueden ser de tipo GST (glutación-S-transferasa), Inteína - CBD (Chitine-Binding Domain), (comercializado por New England Biolabs, <http://www.neb.com>), MBD (Maltose Binding Domain), polipéptidos que contienen residuos histidina contiguos que permiten facilitar la purificación del polipéptido con el que está fusionado. El experto en la materia puede concebir cualquier otra proteína de fusión que permite asociar la función de la DSR-E descrita en este contexto y, según un aspecto particular, reivindicada, con otra función, como por ejemplo, y sin ser limitante, una secuencia que aumenta la estabilidad de la enzima producida por expresión en un huésped recombinante o una secuencia capaz de aumentar la especificidad o la eficacia de acción de esta enzima, o una secuencia que pretende asociar otra actividad enzimática afin.

40 Dichas proteínas de fusión forman también parte de la invención en cuanto contienen el dominio CD2 y el sitio de unión al glucano. De la misma forma, los fragmentos de la secuencia ID No. 2, que comprenden al menos la secuencia ID No. 1 y el dominio de unión a glucano, solos o integrados en una secuencia polipeptídica más amplia forman parte de la invención, a partir del momento en que la actividad enzimática de dextrano-sacarasa se conserva.

45 Las variantes de las secuencias polipeptídicas definidas anteriormente forman también parte de la invención. Además de los polipéptidos obtenidos por sustitución conservativa de los aminoácidos tal como se ha definido anteriormente, las variantes incluyen polipéptidos cuya actividad enzimática está mejorada por ejemplo por mutagénesis dirigida o aleatoria, por evolución molecular, o por duplicación del dominio catalítico CD2.

50 La estructura particular de esta enzima identificada en la presente solicitud resulta de un proceso que comprende:

- 55 a) la identificación y el aislamiento de la dextrano-sacarasa de *L. mesenteroides* que cataliza la producción de GOS de interés que portan las ramificaciones  $\alpha(1\rightarrow2)$ ;  
b) la secuenciación de fragmentos de la enzima;  
c) la síntesis de cebadores de amplificación capaces de amplificar el gen correspondiente de la cepa productora o de fragmentos de ésta;  
d) la secuenciación de los fragmentos amplificados;  
e) la clonación en vectores específicos y su expresión en huéspedes apropiados.

Las modalidades del procedimiento implementado se detallan en la parte experimental a continuación. La primera etapa consiste en una separación de las proteínas por electroforesis en gel de poliacrilamida, e identificación de las bandas que presentan la actividad de dextrano sacarasa mediante una reacción enzimática *in situ* en presencia de sustrato y de aceptor. La naturaleza de los GOS sintetizados se identifica a continuación en cada banda mediante análisis por HPLC de acuerdo con los métodos descritos en (1). El tiempo de retención de los oligósidos en HPLC depende de la naturaleza y de la organización de sus enlaces osídicos. Es posible, en particular, distinguir aquellos constituidos por residuos enlazados en  $\alpha(1\rightarrow6)$ , en  $\alpha(1\rightarrow6)$  con una ramificación  $\alpha(1\rightarrow2)$  en el extremo no reductor de la molécula, y los buscados compuestos por una cadena lineal  $\alpha(1\rightarrow6)$  con ramificaciones  $\alpha(1\rightarrow2)$ .

Los inventores han, por lo tanto, aislado e identificado la dextrano-sacarasa de *L. mesenteroides* NRRL B-1299 productora de GOS de interés.

Un procedimiento de ingeniería inversa implementado en las etapas b) a e) anteriores permitió a continuación proporcionar la secuencia nucleotídica que codifica la enzima y que permite producirla en cantidad industrial y, llegado el caso, modificarla, mejorando sus prestaciones mediante las técnicas a disposición del experto en la materia. A modo de ejemplo, se pueden citar mutagénesis dirigida o aleatoria, o la evolución molecular (DNA shuffling) (3).

Otro aspecto de la presente solicitud se refiere a una molécula de ácido nucleico aislada que codifica una enzima con actividad glucosiltransferasa capaz de formar dextranos u oligósidos que presentan ramificaciones  $\alpha(1\rightarrow2)$  y que comprenden al menos una secuencia que codifica un dominio de unión a glucano, y al menos una secuencia nucleotídica que codifica un dominio catalítico situado en 3' respecto a la anterior, codificando dicha secuencia un dominio catalítico que tiene al menos un 50 % y preferentemente al menos un 70 % de similitud con la secuencia ID No. 3 que tiene la estructura presentada en la figura 1 b.

Por similitud, se entiende el hecho de que, para un mismo marco de lectura, un triplete dado se traduce por el mismo aminoácido. Este término incluye, por lo tanto, las modificaciones de bases que resultan de la degeneración del código genético.

El porcentaje de similitud se determina comparando una secuencia dada con la secuencia de referencia. Cuando éstas son de longitudes diferentes, el porcentaje de similitud está basado en el porcentaje de nucleótidos de la secuencia más corta similares a los de la secuencia más larga.

El grado de similitud puede determinarse convencionalmente mediante utilización de softwares tales como ClustalW (Thompson et al., Nucleic Acid Research (1994), 22: 4673-4680) distribuidos por Julie Thompson ([Thompson@EMBL-Heidelberg.DE](mailto:Thompson@EMBL-Heidelberg.DE)) y Toby Gibson ([Gibson@EMBL-Heidelberg.DE](mailto:Gibson@EMBL-Heidelberg.DE)) del Laboratorio Europeo de Biología Molecular, Meyerhofstrasse 1, D-69117 Heidelberg, Alemania. ClustalW también puede cargarse desde varios sitios web incluyendo IGBMC (Instituto de Genética y de Biología Molecular y Celular, B.P. 163, 67404 Illkirch cedex Francia; <ftp://ftp-igbmc.u-strabg.fr/pub/>) y EBI (<ftp://ftp.ebi.ac.uk/pub/software/>) y todos los sitios que reenvían al Instituto de Bioinformática (Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, RU).

Los ácidos nucleicos aislados descritos en el presente contexto pueden comprender concretamente otras secuencias destinadas a mejorar la expresión y/o la actividad de la enzima producida.

Puede tratarse, a modo de ejemplo:

- de secuencias que codifican una secuencia señal para su secreción;
- de una duplicación de la secuencia que codifica el dominio catalítico CD2.

Preferentemente, un ácido nucleico aislado según la invención comprende:

- a) dos secuencias que codifican dominios catalíticos que tienen al menos un 50 %, y preferentemente al menos un 80 % de similitud con la secuencia ID n° 3 que tiene la estructura presentada en la figura 1b;
- b) una secuencia que codifica el dominio de unión a glucano, estando este último situado preferentemente entre las dos secuencias en a).

Un ácido nucleico tal como se describe en este contexto podrá comprender además:

- un promotor, capaz de su expresión en una célula huésped seleccionada,
- una secuencia que codifica un péptido señal, y/o
- una o varias secuencias variables,

estando esta o estas secuencias todas situadas en la parte 5' de las secuencias que codifican el o los dominios catalíticos.

Un ejemplo particular de un ácido nucleico aislado comprende más particularmente:

- a) la secuencia ID No. 4,
- b) una secuencia que presenta al menos un 80 % de similitud con la secuencia ID n° 4 que tiene la estructura presentada en la figura 1b, o
- c) la cadena complementaria de la secuencia a) o b), o
- d) una secuencia que hibrida con a), b) o c).

La hibridación en d) se realiza en condiciones estándar, y preferentemente en condiciones rigurosas. Por hibridación en condiciones rigurosas, se entiende el hecho de que existe una identidad de secuencias de al menos un 80 % de la secuencia que se busca hibridar y preferentemente una identidad de al menos un 90 % de la secuencia que se busca hibridar, en condiciones descritas por ejemplo en Sambrook y Russel (3ª edición, 2001, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY).

La invención se refiere más particularmente a un ácido nucleico aislado tal como se define en las reivindicaciones 3 y 4.

La solicitud también se refiere a un gen que codifica una dextrano-sacarasa capaz de formar al menos un 15 % de ramificaciones  $\alpha(1\rightarrow2)$ . Además de la secuencia codificante, el gen comprende las secuencias que permiten el inicio de la transcripción así como las secuencias que permiten la fijación de ARN mensajero al ribosoma (RBS). La secuencia ID No. 5 representa una estructura del gen tal como se aísla de *L. mesenteroides* NRRL B-1299.

Los nucleótidos cadena arriba de la ATG de inicio de la traducción se numeran de 1 a 232.

Se puede identificar la existencia de una secuencia RBS entre los nucleótidos 218 y 223, así como las secuencias consenso - 35 y - 10 situadas entre los nucleótidos 82 y 86 (TTGAA), por un lado y 100 y 105 (ATAAAT), por el otro.

Cualquier secuencia de ácido nucleico hibridable con el ADN de la secuencia ID No. 4 o su cadena complementaria es susceptible de codificar una enzima que tiene las propiedades y características de la enzima descrita en este contexto. Esto se aplica tanto a las secuencias naturales que existen en microorganismos diferentes de *L. mesenteroides* NRRL B-1299 y aisladas de bancos genómicos de microorganismos, como a las preparadas mediante ingeniería genética o por síntesis química.

En particular, las secuencias cadena arriba del ATG de inicio de la traducción y necesarias para la expresión de la proteína pueden sustituirse ventajosamente por secuencias de inicio de la transcripción y/o de fijación al ribosoma adaptados al sistema de expresión seleccionado por la secuencia codificante.

Una secuencia de ácidos nucleicos susceptible de hibridar en condiciones rigurosas con el ácido nucleico aislado de acuerdo con la presente solicitud comprende también fragmentos, derivados, o variantes alélicas de la secuencia de ácidos nucleicos de la presente solicitud que codifica una proteína que tiene la actividad enzimática descrita anteriormente. De este modo, los fragmentos se definen como fragmentos de moléculas de ácidos nucleicos suficientemente largos para codificar una proteína que tenga conservada su actividad enzimática. Ésta incluye también fragmentos desprovistos de la secuencia que codifica el péptido señal responsable de la secreción de la proteína.

El término "derivado" significa secuencia, diferente de la secuencia original, en una o varias posiciones, pero que presenta un alto grado de similitud con estas secuencias. En este contexto, similitud significa una identidad de al menos un 80 % de los nucleótidos, y preferentemente de al menos un 90 % con la secuencia original. Las modificaciones en este caso se refieren a deleciones, sustituciones, inserciones o recombinaciones, a partir del momento en que la enzima codificada por estas secuencias homólogas presenta la actividad enzimática se los polipéptidos descritos en este contexto.

Las secuencias de ácidos nucleicos de acuerdo con la presente solicitud tal como se han descrito anteriormente y calificadas como derivados de estas moléculas tal como se han descrito anteriormente, son generalmente variantes que ejercen la misma función biológica. Estas variaciones pueden ser variaciones naturales, concretamente aquellas observables de una especie a otra y que resultan de una variabilidad interespecífica o, por el contrario, ser introducidas por medio de una mutagénesis dirigida, aleatoria o mediante DNA Shuffling (evolución molecular).

De la misma forma, forman parte de la presente divulgación los ácidos nucleicos aislados que codifican una glucosiltransferasa capaz de catalizar la síntesis de dextrano o de oligosacárido que portan al menos un 20 % y preferentemente al menos un 30 % de ramificaciones de tipo  $\alpha(1\rightarrow2)$  y obtenidos por evolución molecular (DNA shuffling) y que comprenden:

- una etapa de modificación aleatoria de una de las secuencias descritas anteriormente y, en particular, secuencias ID No. 3 y 4 y de establecimiento de variantes;
- una etapa de expresión de estas secuencias modificadas en una célula huésped apropiada, un huésped que alberga una variante;

- una etapa de cribado de los huéspedes que expresan una enzima capaz de formar más del 20 % y preferentemente más del 30 % de enlaces  $\alpha(1 \rightarrow 2)$  en un sustrato apropiado y una etapa de aislamiento del o de los genes mejorados.

5 Según otro aspecto de la invención, esta última se refiere también a un proceso tal como se define en la reivindicación 11.

Un ácido nucleico aislado de acuerdo con la presente solicitud también podrá comprender:

- 10 a) una secuencia que tiene al menos un 80 % de similitud con la secuencia que codifica una dextrano-sacarasa expresada por el plásmido pCR-T7-dsrE en *E. coli* depositado en la CNCM el 15 de marzo de 2001 con el número I-2649 (*E. coli* JM 109 [pCR-T7-dsrID]), o
- b) una secuencia complementaria de la secuencia en a).

15 La denominación de la cepa transformada por el plásmido recombinante pCR-T7-dsrE y depositada en la CNCM es la indicada anteriormente entre paréntesis. Ésta no afecta al cambio de denominación del gen operado posteriormente al depósito de dicha cepa por las razones mencionadas anteriormente.

20 La solicitud divulga también los fragmentos de ácidos nucleicos, tal como se han definido anteriormente, hibridables con la secuencia ID No. 4, y utilizables como sondas de hibridación para la detección de secuencias que codifican enzimas descritas en este contexto. Estos fragmentos pueden prepararse mediante todas las técnicas conocidas por el experto en la materia.

25 Además de las sondas de hibridación, cebadores de amplificación forman también parte de la presente solicitud. Dichos cebadores son fragmentos hibridables con la SEQ ID No. 4 o con su cadena complementaria y permiten la amplificación de secuencias específicas que codifican dextrano-sacarosas presentes en un organismo procarionta o eucariota, animal o vegetal.

30 La utilización de dichos cebadores de amplificación permite la implementación de un procedimiento de identificación de la existencia eventual de un gen que codifica una enzima capaz de catalizar la síntesis de GOS con ramificaciones  $\alpha(1 \rightarrow 2)$  en dicho organismo, formando dicho procedimiento también parte de la presente divulgación.

35 También se divulgan vectores de expresión que comprenden un ácido nucleico tal como se ha descrito anteriormente, bajo el control de secuencia que permite su expresión y, preferentemente, su excreción en células procariontas o eucariotas. Como células procariontas, se seleccionarán preferentemente bacterias seleccionadas entre un grupo que comprende *E. coli*, *Lactococcus*, *Bacillus*, *Leuconostoc*. Como células eucariotas, se seleccionarán preferentemente las eucariotas seleccionadas entre un grupo que contiene levaduras, hongos o plantas.

40 El vector comprende un promotor adaptado a la expresión del ácido nucleico aislado de acuerdo con la presente descripción en el sistema de expresión seleccionado. A modo de ejemplo, el promotor del bacteriófago T7 podría seleccionarse ventajosamente para una expresión en *E. coli*. La invención se refiere más particularmente a un vector tal como se define en la reivindicación 5 o 6.

45 La solicitud divulga también células huésped, procariontas o eucariotas, transformadas con un ácido nucleico tal como se ha descrito en este contexto preferentemente comprendido en un vector de expresión que porta un promotor, adaptado para una expresión en las células huésped seleccionadas. Las células transformadas se seleccionan entre el grupo de las bacterias Gram- tales como *E. coli*, o entre el grupo de las bacterias Gram+ tales como *Lactococcus*, *Bacillus*, *Leuconostoc* o entre las eucariotas en un grupo que comprende levaduras u hongos, o plantas.

50 Un ejemplo particular de una célula transformada de acuerdo con la presente descripción es la cepa de *E. coli* que alberga un plásmido denominado PCR-T7dsrE portador de la secuencia ID No. 4 bajo el control del promotor del bacteriófago T7 y depositada en la CNCM el 15 de marzo de 2001 con el número I-2649. La invención se refiere más particularmente a células tal como se definen en las reivindicaciones 7 u 8.

55 La presente solicitud, por otro lado, divulga un procedimiento de producción de una glucosiltransferasa capaz de formar dextranos u oligósidos que presentan al menos un 15 % y preferentemente al menos un 20 % de ramificaciones de tipo  $\alpha(1 \rightarrow 2)$  osídicas y que comprende:

- 60 a) la inserción de un ácido nucleico o de un vector tal como se ha descrito anteriormente en una célula huésped capaz de expresar y preferentemente de secretar glucosiltransferasa;
- b) la caracterización de la actividad enzimática buscada mediante todos los métodos accesibles para el experto en la materia;
- c) la purificación de la enzima a partir de un extracto celular.

65 Según otros aspectos de la invención, esta última se refiere también a un proceso tal como se define en la reivindicación 9 o 10.

Mediante método de caracterización de la actividad enzimática conocida por el experto en la materia, se entenderán los métodos descritos en la bibliografía, por ejemplo en la referencia (2), así como nuevos métodos susceptibles de ser desarrollados que permitan identificar y discriminar los glucooligosacáridos que presentan la tasa de ramificación buscada.

Se trata, de hecho, de cualquier procedimiento de cribado que permita identificar la presencia de ramificaciones  $\alpha(1\rightarrow 2)$  en un GOS.

A modo de ejemplo, se utilizarán:

- HPLC para la cual la migración de los GOS varía en función de la naturaleza y el posicionamiento de las ramificaciones, concretamente aquellas que tienen el enlace  $\alpha(1\rightarrow 2)$  en el extremo reductor y aquellas que tienen este enlace en la penúltima glucosa, y/o
- resonancia magnética nuclear (RMN),
- la existencia de una reacción positiva con anticuerpos monoclonales específicos de los enlaces  $\alpha(1\rightarrow 2)$  en el extremo reductor y/o de anticuerpos monoclonales específicos de los enlaces  $\alpha(1\rightarrow 2)$  en la penúltima glucosa del GOS.

La solicitud divulga también un procedimiento de obtención de una glucosiltransferasa capaz de presentar oligósidos o dextranos que presentan una tasa de ramificación  $\alpha(1\rightarrow 2)$  superior al 15 % y preferentemente superior al 30 % de la totalidad de los enlaces osídicos y que comprende una etapa de modificación de la secuencia ID No. 4 por adición, delección, mutación a partir del momento en que:

- el marco de lectura no está modificado, y
- los siguientes aminoácidos se conservan después de la traducción:

W en posiciones 425 o 2122, codificado por el triplete TGG en posiciones 1273 y 6364,

E en posiciones 430, 565, 2127 y 2248 codificados por los tripletes GAA en posiciones 1288, 1693, 6379 y 6742 respectivamente,

D en posiciones 487, 489, 527, 638, 2170 y 2210 codificados por los tripletes GAT en posiciones 1459, 1465, 1579, 1912, 6508 y 6628 respectivamente,

D en posiciones 2172 y 2322 codificados por los tripletes GAT en posiciones 6514 y 6964,

H en posición 637 y 2321, codificados respectivamente por los tripletes CAT en posición 1909 y CAC en posición 6961,

Q en posiciones 1019 y 2694 codificados respectivamente por los tripletes CAA (posición 3055) y CAG (posición 8080).

Un procedimiento de producción de una glucosiltransferasa de acuerdo con la presente solicitud que tiene las mismas características que anteriormente también puede comprender:

- una etapa de modificación aleatoria de la secuencia ID n° 4 y de establecimiento de un banco de variantes,
- una etapa de expresión de estas secuencias modificadas en una célula huésped apropiada, un huésped que alberga una variante,
- una etapa de cribado de huésped que expresan una enzima capaz de formar más del 15 % y preferentemente más del 30 % de enlace  $\alpha(1 \rightarrow 2)$  en un sustrato apropiado,
- una etapa de aislamiento del o de los genes mejorados.

En otra realización, el procedimiento consiste en modificar la secuencia ID No. 3 mediante duplicación de todo o parte del dominio catalítico CD2.

Se podrá entender que los procedimientos anteriores pretenden, no solamente la obtención de una glucosiltransferasa capaz de formar oligósidos que presenten una tasa de ramificación  $\alpha(1 \rightarrow 2)$  constante y reproducible, superior al 15 % de las ramificaciones totales, sino también mejorar la tasa de ramificación  $\alpha(1 \rightarrow 2)$  con el objetivo de modificar las propiedades de los oligósidos obtenidos en el sentido de una mejora de sus propiedades dietéticas o de su capacidad para mantener o restablecer la flora bacteriana asociada a ciertos órganos del cuerpo humano o animal.

La presente solicitud divulga finalmente glucosiltransferasas susceptibles de ser obtenidas mediante un procedimiento citado anteriormente y capaces de formar al menos un 15 % y preferentemente al menos un 30 % de ramificaciones de tipo  $\alpha(1 \rightarrow 2)$  osídicas en glucooligosacáridos.

La solicitud divulga finalmente la utilización de las glucosiltransferasas descritas en este contexto así como las susceptibles de ser obtenidas mediante los procedimientos anteriores, en la fabricación de una composición con efecto prebiótico o en la fabricación de una composición dermatológica, cosmética o farmacéutica. La invención se refiere, más particularmente, a la utilización definida en la reivindicación 12.

A modo de ejemplos no limitantes, se puede citar la mejora del tránsito intestinal en animales y en el ser humano, la mejora de la asimilación del calcio y/o del magnesio y de los minerales en general, la prevención del cáncer de colon, la prevención o el tratamiento de las afecciones de la piel tales como acné, caspa, olores corporales.

La ventaja de los polipéptidos y de los ácidos nucleicos que codifican estos polipéptidos descritos en este contexto se sitúa no solamente al nivel de la mejora en términos de calidad, de rendimiento, de reproducibilidad, y de precio de coste de las glucosiltransferasas capaces de formar oligosacáridos con ramificaciones de tipo  $\alpha(1 \rightarrow 2)$  osídicas pero también en la perspectiva de producir nuevas enzimas cuya funcionalidad esté mejorada.

Las figuras, ejemplos y la descripción detallada a continuación permiten, sin limitarla, ilustrar las características y las funcionalidades particulares de los polipéptidos con actividad enzimática y de las secuencias que los codifican. Éstas permiten en particular ilustrar de forma más precisas la especificidad del dominio catalítico presente en la parte carboxílica de la enzima descrita en este contexto o según la invención y su evolución potencial para la obtención de enzimas mejoradas.

Leyenda de las figuras:

Figura 1: estructura de las glucosiltransferasas nativas y de las proteínas recombinantes derivadas: la figura 1a) representa la estructura de las glucosiltransferasas y de las dextrano-sacarosas descritas en la bibliografía (1). PS: péptido señal; ZV: región variable, CD: dominio catalítico, GBD: dominio de unión a glucano. La figura 1b) representa la estructura de la glucosiltransferasa según la invención. Las figuras 1c) a 1i) representan diferentes construcciones que constan de las deleciones en comparación con la proteína DSR-E nativa.  $\Delta$ (PS) corresponde al control constituido por la forma completa clonada en el sistema pBAD-TOPO Thiofusion (Invitrogen).

Figura 2: esquema resumen del método de clonación de la secuencia nucleotídica que codifica una glucosiltransferasa de acuerdo con la presente solicitud con ayuda de una biblioteca genómica utilizando una sonda de PCR descrita en la tabla I y una sonda *HindIII/EcoRV* respectivamente.

Figura 3: comparación de las secuencias señal de diferentes glucosiltransferasas de *L. mesenteroides*. Los aminoácidos conservados están en negrita. DSR-B: *L. mesenteroides* NRRL B-1299 (4); DSR-S: *L. mesenteroides* NRRL B-512F (5); ASR: *L. mesenteroides* NRRL B-1355 (6).

Figura 4: alineamiento de las 11 secuencias repetidas de la enzima DSR-E y observadas en la zona variable.

Figura 5: alineamiento de las secuencias conservadas del dominio catalítico.

- Bloque A: aminoácidos esenciales de la parte N-terminal del dominio catalítico;
- Bloque B: aminoácidos de la parte del dominio catalítico de unión a sacarosa;
- Bloques C, D, E: bloques que contienen los tres residuos de aminoácidos implicados en la tríada catalítica (6);
- Bloque F: secuencia que contiene la glutamina 937 de GTF-I estudiada por Monchois et al. (7).

Los aminoácidos conservados completamente se indican en negrita. «\*»: sustituciones conservativas; «:»: sustituciones semiconservativas; —: GAP. Las numeraciones son las de la secuencia ID No. 2.

Figura 6: caracterización por HPLC de los productos sintetizados por la enzima recombinante DSR-E.

6A: análisis por HPLC de los glucooligosacáridos obtenidos con las dextrano-sacarosas de *L. mesenteroides* NRRL B-1299.

6B: análisis por HPLC de los glucooligosacáridos obtenidos mediante DSR-E recombinante. La identificación de los diferentes picos siguientes:

- 1: fructosa,
- 2: maltosa,
- 3: sacarosa,
- 4: panosa,
- 5: R4,
- 6: OD4,
- 7: R5,
- 8: OD5,
- A, B, C: picos no identificados.

6C: DSR-E recombinante delecionada del dominio catalítico de la parte carboxílica de la enzima ( $\Delta$  DSR-E).

Figura 7: análisis por HPLC de los productos de la reacción de aceptor en maltosa, sintetizados por las diferentes formas completas y delecionadas de la proteína DSR-E.

L. m. B-1299: mezcla de dextrano-sacarosas producida por *L. mesenteroides* NRRL B-1299.

La identificación de los diferentes picos se realiza de la siguiente manera:

F: fructosa

M: maltosa  
 S: sacarosa  
 P: panosa  
 R4, R5: GOS que constan de los enlaces  $\alpha(1\rightarrow2)$   
 OD4, OD5: GOS desprovistos de enlaces  $\alpha(1\rightarrow2)$ .

Materiales y métodos:

1) Cepas bacterianas, plásmidos y condiciones de crecimiento:

Todas las cepas se conservan a  $-80\text{ }^{\circ}\text{C}$  en tubos que contienen un 15 % de glicerol (v/v).

*Leuconostoc mesenteroides* B-1299 (NRRL, Peoria, EE. UU.) se cultiva a  $27\text{ }^{\circ}\text{C}$ , con agitación (200 RPM) en medio estándar (sacarosa  $40\text{ g.l}^{-1}$ , fosfato de potasio  $20\text{ g.l}^{-1}$ , extracto de levadura  $20\text{ g.l}^{-1}$ ,  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$   $0,2\text{ g.l}^{-1}$ ,  $\text{MnSO}_4\cdot \text{H}_2\text{O}$   $0,01\text{ g.l}^{-1}$ , NaCl  $0,01\text{ g.l}^{-1}$ ,  $\text{CaCl}_2$   $0,02\text{ g.l}^{-1}$ ,  $\text{FeSO}_4\cdot 7\text{H}_2\text{O}$   $0,01\text{ g.l}^{-1}$ ), ajustándose el pH a 6,9. *Escherichia coli* DH5 $\alpha$  y JM109 se cultivaron en medio LB (Luria-Bertani).

La selección de los clones recombinantes de pUC18 o pGEM-T Easy se efectúa en placas de LB-agar suplementado con  $100\text{ }\mu\text{g.ml}^{-1}$  de ampicilina,  $0,5\text{ mM}$  de isopropil- $\beta$ -D-tiogalactopiranosido (IPTG) y  $40\text{ }\mu\text{g.ml}^{-1}$  de 5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosido (X-gal). Se utilizaron células de *E. coli* TOP 10 para el sistema de clonación de producto de PCR TOPO Cloning (Invitrogen), y se cultivaron en medio LB suplementado con kanamicina a la concentración de  $50\text{ }\mu\text{g.ml}^{-1}$ .

En lo que concierne a la expresión de *dsrE*, el kit de clonación ECHO Cloning System (Invitrogen) permite la clonación de un producto de PCR en un vector donador (pUNI/V5-His-TOPO), precediendo a una etapa de recombinación con un vector aceptor adaptado (pCR-T7-E). Este sistema requiere células de *E. coli* PYR1, TOP 10 y BL21(DE3)pLysS cultivadas en medio LB suplementado con  $50\text{ }\mu\text{g.ml}^{-1}$  de kanamicina, así como con  $34\text{ }\mu\text{g.ml}^{-1}$  de cloranfenicol para la cepa BL21(DE3)pLysS.

Los plásmidos pUC18 digeridos y desfosforilados provienen de Pharmacia (Amersham Pharmacia Biotech) y han sido utilizados para la constitución de un banco de ADN genómico de *L. mesenteroides* NRRL B-1299. La clonación de producto de PCR, por su parte, necesitó el empleo de plásmido pGEM-T Easy (Promega), y para los fragmentos de más de 2 kpb, de plásmido TOPO-XL (Invitrogen).

El sistema pBAD-TOPO Thiofusion (Invitrogen), utilizado para construir las diferentes formas delecionadas de la proteína DSR-E, pone en juego el promotor araBAD, cuyos mecanismos de control implican la proteína reguladora AraC. En ausencia de inductor, es decir L-arabinosa, la proteína dimérica AraC se asocia a las estructuras reguladoras del operón y conlleva la formación de un bucle de ADN, bloqueando dicho bucle de este modo la transcripción de los genes colocados bajo el control del promotor araBAD. En presencia de L-arabinosa en cambio, AraC forma un complejo, lo que libera el bucle de ADN y permite el inicio de la transcripción. La expresión basal puede limitarse añadiendo glucosa al medio de cultivo: esto actúa disminuyendo el nivel de AMP cíclico, y, por lo tanto, la activación concomitante de la proteína CAP (cAMP activator protein). El nivel de activación obtenido está en función de la concentración de L-arabinosa, de modo que las condiciones óptimas de producción de la proteína de interés pueden seleccionarse con precisión.

Además, el empleo de este vector permite situar una etiqueta de tiorredoxina de 12 kDa en el lado N-terminal de la proteína de interés. Esta fusión pretende favorecer la traducción del gen que codifica dicha proteína de interés. La proteína etiqueta permite además aumentar la solubilidad de la proteína a la que se encuentra fusionada. El sistema pBAD-TOPO Thiofusion está diseñado para permitir eliminar fácilmente la etiqueta de tiorredoxina, mediante simple escisión con ayuda de enteroquinasa. Finalmente, gracias a este sistema de expresión, se inserta una etiqueta de histidina en el lado del extremo C-terminal de la proteína de interés. Dicha etiqueta es útil para purificar dicha proteína por afinidad.

En el marco de la utilización de este sistema, la cepa *E. coli* TOP 10 se cultivó en medio LB suplementado con  $100\text{ }\mu\text{g.ml}^{-1}$  de ampicilina.

2) Electroforesis en gel, localización y caracterización de la enzima:

Después de un cultivo de *L. mesenteroides* NRRL B-1299 de 7 h, el medio se centrifugó (7000 RPM,  $4\text{ }^{\circ}\text{C}$ , 30 min) y las células, donde se encuentra el 90 % de la actividad enzimática, se concentraron 10 veces en una solución de tampón acetato (20 mM, pH 5,4), se calentaron 5 minutos a  $95\text{ }^{\circ}\text{C}$  en presencia de solución de desnaturalización (Tris HCl 62,5 mM, SDS al 4 %, urea 6 M, azul de bromofenol al 0,01 % y  $\beta$ -mercaptoetanol 200 mM). Se depositaron  $300\text{ }\mu\text{l}$  de la mezcla sobre gel de poliacrilamida al 7 %. Después de la migración, las proteínas totales se revelaron por tinción con negro amido, mientras que la actividad dextrano-sacarasa se detectó mediante tinción del polímero con reactivo de Schiff después de la síntesis de dextrano *in situ*. Las bandas que corresponden a dextrano-sacarasa activas se escindieron y se incubaron por separado en 2 ml de solución de acetato de sodio 20 mM pH

5,4 que contiene 100 g.l<sup>-1</sup> de sacarosa y 50 g.l<sup>-1</sup> de maltosa. Después del consumo total de la sacarosa, la reacción se detuvo mediante calentamiento a 95 °C durante 5 minutos, y el medio de reacción se centrifugó 5 minutos a 15000 g para eliminar el dextrano insoluble. Las muestras se analizaron mediante cromatografía en fase inversa (columna C18, Ultrasep 100, 6 µm, 5x300mm, Bishoff Chromatography) utilizando agua ultrapura como eluyente, a un caudal constante de 0,5 ml.min<sup>-1</sup>. Los oligosacáridos se separaron durante 30 minutos a temperatura ambiente, y se detectan por refractometría. La secuenciación peptídica se realizó en las bandas proteicas seleccionadas por el Laboratorio de Microsecuenciación, Instituto Pasteur, París.

3) Técnicas de biología molecular utilizadas:

La purificación del plásmido de *E. coli* y la purificación del ADN genómico de *L. mesenteroides* se realizaron utilizando respectivamente el QiaPrep Spin Plasmid kit y el Cell Culture DNA maxi kit (QiaGen). Los procedimientos de amplificación y de clonación se realizaron utilizando las técnicas estándar (Sambrook y Russel, 2001, anteriormente). Las enzimas de restricción y de modificación, que provienen de las compañías comerciales New England Biolabs o Gibco BRL, se utilizaron de acuerdo con los protocolos de los fabricantes. La PCR se realizó con cebadores seleccionados basándose en la secuencia proteica obtenida en una banda de gel de electroforesis aislada (véase anteriormente, electroforesis en gel y localización de la enzima). Se seleccionaron dos péptidos:

- 29-FYFESGK, y
- 24-FESQNNNP

y se utilizaron para sintetizar oligonucleótidos degenerados e indicados en la tabla I a continuación.

En esta tabla donde las numeraciones son las de la secuencia ID no. 4, parece que la presencia de un residuo serina en los dos péptidos necesita la síntesis de dos cebadores para cada péptido en la medida en que la serina puede estar codificada por seis codones diferentes. ECHO-dir y ECHO-inv son los cebadores utilizados que han permitido la amplificación de *dsrE* mediante PCR para su clonación en el sistema de expresión ECHO Cloning (Invitrogen).

TABLA I:

Designación	Descripción	Secuencia 5'-3'
29-dir1 29-dir2	FYFESGK	TT(C/T)TA(C/T)TT(C/T)GA(A/G)TCAGG(C/G)AA(A/G) TT(C/T)TA(C/T)TT(C/T)GA(A/G)AGCGG(C/G)AA(A/G)
24-inv1 24-inv2	FESQNNNP	(T/G)GG(G/A)TT(G/A)TT(G/A)TTTTGTGA(T/C)TCAAA (T/G)GG(G/A)TT(G/A)TT(G/A)TTTTGGCT(T/C)TCAAA
IPCR-inv IPCR-dir	secuencia nt 5769-5798 secuencia nt 8311-8342	CCCTTTACAAGCTGATTTTGCTTATCTGCG GGGTCAAATCCTTACTATACATTGTACACGG
ECHO-dir ECHO-inv	secuencia nt -6 - 39 secuencia nt 8457-8504	AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG ATTTGAGGTAATGTTGATTTATCACCATCAAGCTTGAAATATTGACC

PCR:

La PCR se realizó utilizando un termociclador Perkin-Elmer, modelo 2400, y con 50 nanogramos de ADN genómico. Las cantidades de cebadores utilizadas eran de 10 µM de 29-Dir1 y de 24-Inv1. A la mezcla de reacción, se le añadieron 250 µM de cada desoxinucleótido trifosfato, y la Taq Polimerasa.

Después de una amplificación de 25 ciclos a 94 °C durante 30 segundos y a continuación a 50 °C durante 30 segundos, a continuación a 72 °C durante 5 minutos, se obtuvo un fragmento de 666 pares de bases.

Ciertos fragmentos se amplificaron con ayuda del sistema «Expand Long Template PCR» (Roche Boehringer Mannheim), de acuerdo con las indicaciones de utilización del proveedor. Este sistema permite amplificar fragmentos de gran tamaño, hasta aproximadamente 20 kpb, de manera muy eficaz. La combinación de dos ADN polimerasas permite concretamente minimizar la tasa de error durante las etapas de alargamiento.

Hibridación Southern y biblioteca genómica de *L. mesenteroides* NRRL B-1299:

El ADN cromosómico de *L. mesenteroides* NRRL B-1299 se digirió con diferentes enzimas de restricción, a continuación se separó por electroforesis en gel de agarosa al 0,8 % en tampón TAE 0,5X.

Bibliotecas genómicas de la bacteria se transfirieron en membranas de nylon hybond N+ (Amersham PharmaciaBiotech). La hibridación se realizó utilizando el fragmento de 666 pares de bases con desoxiadenosina-trifosfato marcado con <sup>32</sup>P. La reacción de marcado se realizó utilizando el kit de marcado "Mega Prime DNA Labelling System Kit" (Amersham PharmaciaBiotech), seguida por la purificación de la sonda en columnas MicroSpin S-200HR. La pre-hibridación y la hibridación se realizaron en condiciones fuertemente rigurosas (65 °C durante la noche, de acuerdo con los métodos habituales) (Sambrook y Russel, 2001, anteriormente).

#### PCR inversa:

La reacción de PCR inversa permite obtener un fragmento de ADN lineal a partir de una matriz circular utilizando cebadores divergentes.

El ADN genómico de *L. mesenteroides* NRRL B-1299 se digirió por EcoRV en las condiciones recomendadas por el proveedor.

Después de la re-circularización, los productos de digestión se utilizaron como matriz en una reacción de PCR inversa [Extrapol II DNA polymerase (Eurobio), volumen de reacción de 50 µl, parámetros de la reacción de PCR inversa: 25 ciclos; 94 °C, 30 segundos; 51 °C, 30 segundos; 72 °C, 3 minutos]. Los dos cebadores se seleccionaron en función de la secuencia del inserto de pSB2 como se indica en la figura 2.

La figura 2 resume las modalidades de obtención de los diferentes plásmidos portadores de los fragmentos de *dsrE* mediante cribado de la biblioteca genómica y utilización de las sondas descritas anteriormente.

#### Secuencia de ADN y análisis:

Después de la secuenciación de los péptidos, se sintetizaron cebadores degenerados diseñados teniendo en cuenta la frecuencia de utilización de los codones en los genes de dextrano-sacarasa de *L. mesenteroides* NRRL B-1299, y se permitió la amplificación de un fragmento de 666 pb. La secuenciación de este fragmento reveló fuertes homologías con los genes de dextrano-sacarasa ya conocidos, siendo todo totalmente nuevo.

La utilización de este fragmento como sonda homóloga en experimentos de Southern, permitió identificar señales positivas en diferentes pistas de ADN genómico digerido. De este modo, se cribó un primer banco *HindIII*, y se purificó un plásmido recombinante, denominado pSB2, que contenía un inserto de 5,6 kpb. El análisis de la secuencia de este fragmento *HindIII* reveló un marco abierto de lectura que cubre la totalidad del inserto. A continuación, se cribó un banco EcoRV con una sonda *HindIII*/EcoRV aislada en el extremo N-terminal del inserto *HindIII* de 5,6 kpb. Un plásmido recombinante pSB3, ensayado positivamente por dot-blot, demostró contener un inserto de 3,8 kpb que, después de la secuenciación, mostró contener el codón de inicio de la traducción y la región promotora del nuevo gen de dextrano-sacarasa denominado *dsrE*.

Con el objetivo de obtener el codón de terminación de *dsrE*, se realizó una PCR inversa en el ADN genómico de *L. mesenteroides* NRRL B-1299 digerido mediante EcoRV y religado sobre sí mismo, utilizando cebadores oligonucleotídicos divergentes diseñados a partir de la secuencia del inserto pSB2. Un fragmento único con el tamaño esperado de 1 kpb se amplificó y a continuación se clonó en un pGEM-T Easy, para obtener el plásmido pSB4. Después de la secuenciación, la secuencia amplificada situada cadena abajo del sitio *HindIII* consta de 221 pb y contiene el codón de terminación del marco de lectura de *dsrE*, situado 30 pb cadena abajo del sitio de restricción *HindIII*.

La secuenciación de los diferentes fragmentos portados por los tres plásmidos fue realizada por la compañía Génome Express y esto, en las dos cadenas. Los análisis de las secuencias de nucleótidos se realizaron utilizando el "ORF Finder" (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), Blast (<http://www.ncbi.nlm.nih.gov/blast/blast.cgi>, Altschul et al., 1997) ClustalW (<http://www2.ebi.ac.uk/clustalw>, Thompson et al., 1994), PRODOM (<http://protein.toulouse.inra.fr/pro-dom.html>, Corpet et al., 2000), PFAM (<http://pfam.wustl.edu/hmmsearch.shtml>, Bateman et al., 2000) y SAPS (<http://bioweb.pasteur.fr/seqanal/interfaces/saps.html>, Brendel et al., 1992), siendo el conjunto de estos softwares accesible por internet.

#### Expresión de la proteína:

Se utilizaron dos sistemas de clonación y de expresión con fines de la producción de proteínas recombinantes en *E. coli*, a saber los sistemas ECHO-Cloning y pBAD-TOPO Thiofusion (Invitrogen).

A modo de ejemplo, se describe brevemente a continuación el método de clonación de la secuencia nucleotídica que codifica la proteína DSR-E por medio del sistema ECHO-Cloning.

Se utilizaron dos cebadores tal como se proponen en la tabla I anterior en el marco de la amplificación con ayuda del sistema "Expand Long Template" en las siguientes condiciones: 94 °C durante 3 minutos seguidos de 25 ciclos a 94 °C durante 30 segundos, 55 °C durante 30 segundos, y 68 °C durante 7 minutos. Los productos de PCR se clonaron

a continuación en el vector pUNI/V5-His-TOPO, que permite la obtención de un vector donador (pUNI-dsrE) a recombinar con un vector aceptor (pCR-T7-E) y adaptado a la expresión en *E. coli*. El plásmido final se designó como pCR-T7-dsrE.

5 Esta construcción, que coloca el gen *dsrE* bajo el control del promotor del bacteriófago T7, permitió la expresión inducible del gen *dsrE*.

Después de la inducción con 1 mM de IPTG, las células de *E. coli* BL21 transformadas se recogieron mediante centrifugado después de 4 horas de crecimiento, y se resuspendieron a una densidad óptica final de 80 a 600 nm en tampón acetato de sodio 20 mM pH 5,4 y Triton X100 al 1 % (v/v) en presencia de PMSF 1 mM para impedir la proteólisis en los extractos celulares después de la sonicación.

Experimentos similares, realizados con el sistema pBAD-TOPO Thiofusion, permitieron construir el vector recombinante pBAD-TOPO-*dsrE*.

15 Ensayos enzimáticos:

Las reacciones enzimáticas se realizaron en las condiciones estándar a 30 °C en tampón acetato de sodio 20 mM pH 5,4, NaN<sub>3</sub> 1 g/l<sup>-1</sup> y sacarosa 100 g/l<sup>-1</sup>. La actividad de la enzima DSR-E se determinó midiendo la velocidad de liberación de los azúcares reductores, representados en este caso particular por fructosa, con ayuda del método con ácido dinitro-salicílico bien conocido por el experto en la materia. Una unidad se definió como la cantidad de enzima que catalizaba la formación de 1 μmol de fructosa por minuto en las condiciones estándar. Los oligosacáridos se sintetizaron en un medio de reacción que contenía 100 g/l de maltosa, 200 g/l de sacarosa y 0,5 unidades/ml de DSR-E.

Como para la síntesis de dextrano, la reacción enzimática continuó durante 24 horas en presencia de 100 g/l de glucosa. El dextrano producido se precipitó en presencia de etanol al 50 % (v/v) y se lavó dos veces en etanol al 50 % (v/v) antes de la liofilización. A continuación se disolvió a 10 mg/ml en D<sub>2</sub>O y se analizó mediante espectrometría de RMN del <sup>13</sup>C.

30 Separación por HPLC:

Se tomaron muestras de 100 μl y se calentaron a 95 °C durante 5 minutos, a continuación se diluyeron en agua ultra-pura para obtener una concentración final de azúcares totales inferior a 5 g.l<sup>-1</sup>. Después del centrifugado, los sustratos residuales y las diferentes especies formadas se analizaron por HPLC en columna C18 (Ultrasep 100, 6 μm, 5x300 mm, Bishoff Chromatography).

La separación de los oligósidos se realizó a temperatura ambiente, durante 30 minutos, en agua ultra-pura aplicada como eluyente a un caudal de 0,5 ml/min. La detección se consiguió por refractometría.

Dichas condiciones permitieron separar las siguientes especies: fructosa, maltosa, leucrosa, sacarosa, así como los oligósidos cuyo grado de polimerización no superaba 6.

45 Cálculo de los rendimientos:

El método de cálculo de los rendimientos de las reacciones de síntesis de oligósidos tuvo en cuenta la concentración residual de aceptor, de acuerdo con la siguiente fórmula:

$$R = \{[\text{GOS finales}] - [\text{GOS iniciales}]\} / \{0,474 \times [\text{sacarosa consumida}] + [\text{aceptor consumido}]\}$$

donde R representa el rendimiento real de la reacción de síntesis de los GOS totales, expresándose las concentraciones en g/l.

55 Construcción de las diferencias formas delecionadas de la proteína DSR-E:

Las diferentes formas delecionadas de la proteína DSR-E [figura 1c) a 1i)] se obtuvieron mediante amplificación por PCR de los fragmentos correspondientes del gen *dsrE*, a continuación clonación en el vector pBAD-TOPO Thiofusion, cuya descripción se proporciona anteriormente. Los cebadores utilizados para la amplificación de las regiones seleccionadas a partir del gen *dsrE* se enumeran en la tabla II siguiente. Las posiciones de los cebadores se indican en referencia a la secuencia ID no. 5, con respecto a la secuencia del gen *dsrE*. Las bases mutadas para introducir el sitio de restricción NcoI están en caracteres en negrita y el sitio NcoI resultante está subrayado.

TABLA II:

Designación	Posiciones	Secuencia 5'-3'
pBAD-PS/ZV-dir	344-373	<u>GCCATGGCAAATACGATTGCAGTTGACACG</u>
pBAD-ZV/CD1-dir	971-1001	<u>GCCATGGACGGTAAAACCTATTTTCTTGACG</u>

(continuación)		
Designación	Posiciones	Secuencia 5'-3'
pBAD-CD1/GBD-dir	3656-3682	TCCATGGGTGAAAAACAAGCACCGGC
pBAD-GBD/CD2-dir	6167-6189	ACCATG GATATGTCTACTAATGC
pBAD-CD1/GBD-inv	3638-3658	TAACGTTTAGGCAAGAATCC
pBAD-GBD/CD2-inv	6146-6172	TAATGTATTAGTGAATAAGTATTCCACC
pBAD-ent-inv	8714-8737	AATTTGAGGTAATGTTGATTATC

5 Los cebadores directos e inversos anteriores se diseñaron para garantizar la fusión traduccional de la etiqueta de tiorredoxina en N-terminal y la etiqueta polihistidina en C-terminal de las formas proteicas truncadas respetando los marcos abiertos de lectura de las regiones que codifican dichas formas.

10 En la medida en que el plásmido pBAD-TOPO Thiofusion contiene un sitio de restricción específico de la enzima *NcoI* situado a nivel del extremo 5' de la región que codifica la tiorredoxina, se introdujo un segundo sitio *NcoI* en cada cebador directo para poder, llegado el caso, extraer dicha región.

Las reacciones de amplificación por PCR se realizaron con ayuda del sistema «Expand Long Template» en las siguientes condiciones: una pre-desnaturalización a 94 °C durante 3 minutos, seguida de 25 ciclos a 94 °C durante 30 segundos, 52 °C durante 30 segundos y 68 °C durante 7 minutos.

15 Los productos de amplificación generados de este modo se clonaron a continuación en el vector pBAD-TOPO Thiofusion para fines de la transformación posterior de la cepa *E. coli* TOP 10. Se seleccionaron clones recombinantes, su perfil de restricción se analizó para identificar, para cada una de las formas buscadas, un plásmido recombinante que porta la inserción orientada como se esperaba.

20 *Ejemplo 1: caracterización y purificación de la enzima DSR-E y obtención del gen dsrE*

Las enzimas producidas por los cultivos de *L. mesenteroides* y obtenidas en gel de poli-acrilamida en SDS tal como se describe en la parte Materiales y Métodos se aislaron mediante recorte del gel.

25 Los GOS producidos por las enzimas aisladas de este modo se analizaron por HPLC de acuerdo con los métodos descritos en (1). La enzima cuya actividad se buscaba se dedujo de la naturaleza de los GOS producidos. Después de la proteólisis trípica y separación por HPLC de los péptidos producidos, 2 péptidos: 29-FYFESGK y 24-FESQNNNP se secuenciaron y se utilizaron como modelo para la síntesis de cebadores nucleotídicos degenerados.

30 Las diferentes etapas de amplificación y de clonación se representan en la figura 2. El gen completo se insertó en el plásmido pCR-T7-E y se expresó en *E. coli*.

La producción de una enzima funcional se certifica mediante la producción de los GOS cuyo análisis por HPLC se representa en la figura 6b).

35 Se destacará en particular la importancia de los picos 5 y 7 representativos de los GOS con ramificación  $\alpha(1\rightarrow2)$ .

*Ejemplo 2: caracterización de las secuencias de dsrE y DSR-E*

40 2.1 Secuencia nucleotídica:

La secuencia nucleotídica de la enzima se representa en la secuencia ID No. 4. Está compuesta por un marco de lectura de 8506 nucleótidos.

45 La secuencia nucleotídica del inserto en el plásmido pCR-T7-dsrE contiene un sitio de unión al ribosoma (RBS), 9 bases cadena arriba del codón de inicio ATG y está compuesta por un hexa-nucleótido GAGGAA.

2.2 Análisis de la secuencia de aminoácidos:

50 La secuencia de 8506 nucleótidos de *dsrE* codifica una proteína de 2835 aminoácidos representada en la secuencia ID No. 2. El punto isoeléctrico de esta proteína es de 4,88 y su peso molecular teórico de 313,2 kDa. A pesar de las fuertes similitudes con las dextrano-sacarosas ya conocidas, DSR-E se caracteriza por una estructura original.

55 El alineamiento de la secuencia de aminoácidos con el conjunto de las glucosiltransferasas y de las dextrano-sacarosas conocidas confirma que la estructura en dominio de las glucosiltransferasas y de las dextrano-sacarosas está conservada, a saber: una secuencia señal, una región variable, un dominio catalítico altamente conservado y un dominio de unión a glucano. Esta estructura se representa en la figura 1a).

60 Como indica la figura 1b), un segundo dominio catalítico forma la parte carboxi-terminal de la enzima como se ha confirmado mediante PRODOM y un análisis Blast.

Con un peso molecular de 313,2 kDa, DSR-E tiene aproximadamente 2 veces el peso molecular medio de las otras glucosiltransferasas y dextrano-sacarosas (1), lo que está de acuerdo con la presencia de un segundo dominio catalítico en el extremo C-terminal y también con un dominio de unión a glucano más largo.

5 a) análisis de la secuencia señal:

La secuencia señal y la secuencia nucleotídica que codifica el péptido señal están extremadamente conservadas si se les compara con las otras dextrano-sacarosas como se indica en la figura 3. El sitio de escisión está localizado entre los aminoácidos 40 y 41.

10

b) dominio variable:

Cadena abajo del péptido señal, DSR-E tiene un dominio variable de 207 aminoácidos. Cuando se le compara con los otros dominios variables de las glucosiltransferasas, utilizando un programa de alineamiento de tipo SAPS, se pone de manifiesto la presencia de un motivo de 14 aminoácidos repetido 11 veces como se indica en la figura 4.

15

Este motivo repetido, rico en alanina, treonina y ácido aspártico no ha sido identificado nunca anteriormente.

El papel y el significado de esta región no han sido dilucidados nunca. Diferentes estudios han demostrado que su delección no afecte a la actividad enzimática (4). El papel del motivo repetido de 14 aminoácidos que no existe en las otras glucosiltransferasas aún está, no obstante, por determinar.

20

c) análisis de los dominios catalíticos:

El primer dominio catalítico se extiende de los aminoácidos 248 a 1142 (CD1) de la secuencia ID No. 2, mientras que el segundo está localizado entre los aminoácidos 1980 y 2836 (CD2). Estos dos dominios presentan un 45 % de identidad y un 65 % de similitud entre sí.

25

CD1 y CD2 contienen los aminoácidos ya identificados en las glucosiltransferasas y las dextrano-sacarosas como esenciales por su actividad enzimática, y como se indica en la figura 5.

30

Las tríadas catalíticas de CD1 y CD2 determinadas por analogía con  $\alpha$  amilasa (7) están presentes en las posiciones siguientes: (Asp 527/Glu 565/Asp 638 para CD1 y Asp 2210/Glu 2248/Asp 2322 para CD2).

35

Otros residuos conservados se identificaron como importantes para la actividad enzimática: los residuos Trp 425/Glu 430 para CD1 y Trp 2122/Glu 2127 para CD2, los cuales son análogos a los del dominio N-terminal de GFTI descritos por Monchois et al. (4): Trp 344/Glu 349.

40

En cambio, ciertas secuencias situadas en la región conservada de las glucosiltransferasas y de las dextrano-sacarosas no se encuentran en el CD2 de DSR-E. De este modo, como se indica en la figura 5 a continuación, las secuencias FIHNDTI (2214-2220) y KGVQEKV (2323-2329) divergen de las otras secuencias consenso de las dextrano-sacarosas ya estudiadas que son, respectivamente, NVDADLL y SEVQTVI.

45 d) dominio de unión a glucano:

Quando se compara la secuencia de DSR-E con las secuencias conocidas, parece que el dominio de unión a glucano es sensiblemente más largo. En efecto, este dominio tiene una longitud de aproximadamente 500 aminoácidos en las glucosiltransferasas y las dextrano-sacarosas estudiadas mientras que, en DSR-E, representa 836 aminoácidos. Varios motivos repetidos A y C y, en particular, se han podido identificar una serie de repeticiones AC. En la tabla III siguiente, se indican las secuencias consenso de los motivos repetidos del GBD, concretamente motivos A y C, descritos en la bibliografía con respecto a dextrano-sacarosas de *Leuconostoc* y *Streptococcus* spp.

50

TABLA III:

Motivo	Secuencia consenso
A	WWYFNxDGQAATGLQTIIDGQTVFDDNGxQVKG
B	VNGKTYFSGDGTATQANPKGQTFKDGSGVLRFYNLEGQYVSGSGWY
C	DGKIYFFDPDSGEVVKNRFV
D	GGWKNADGTYSKY
N	YYFxAxQGxxxL

x: aminoácido indiferente.

55 *Ejemplo 3: expresión de dsrE en E. coli*

Se cultivaron células de *E. coli* BL21 (DE3) pLysS pCR-T7-*dsrE* como se ha descrito anteriormente. Después del gel de electroforesis en poli(acrilamida-(page-SDS)), el análisis de los extractos proteicos reveló efectivamente la presencia de varias bandas que tienen la actividad de dextrano sacarasa, midiéndose dicha actividad como anteriormente.

5 La línea *E. coli* JM109 [pCR-T7-*dsrD*] se depositó en la CNCM el 15 de marzo de 2001 con el número I-2649.

Identificación y caracterización de la actividad enzimática:

10 Utilizando una molécula aceptora de glucosa, las dextrano-sacarosas producidas por *E. coli* recombinante se compararon con las producidas por *L. mesenteroides* NRRL B-1299.

15 El análisis por HPLC de los productos de la reacción con DSR-E recombinante muestra (figura 6) tiempos de retención que corresponden a los GOS identificados previamente R4 y R5 (2). Los oligosacáridos de tipo R son las series de GOS lineales, estando el enlace  $\alpha(1\rightarrow2)$  unido al extremo no reductor. La serie OD, GOS lineales que resultan de enlaces glucosídicos  $\alpha(1\rightarrow6)$  con un residuo maltosa en el extremo reductor, se observó en muy bajas cantidades. En cambio, tres nuevos compuestos son detectados en los productos de la enzima recombinante.

20 Identificación de los GOS producidos:

Finalmente, la figura 6b muestra claramente que los picos 5 y 7 que representan los GOS de la serie R son relativamente más importantes con la enzima recombinante que con la enzima nativa cuyos picos que corresponden a panosa y a OD5 son más grandes.

25 *Ejemplo 4: efecto de la delección de CD2 sobre la actividad enzimáticas de DSR-E*

30 El ADN genómico de *L. mesenteroides* NRRL B-1299 se utilizó como matriz para amplificar por PCR el gen *dsrE* delecionado de la secuencia que corresponde al segundo dominio catalítico. Para ello, 2 oligonucleótidos, ECHO-dir (5'-AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG) (SEQ. ID No. 6), que corresponden a la secuencia nucleotídica -6 a 39 y que contienen el codón de inicio de la traducción, y ECHO-inv-del (5'-GTATTAGTGAATAAGTATTCACCATTGCATTTATCGTCAAATAGTACG) (SEQ. ID No. 7) complementaria de la secuencia 5889-5937 y que corresponde a la secuencia peptídica YYFDDKGNGEYCFTNT, se sintetizaron, para fusionar el extremo C-terminal de la proteína delecionada con una etiqueta His presente en el vector de clonación. La reacción de PCR se realizó gracias a un DNA thermal cycler, modelo 2400 (Perkin-Elmer), con el sistema Expand Long Template System (Boehringer Mannheim), siguiendo el ciclo de temperatura: 94 °C durante 3 min, a continuación 25 ciclos con: 30 s a 94 °C, 30 s a 55 °C y 7 min a 68 °C. El producto de PCR se clonó a continuación en el vector donador pUNI, y el plásmido resultante, utilizado en una reacción de recombinación con el vector de expresión pCR-T7- $\Delta dsrE$ .

40 La preparación de los extractos celulares, las reacciones enzimáticas, el análisis de los productos de la reacción son los mismos que en el ejemplo 3 anteriormente.

45 El perfil de HPLC de los GOS obtenidos con la enzima DSR-E delecionada del dominio CD2 aparece en la figura 6 c).

50 Los GOS de tipo R, representados por los picos 5 y 7 visibles en la figura 6 a) y 6 b) están totalmente ausentes de los productos obtenidos con la enzima recombinante delecionada de CD2. Los únicos productos analizables son aquellos que corresponden a oligosidos lineales que resultan de enlaces  $\alpha(1 \rightarrow 6)$  con un residuo maltosa en la parte reductora. Este resultado indica claramente el papel esencial del dominio catalítico situado en la parte carboxi-terminal de la enzima en su capacidad para formar enlaces osídicos  $\alpha(1 \rightarrow 2)$ .

*Ejemplo 5: estudio de las relaciones estructura-función de la proteína DSR-E*

55 El gen *dsrE*, dado que es el primer gen que codifica una dextrano-sacarasa que cataliza la síntesis de enlaces  $\alpha(1\rightarrow2)$  en haber sido clonado, reviste un interés particular. Por lo tanto, es importante caracterizar este gen así como su producto de expresión, en el caso en que se determina cuál es la implicación de los diferentes dominios que componen la proteína DSR-E en la función que le ha sido asignada, a saber de corresponder a una dextrano-sacarasa específica de la síntesis de enlaces  $\alpha(1\rightarrow2)$ .

60 5.1 Formas delecionadas de la proteína DSR-E:

65 El estudio de seis formas diferentes obtenidas por delección de uno o de varios dominios de la proteína DSR-E se proyectó para determinar, en referencia a la figura 1 a continuación: (i) la influencia de la presencia del dominio CD2 estudiando las construcciones GBD-CD2 y  $\Delta(CD2)$ ; (ii) la influencia de la presencia de la zona variable analizando las formas  $\Delta(ZV)$  y CD1-GBD; y (iii) el potencial catalítico intrínseco de los dominios CD1 y CD2 expresados de forma aislada (construcciones CD1 y CD2).

La actividad catalítica de cada una de las diferentes formas se comparó con la observada con el control que corresponde a la forma completa delecionada del único péptido señal  $\Delta(\text{PS})$  [figura 1c)].

## 5.2 Análisis de las construcciones:

Al finalizar el procedimiento experimental de amplificación por PCR y de clonación detallada anteriormente, se obtuvieron varios clones que presentan una inserción de acuerdo con la orientación esperada para cada una de las construcciones previstas, con la excepción de la forma truncada GBD-CD2 para la cual el producto de amplificación deseado no se pudo clonar.

Las secuencias de las inserciones se determinaron para asegurarse de la ausencia de mutaciones susceptibles de modificar, después de la traducción, aminoácidos situados a nivel de posiciones presumidas esenciales para la actividad enzimática de la proteína codificada de este modo.

Se identificó una mutación a nivel de la 31ª base de la inserción con respecto a control  $\Delta(\text{PS})$ , induciendo la sustitución de un ácido aspártico por una asparagina en posición 10 de la zona variable. Al no estar situada al nivel de los motivos repetidos S de la zona variable (figura 4), parecería que la incidencia de esta mutación sobre la función finalmente observada sería despreciable.

Se introdujo una mutación en el producto de amplificación correspondiente a la construcción  $\Delta(\text{CD2})$ , que modifica el residuo aromático F1411 a leucina. Esta mutación está situada en el primer tercio del dominio de unión a glucano GBD, a nivel de una unión entre dos motivos repetidos.

La construcción  $\Delta(\text{ZV})$ , habida cuenta de los errores efectuados por la polimerasa durante la reacción de amplificación por PCR, no presentaba la secuencia esperada. En efecto, la inserción contenía un marco abierto de lectura, correspondiendo dicho marco, en lo esencial, a la forma GBD-CD2 que no se había podido clonar. No obstante, en la forma GBD-CD2 obtenida en definitiva en lugar de  $\Delta(\text{ZV})$ , los 46 residuos N-terminales estaban ausentes. Ahora bien, el dominio GBD cuenta con más de 800 aminoácidos que forman una concatenación de 24 unidades repetidas. Esta concatenación es tal que, en los 46 residuos truncados, solamente los 9 últimos estaban situados a nivel de una de dichas unidades, y en particular a nivel de la primera de ellas. Parecía en consecuencia plausible considerar que la delección de estos aminoácidos no tenía influencia sobre la reacción enzimática catalizada por la forma proteica correspondiente. Esta hipótesis era tanto más verosímil en cuanto a que ha mostrado en otras dextrano-sacarosas, que la pérdida cierto número de unidades repetidas del dominio GBD no reducía la actividad de la proteína resultante de manera significativa (8).

La inserción que codifica la forma CD1-GBD contenía una mutación que afecta al residuo F633, situado en el dominio CD1, y más exactamente, a nivel de una región fuertemente conservada entre las dextrano-sacarosas, localizada a su vez justo antes del segundo ácido aspártico de la tríada catalítica (figura 5). La fenilalanina esperada estaba sustituida por una leucina. Es difícil, en esta fase, estimar el impacto de dicha mutación sobre la actividad catalítica observada.

De la misma manera que para las otras construcciones, se determina la secuencia de las inserciones que codifican los dominios catalíticos CD1 y CD2 aislados.

## 5.3 Productos de expresión y actividades enzimáticas:

Las proteínas que corresponden a las diversas formas delecionadas de DSR-E se expresaron sometiendo a las células recombinantes de *E. coli* a una inducción por L-arabinosa de acuerdo con una concentración del 0,002 %. La actividad enzimática se observó durante las cuatro primeras horas que seguían a la inducción.

Los extractos proteicos obtenidos por sonicación de los sedimentos celulares se analizaron en electroforesis SDS-PAGE (Sambrook y Russel, 2001, anteriormente). Las masas moleculares de las proteínas recombinantes se estimaron a partir de los perfiles electroforéticos obtenidos, correspondiendo dichas masas en lo esencial a las masas esperadas, teniendo en cuenta el aumento de 12 kDa asociado a la etiqueta de tiorredoxina. La tabla IV a continuación resume los valores estimados de las masas moleculares de las diferentes formas truncadas y proporciona, a modo de comparación, las masas esperadas.

TABLA IV:

Forma proteica	Masa esperada (kDa)	Masa esperada + tiorredoxina (kDa)	Masa estimada (kDa)
$\Delta(\text{PS})$	309	321	324
$\Delta(\text{CD2})$	218	230	ND
GBD-CD2	224	/	233
CD1-GBD	193	205	199
CD1	99	111	111
CD2	95	107	ND

ND: no determinada.

En la tabla V siguiente, se indican la naturaleza y la posición de los aminoácidos que marcan el inicio y el fin de las formas proteicas construidas en el marco de este estudio. Las diferentes posiciones hacen referencia a la secuencia ID no. 2 que corresponde a la proteína DSR-E.

5

TABLA V:

Forma proteica	Aminoácido de inicio	Aminoácido de fin	Longitud total
$\Delta$ (PS)	N41	I2835	2795
$\Delta$ (CD2)	M1	L1980	1980
GBD-CD2	M1188	I2835	1648
CD1-GBD	I248	L1980	1733
CD1	I248	Q1141	894
CD2	D1981	I2835	855

La forma GBD-CD2 no poseía la etiqueta de tiorredoxina. En efecto, esta forma derivaba de los riesgos experimentales ocasionados por el procedimiento de amplificación por PCR de la secuencia que se supone que codifica la forma  $\Delta$ (ZV). Habida cuenta de las deleciones de secuencias generadas entonces, la etiqueta de tiorredoxina, en principio situada en 5' de la proteína de interés, no se pudo fusionar con la región GBD-CD2.

10

La calidad de los geles de electroforesis no permitió determinar si el nivel de expresión de las diferentes formas era cuantitativamente idéntico y, en consecuencia, si dichas formas estaban presentes en las mismas proporciones en los extractos celulares.

15

También las medidas de actividad proporcionadas se establecieron sobre la base de un volumen dado de extractos celulares, pero no se pudieron llevar a la cantidad de cada proteína contenida realmente en dicho volumen de extractos.

20

La síntesis de polímeros de dextrano *in situ* por incubación de los geles de electroforesis en una solución de sacarosa, y la tinción con reactivo de Schiff posterior confirmaron la presencia de proteínas que poseen una actividad glucano-sacarasa en los extractos celulares que corresponden a  $\Delta$ (PS),  $\Delta$ (CD2), GBD-CD2 y CD1-GBD.

25

En la tabla VI a continuación, se presentan las actividades enzimáticas máximas observadas para cada construcción. Los resultados confirmaron los datos extraídos de los experimentos de tinción de los geles con reactivo de Schiff, a saber el hecho de que los extractos celulares relativos a las formas  $\Delta$ (PS),  $\Delta$ (CD2), GBD-CD2 y CD1-GBD presentaban una actividad sacarasa, al contrario que los dos dominios catalíticos tomados de forma aislada. Este resultado estaba conforme a la bibliografía, dado que se había demostrado para otras dextrano-sacarosas, que la ausencia del dominio GBD inducía una pérdida de actividad enzimática drástica (8, 9, 10).

30

TABLA VI:

Forma proteica	$\Delta$ (PS)	$\Delta$ (CD2)	GBD-CD2	CD1-GBD	CD1	CD2
Actividad máxima (U/l)	1063	181	86	235	5,3	0

La forma CD1 poseía una actividad intrínseca de nivel suficiente para ser detectada. En cuanto a la forma GBD-CD2, ésta presentaba un nivel de actividad no despreciable, de la que se puede concluir que la organización estructural correspondiente, a saber un dominio catalítico cadena abajo del dominio de unión a glucano, seguía siendo enzimáticamente activa.

35

#### 5.4 Efectos de las deleciones sobre la síntesis de oligósidos:

En la medida en que la especificidad de síntesis de los enlaces  $\alpha(1\rightarrow2)$  se conserva durante la reacción en presencia de un aceptor, en primer lugar se efectuaron experimentos de síntesis de oligósidos a partir de maltosa (figura 7).

Cuando las reacciones se llevaron a término, es decir cuando toda la sacarosa se ha consumido, se calcularon los rendimientos de síntesis en oligósidos. Los resultados aparecen en la tabla VII siguiente. Solamente la reacción que implica el extracto celular que contiene la forma proteica CD1 no pudo permitir dicho cálculo. El efecto de la temperatura condujo probablemente a la inactivación de la muy baja actividad presente en el extracto proteico.

45

TABLA VII:

Forma proteica	Rendimiento en oligósidos de la serie OD (%)	Rendimiento en oligósidos de la serie R (%)	Rendimiento en oligósidos totales
Enzima nativa	36	28	64
$\Delta$ (PS)	41	14	55
$\Delta$ (CD2)	67	1	68
GBD-CD2	45	47	92
CD1-GBD	100	0	100

Como se indica en la figura 7 a continuación, la presencia de oligósidos de la serie R solamente se detectaba con las formas enzimáticas que poseen el dominio catalítico CD2, con la excepción del caso en que dicho dominio estaba aislado y entonces inactivado completamente. En efecto, los tiempos de retención de los oligósidos sintetizados por la forma delecionada del segundo dominio catalítico, así como por la forma CD1-GBD, correspondían únicamente a los de la serie OD, es decir a los GOS desprovistos de enlaces  $\alpha(1\rightarrow 2)$ . Estos resultados indicaban, por lo tanto, que el dominio CD2 se requería para la formación de los enlaces  $\alpha(1\rightarrow 2)$ .

Los productos obtenidos con la forma GBD-CD2 permitieron respaldar estas observaciones. Esta construcción, que poseía CD2 como único dominio catalítico, era capaz de catalizar de manera preponderante la síntesis de oligósidos de la serie R, que poseen enlaces  $\alpha(1\rightarrow 2)$ . Este resultado demostraba, por lo tanto, que la especificidad en términos de función de la enzima DSR-E reside en la secuencia muy original de este dominio, y no en la asociación de dos dominios catalíticos. Además, la forma proteica GBD-CD2 permitía también la síntesis de enlaces  $\alpha(1\rightarrow 6)$ . Sin embargo, los bajos rendimientos obtenidos para estos oligósidos indicaban que se habían convertido preferentemente en oligósidos de grado de polimerización superior que pertenecen a la serie R, lo que impedía su acumulación en el medio de reacción, a diferencia de las moléculas de la serie R que, éstas, no se convirtieron (2).

Comparando los perfiles de los productos obtenidos, tal como se muestran en la figura 7, aparece que la forma completa  $\Delta(\text{PS})$  sintetizaba mayoritariamente oligósidos lineales. En efecto, la molécula R4 estaba ausente y el oligósido R5 presente escasamente. El dominio catalítico CD1 permitía, por su parte, catalizar la síntesis exclusiva de enlaces  $\alpha(1\rightarrow 6)$  y su actividad parecía preponderante con respecto a la del dominio CD2. También, en la forma completa de la enzima, la implicación del dominio CD2 sería, por lo tanto, menos importante debido a: (i) parámetros catalíticos intrínsecos más bajos; y/o (ii) de una configuración global de la enzima desfavorable para su actividad.

Además, la enzima completa  $\Delta(\text{PS})$  catalizaba la síntesis de oligósidos de la serie R con un rendimiento inferior al observado con la mezcla de dextrano-sacarosas producida por *L. mesenteroides* NRRL B-1299 (figura 7). El rendimiento obtenido, del 28 %, estaba situado entre los observados con la forma completa  $\Delta(\text{PS})$  por un lado, y la forma GBD-CD2 por otro lado. Ahora bien, se sabe que la cepa silvestre produce varias formas de dextrano-sacarosas susceptibles de sintetizar enlaces osídicos, y concretamente enlaces  $\alpha(1\rightarrow 2)$ . Se formuló una hipótesis, de acuerdo con la cual dichas formas eran productos de degradación de DSR-E. De este modo, en la medida en que formas truncadas de DSR-E, tales como GBD-CD2, podían catalizar la síntesis de oligósidos de la serie R más eficazmente, parecería que los rendimientos obtenidos con la mezcla heterogénea producida por *L. mesenteroides* NRRL B-1299 fueran imputables a la contribución de las actividades catalíticas del conjunto de estas diferentes formas enzimáticas.

En conclusión, los inventores consiguieron, aislando una dextrano sacarasa particular producida por *L. mesenteroides*, caracterizar una estructura particular e inesperada de esta enzima capaz de producir oligósidos de interés y que presentan ramificaciones de tipo  $\alpha(1\rightarrow 2)$ . La identificación y la caracterización de esta secuencia permiten, por un lado, construir células u organismos recombinantes que expresan de forma específica esta enzima y, por otro lado, prever su modificación mediante mutagénesis dirigida o aleatoria o mediante evolución molecular (DNA Shuffling) para mejorar aún más sus características.

Este aporte permite además mejorar el rendimiento y la reproducibilidad de la producción de los GOS de interés para las diferentes aplicaciones citadas anteriormente.

#### Referencias bibliográficas

(1) Monchois V., Willemot R.M., Monsan P. (1999). Glucansucrases: mechanism of action and structure-function relationships. *FEMS microbiol. Rev.* 23, 131-151.

(2) Dols M., Remaud-Simeon M., Willemot R.M., Vignon M.R., Monsan P.F. (1998). Structural characterization of the maltose acceptor-products synthesised by *Leuconostoc mesenteroides* NRRL B-1299 dextranase. *Carbohydrate Research* 305, 549-559.

(3) Arnold F.H. (2001). *Nature* 409 n° 6817, 253.

(4) Monchois V. Vignon M., Russel R.R.B. (1999). Isolation of key amino-acid residues at the N-terminal end of the core region of *Streptococcus downei* glucansucrase GTF-I. *Appl. Microbiol. Biotechnol.* 52, 660-665.

(5) Wilke-Douglas M., Perchorowicz J.T., Houck C.M., Thomas B.R. (1989). Methods and compositions for altering physical characteristics of fruit and fruit products. Patente PCT, WO 89/12386.

(6) Arguello-MOrales M.A., Remaud-Simeon M., Pizzut S., Sarçabal P., Willemot R.M., Monsan P. (2000). Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355. *FEMS Microb. Lett.* 182, 81-85.

(7) Devulapalle K.S., Goodman S., Gao Q., Hemsley A., Mooser G. (1997). Knowledge-based model of a glucosyltransferase from oral bacterial group of mutants streptococci. *Protein Sci.* 6, 2489-2493.

5 (8) Kato C., y Kuramitsu H.K. (1990). Carboxy-terminal deletion analysis of the *Streptococcus mutans* glucosyltransferase-I enzyme. *FEMS Microbiol. Lett.* 72, 299-302.

(9) Lis M., Shiroza T., y Kuramitsu H.K. (1995). Role of the C-terminal direct repeating units of the *Streptococcus mutans* glucosyltransferase-S in glucan binding. *Appl. Env. Microbiol.* 61, 2040-2042.

10 (10) Monchois V., Remaud-Simeon M., Russel R.R.B., Monsan P., y Willemot R.M. (1997). Characterization of *Leuconostoc mesenteroides* NRRL B-512F dextranucrase (DSR-S) and identification of amino-acid residues playing a key role in enzyme activity. *Appl. Microbiol. Biotechnol.* 48, 465-472

LISTADO DE SECUENCIAS

15 <110> Instituto Nacional de Ciencias Aplicadas de Toulouse (INSA)

<120> MOLÉCULA DE ÁCIDOS NUCLEICOS QUE CODIFICA UNA DEXTRANO-SACARASA QUE CATALIZA LA SÍNTESIS DE DEXTRANO QUE PORTA RAMIFICACIONES DE TIPO ALFA-1,2 OSÍDICAS

20 <130> B4787A (INPI) CNRS/INSA TOULOUSE

<140> xxxxxxxx

<141> 19-12-2001

25 <150> 0103631

<151> 16-03-2001.

<160> 17

30 <170> PatentIn Ver. 2.1

<210> 1

<211> 855

35 <212> PRT

<213> Aminoácidos (dominio catalítico nº 2)

<400> 1

ES 2 726 941 T3

Asp Met Ser Thr Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn His  
 1 5 10 15  
 Asp Ser Ser Ser Phe Asp His Thr Val Asp Gly Phe Leu Thr Ala Asp  
 20 25 30  
 Thr Trp Tyr Arg Pro Lys Ser Ile Leu Ala Asn Gly Thr Thr Trp Arg  
 35 40 45  
 Asp Ser Thr Asp Lys Asp Met Arg Pro Leu Ile Thr Val Trp Trp Pro  
 50 55 60  
 Asn Lys Asn Val Gln Val Asn Tyr Leu Asn Phe Met Lys Ala Asn Gly  
 65 70 75 80  
 Leu Leu Thr Thr Ala Ala Gln Tyr Thr Leu His Ser Asp Gln Tyr Asp  
 85 90 95  
 Leu Asn Gln Ala Ala Gln Asp Val Gln Val Ala Ile Glu Arg Arg Ile  
 100 105 110  
 Ala Ser Glu His Gly Thr Asp Trp Leu Gln Lys Leu Leu Phe Glu Ser  
 115 120 125  
 Gln Asn Asn Asn Pro Ser Phe Val Lys Gln Gln Phe Ile Trp Asn Lys  
 130 135 140  
 Asp Ser Glu Tyr His Gly Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr  
 145 150 155 160

ES 2 726 941 T3

Leu Lys Tyr Gly Asn Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr  
 165 170 175  
 Arg Gln Pro Gly Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp  
 180 185 190  
 Asn Ser Asn Pro Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr  
 195 200 205  
 Leu Met Asn Phe Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe  
 210 215 220  
 Asp Ser Ile Arg Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile  
 225 230 235 240  
 Gln Arg Thr Tyr Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln Ser  
 245 250 255  
 Glu Ala Lys Ala Asn Gln His Ile Ser Leu Val Glu Ala Gly Leu Asp  
 260 265 270  
 Ala Gly Thr Ser Thr Ile His Asn Asp Ala Leu Ile Glu Ser Asn Leu  
 275 280 285  
 Arg Glu Ala Ala Thr Leu Ser Leu Thr Asn Glu Pro Gly Lys Asn Lys  
 290 295 300  
 Pro Leu Thr Asn Met Leu Gln Asp Val Asp Gly Gly Thr Leu Ile Thr  
 305 310 315 320  
 Asp His Thr Gln Asn Ser Thr Glu Asn Gln Ala Thr Pro Asn Tyr Ser  
 325 330 335  
 Ile Ile His Ala His Asp Lys Gly Val Gln Glu Lys Val Gly Ala Ala  
 340 345 350  
 Ile Thr Asp Ala Thr Gly Ala Asp Trp Thr Asn Phe Thr Asp Glu Gln  
 355 360 365  
 Leu Lys Ala Gly Leu Glu Leu Phe Tyr Lys Asp Gln Arg Ala Thr Asn  
 370 375 380  
 Lys Lys Tyr Asn Ser Tyr Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu  
 385 390 395 400  
 Thr Asn Lys Asp Thr Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln  
 405 410 415  
 Asp Asp Gly Gln Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu  
 420 425 430  
 Val Ser Leu Met Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr  
 435 440 445  
 Met Ser Val Asp Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys  
 450 455 460  
 Asp Ala Met Thr Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu  
 465 470 475 480

ES 2 726 941 T3

Gly Leu Gly Val Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn Asp  
485 490 495

Ser Asp Lys Val Thr Leu Asp Met Gly Ala Ala His Lys Asn Gln Lys  
500 505 510

Tyr Arg Ala Val Ile Leu Thr Thr Arg Asp Gly Leu Ala Thr Phe Asn  
515 520 525

Ser Asp Gln Ala Pro Thr Ala Trp Thr Asn Asp Gln Gly Thr Leu Thr  
530 535 540

Phe Ser Asn Gln Glu Ile Asn Gly Gln Asp Asn Thr Gln Ile Arg Gly  
545 550 555 560

Val Ala Asn Pro Gln Val Ser Gly Tyr Leu Ala Val Trp Val Pro Val  
565 570 575

Gly Ala Ser Asp Asn Gln Asp Ala Arg Thr Ala Ala Thr Thr Thr Glu  
580 585 590

Asn His Asp Gly Lys Val Leu His Ser Asn Ala Ala Leu Asp Ser Asn  
595 600 605

Leu Ile Tyr Glu Gly Phe Ser Asn Phe Gln Pro Lys Ala Thr Thr His  
610 615 620

Asp Glu Leu Thr Asn Val Val Ile Ala Lys Asn Ala Asp Val Phe Asn  
625 630 635 640

Asn Trp Gly Ile Thr Ser Phe Glu Met Ala Pro Gln Tyr Arg Ser Ser  
645 650 655

Gly Asp His Thr Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe  
660 665 670

Thr Asp Arg Tyr Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr  
675 680 685

Asp Gly Asp Leu Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met  
690 695 700

Gln Val Met Ala Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly  
705 710 715 720

Lys Glu Val Val Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp Asp  
725 730 735

Ala Thr Gly Phe Gly Thr Gln Leu Tyr Val Thr Asn Ser Val Gly Gly  
740 745 750

Gly Gln Tyr Gln Glu Lys Tyr Ala Gly Gln Tyr Leu Glu Ala Leu Lys  
755 760 765

Ala Lys Tyr Pro Asp Leu Phe Glu Gly Lys Ala Tyr Asp Tyr Trp Tyr  
770 775 780

Lys Asn Tyr Ala Asn Asp Gly Ser Asn Pro Tyr Tyr Thr Leu Ser His  
785 790 795 800

ES 2 726 941 T3

Gly Asp Arg Glu Ser Ile Pro Ala Asp Val Ala Ile Lys Gln Trp Ser  
 805 810 815  
 Ala Lys Tyr Met Asn Gly Thr Asn Val Leu Gly Asn Gly Met Gly Tyr  
 820 825 830  
 Val Leu Lys Asp Trp His Asn Gly Gln Tyr Phe Lys Leu Asp Gly Asp  
 835 840 845  
 Lys Ser Thr Leu Pro Gln Ile  
 850 855

<210> 2 <211> 2835

<212> PRT

5 <213> Aminoácidos (proteína completa DSR-E)

<400> 2

Met Arg Asp Met Arg Val Ile Cys Asp Arg Lys Lys Leu Tyr Lys Ser  
 1 5 10 15  
 Gly Lys Val Leu Val Thr Ala Gly Ile Phe Ala Leu Met Met Phe Gly  
 20 25 30  
 Val Thr Thr Ala Ser Val Ser Ala Asn Thr Ile Ala Val Asp Thr Asn  
 35 40 45  
 His Ser Arg Thr Ser Ala Gln Ile Asn Lys Ser Ala Val Asp Lys Val  
 50 55 60  
 Asn Asp Asp Lys Thr Thr Leu Gly Ala Ala Lys Val Val Ala Val Ala  
 65 70 75 80  
 Thr Thr Pro Ala Thr Pro Val Ala Asp Lys Thr Val Ser Ala Pro Ala  
 85 90 95  
 Ala Asp Lys Ala Val Asp Thr Thr Ser Ser Thr Thr Pro Ala Thr Asp  
 100 105 110  
 Lys Ala Val Asp Thr Thr Pro Thr Thr Pro Ala Ala Asp Lys Ala Val  
 115 120 125  
 Asp Thr Thr Pro Thr Thr Pro Ala Ala Asp Lys Ala Val Asp Thr Thr  
 130 135 140  
 Pro Thr Thr Pro Ala Ala Asn Lys Ala Val Asp Thr Thr Pro Ala Thr  
 145 150 155 160  
 Ala Ala Thr Asp Lys Ala Val Ala Thr Pro Ala Thr Pro Ala Ala Asp  
 165 170 175  
 Lys Leu Ala Asn Thr Thr Pro Ala Thr Asp Lys Ala Val Ala Thr Thr  
 180 185 190  
 Pro Ala Thr Pro Val Ala Asn Lys Ala Ala Asp Thr Ser Ser Ile His  
 195 200 205  
 Asp Gln Pro Leu Asp Thr Asn Val Pro Thr Asp Lys Ser Ala Asn Leu  
 210 215 220

ES 2 726 941 T3

Val Ser Thr Thr Gln Lys Ser Thr Asp Asn Gln Gln Val Lys Ser Thr  
 225 230 235 240

Glu Thr Ser His Leu Gln Glu Ile Asn Gly Lys Thr Tyr Phe Leu Asp  
 245 250 255

Asp Asn Gly Gln Val Lys Lys Asn Phe Thr Ala Ile Ile Asp Gly Lys  
 260 265 270

Val Leu Tyr Phe Asp Lys Thr Ser Gly Glu Leu Thr Ala Asn Ala Pro  
 275 280 285

Gln Val Thr Lys Gly Leu Val Asn Ile Asp Asn Ala His Asn Ala Ala  
 290 295 300

His Asp Leu Thr Ala Asp Asn Phe Thr Asn Val Asp Gly Tyr Leu Thr  
 305 310 315 320

Ala Asn Ser Trp Tyr Arg Pro Lys Asp Ile Leu Lys Asn Gly Thr Thr  
 325 330 335

Trp Thr Pro Thr Thr Ala Glu Asp Phe Arg Pro Leu Leu Met Ser Trp  
 340 345 350

Trp Pro Asp Lys Asn Thr Gln Val Ala Tyr Leu Gln Tyr Met Gln Ser  
 355 360 365

Val Gly Met Leu Pro Asp Asp Val Lys Val Ser Asn Asp Asp Asn Met  
 370 375 380

Ser Thr Leu Thr Asp Ala Ala Met Thr Val Gln Lys Asn Ile Glu Ser  
 385 390 395 400

Arg Ile Gly Val Ser Gly Lys Thr Asp Trp Leu Lys Gln Asp Met Asn  
 405 410 415

Lys Leu Ile Asp Ser Gln Ala Asn Trp Asn Ile Asp Ser Glu Ser Lys  
 420 425 430

Gly Asn Asp His Leu Gln Gly Gly Ala Leu Leu Tyr Val Asn Asp Asp  
 435 440 445

Lys Thr Pro Asn Ala Asn Ser Asp Tyr Arg Leu Leu Asn Arg Thr Pro  
 450 455 460

Thr Asn Gln Thr Gly Gln Ile Thr Asp Pro Ser Lys Gln Gly Gly Tyr  
 465 470 475 480

Glu Met Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro Val Val Gln  
 485 490 495

Ala Glu Gln Leu Asn Trp Leu His Tyr Met Met Asn Ile Gly Thr Ile  
 500 505 510

Ala Gln Asn Asp Pro Thr Ala Asn Phe Asp Gly Tyr Arg Val Asp Ala  
 515 520 525

Val Asp Asn Val Asp Ala Asp Leu Leu Gln Ile Ala Gly Asp Tyr Phe  
 530 535 540

ES 2 726 941 T3

Lys Ala Ala Tyr Gly Thr Gly Lys Thr Glu Ala Asn Ala Asn Asn His  
 545 550 555 560  
 Ile Ser Ile Leu Glu Asp Trp Asp Asn Asn Asp Ser Ala Tyr Ile Lys  
 565 570 575  
 Ala His Gly Asn Asn Gln Leu Thr Met Asp Phe Pro Ala His Leu Ala  
 580 585 590  
 Leu Lys Tyr Ala Leu Asn Met Fro Leu Ala Ala Gln Ser Gly Leu Glu  
 595 600 605  
 Pro Leu Ile Asn Thr Ser Leu Val Lys Arg Gly Lys Asp Ala Thr Glu  
 610 615 620  
 Asn Glu Ala Gln Pro Asn Tyr Ala Phe Ile Arg Ala His Asp Ser Glu  
 625 630 635 640  
 Val Gln Thr Val Ile Ala Gln Ile Ile Lys Asp Lys Ile Asn Thr Lys  
 645 650 655  
 Ser Asp Gly Leu Thr Val Thr Pro Asp Glu Ile Lys Gln Ala Phe Thr  
 660 665 670  
 Ile Tyr Asn Ala Asp Glu Leu Lys Ala Asp Lys Glu Tyr Thr Ala Tyr  
 675 680 685  
 Asn Ile Pro Ala Ser Tyr Ala Val Leu Leu Thr Asn Lys Asp Thr Val  
 690 695 700  
 Pro Arg Val Tyr Tyr Gly Asp Leu Phe Ser Asp Asp Gly Gln Tyr Met  
 705 710 715 720  
 Ser Gln Lys Ser Pro Tyr Tyr Asp Ala Ile Thr Ser Leu Leu Lys Ser  
 725 730 735  
 Arg Ile Lys Tyr Val Ala Gly Gly Gln Ser Met Asn Met Thr Tyr Leu  
 740 745 750  
 His Glu Cys Phe Asp Pro Ala Lys Asn Glu Thr Lys Pro Gln Gly Val  
 755 760 765  
 Leu Thr Ser Val Arg Tyr Gly Lys Gly Ala Met Thr Ala Asp Asp Leu  
 770 775 780  
 Gly Asn Ser Asp Thr Arg Gln Gln Gly Ile Gly Leu Val Ile Asn Asn  
 785 790 795 800  
 Lys Pro Phe Leu Asn Leu Asn Asp Asp Glu Gln Ile Val Leu Asn Met  
 805 810 815  
 Gly Ala Ala His Lys Asn Gln Ala Tyr Arg Pro Leu Met Leu Thr Thr  
 820 825 830  
 Lys Ser Gly Leu Gln Ile Tyr Asp Lys Asp Ala Gly Ala Pro Val Val  
 835 840 845  
 Tyr Thr Asn Asp Ala Gly Gln Leu Ile Phe Lys Ser Asp Met Val Tyr  
 850 855 860

ES 2 726 941 T3

Gly Val Ser Asn Pro Gln Val Ser Gly Tyr Phe Ala Ala Trp Val Pro  
865 870 875 880

Val Gly Ala Ser Asp Ser Gln Asp Ala Arg Thr Gln Ser Ser Gln Ser  
885 890 895

Glu Thr Lys Asp Gly Asp Val Tyr His Ser Asn Ala Ala Leu Asp Ser  
900 905 910

Asn Val Ile Tyr Glu Gly Phe Ser Asn Phe Gln Ala Met Pro Glu Lys  
915 920 925

Asn Asp Asp Phe Thr Asn Val Lys Ile Ala Gln Asn Ala Lys Leu Phe  
930 935 940

Lys Asp Leu Gly Ile Thr Ser Phe Glu Leu Ala Pro Gln Tyr Arg Ser  
945 950 955 960

Ser Thr Asp Asn Ser Phe Leu Asp Ser Val Ile Gln Asn Gly Tyr Ala  
965 970 975

Phe Thr Asp Arg Tyr Asp Val Gly Tyr Asn Thr Pro Thr Lys Tyr Gly  
980 985 990

Thr Val Asp Gln Leu Leu Asp Ser Leu Arg Ala Leu His Ala Gln Gly  
995 1000 1005

Ile Gln Ala Ile Asn Asp Trp Val Pro Asp Gln Ile Tyr Asn Leu Pro  
1010 1015 1020

Gly Glu Gln Ile Val Thr Ala Val Arg Thr Asn Gly Ser Gly Lys Tyr  
1025 1030 1035 1040

Asp Tyr Asp Ser Val Ile Asn Asn Thr Leu Tyr Asp Ser Arg Thr Val  
1045 1050 1055

Gly Gly Gly Glu Tyr Gln Glu Lys Phe Gly Gly Leu Phe Leu Asp Gln  
1060 1065 1070

Leu Lys Lys Asp Tyr Pro Ser Leu Phe Glu Thr Lys Gln Ile Ser Thr  
1075 1080 1085

Asn Gln Pro Met Asn Pro Asp Val Lys Ile Lys Glu Trp Ser Ala Lys  
1090 1095 1100

Tyr Phe Asn Gly Ser Asn Ile Gln Gly Arg Gly Ala Trp Tyr Val Leu  
1105 1110 1115 1120

Lys Asp Trp Ala Thr Asn Gln Tyr Phe Asn Val Ser Ser Asp Asn Gly  
1125 1130 1135

Phe Leu Pro Lys Gln Leu Leu Gly Glu Lys Thr Ser Thr Gly Phe Ile  
1140 1145 1150

Thr Glu Asn Gly Lys Thr Ser Phe Tyr Ser Thr Ser Gly Tyr Gln Ala  
1155 1160 1165

Lys Asp Thr Phe Ile Gln Asp Gly Thr Asn Trp Tyr Tyr Phe Asp Asn  
1170 1175 1180

ES 2 726 941 T3

Ala Gly Tyr Met Leu Thr Gly Lys Gln Asn Ile His Asp Lys Asn Tyr  
1185 1190 1195 1200

Tyr Phe Leu Pro Asn Gly Val Glu Leu Gln Asp Ala Tyr Leu Phe Asp  
1205 1210 1215

Gly Asn Gln Glu Phe Tyr Tyr Asn Lys Ala Gly Glu Gln Val Met Asn  
1220 1225 1230

Gln Tyr Tyr Gln Asp Ser Gln Asn Gln Trp His Tyr Phe Phe Glu Asn  
1235 1240 1245

Gly Arg Met Ala Ile Gly Leu Thr Glu Val Pro Asn Ala Asp Gly Thr  
1250 1255 1260

His Val Thr Gln Tyr Phe Asp Ala Asn Gly Val Gln Ile Lys Gly Thr  
1265 1270 1275 1280

Ala Ile Lys Asp Gln Asn Asn Gln Leu Arg Tyr Phe Asp Glu Ala Thr  
1285 1290 1295

Gly Asn Met Val Val Asn Ser Trp Gly Gln Leu Ala Asp Lys Ser Trp  
1300 1305 1310

Leu Tyr Leu Asn Ala Gln Gly Val Ala Val Thr Gly Asn Gln Lys Ile  
1315 1320 1325

Asp Gly Glu Glu Tyr Tyr Phe Asn Ala Asp Gly Lys Gln Val Lys Gly  
1330 1335 1340

Asn Ala Ile Ile Asp Asn Asn Gly Asp Gln Arg Tyr Tyr Asp Gly Asp  
1345 1350 1355 1360

Lys Gly Val Met Val Val Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser  
1365 1370 1375

Trp Leu Tyr Leu Asn Asp Lys Gly Ile Ala Val Thr Gly Arg Gln Val  
1380 1385 1390

Ile Asn Asn Gln Val Asn Phe Phe Gly Asn Asp Gly Lys Gln Ile Lys  
1395 1400 1405

Asp Ala Phe Lys Leu Leu Ser Asp Gly Ser Trp Val Tyr Leu Asp Asp  
1410 1415 1420

Lys Gly Leu Ile Thr Thr Gly Ala Lys Val Ile Asn Gly Leu Asn Met  
1425 1430 1435 1440

Phe Phe Asp Lys Asp Gly His Gln Ile Lys Gly Asp Ala Ser Thr Asp  
1445 1450 1455

Ala Asn Gly Lys Arg His Tyr Tyr Asp Lys Asn Asp Gly His Leu Val  
1460 1465 1470

Thr Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser Trp Leu Tyr Leu Glu  
1475 1480 1485

Glu Gln Gly Asp Ala Val Thr Gly Gln Arg Val Ile Asp Gly Lys Thr  
1490 1495 1500

ES 2 726 941 T3

Arg Tyr Phe Asp Glu Asp Gly Lys Gln Ile Lys Asn Ser Leu Lys Thr  
1505 1510 1515 1520

Leu Ala Asn Gly Asp Lys Ile Tyr Leu Asp Gly Asp Gly Val Ala Ala  
1525 1530 1535

Thr Gly Leu Gln His Val Gly Asp Lys Ile Met Tyr Phe Asp Glu Asp  
1540 1545 1550

Gly Lys Gln Val Val Gly Lys Phe Val Ser Ala Lys Asp Gly Ser Trp  
1555 1560 1565

Tyr Tyr Leu Asn Gln Asp Gly Val Ala Ala Val Gly Pro Ser Ser Ile  
1570 1575 1580

Asn Gly Gln Ser Leu Tyr Phe Asp Gln Asp Gly Lys Gln Val Lys Tyr  
1585 1590 1595 1600

Asn Glu Val Arg Asn Ser Asp Gly Thr Thr Asn Tyr Tyr Thr Gly Leu  
1605 1610 1615

Thr Gly Glu Lys Leu Thr Gln Asp Phe Gly Glu Leu Pro Asp Gly Ser  
1620 1625 1630

Trp Ile Tyr Leu Asp Ala Gln Gly His Thr Val Thr Gly Ala Gln Ile  
1635 1640 1645

Ile Asn Gly Gln Asn Leu Tyr Phe Lys Ala Asp Gly Gln Gln Val Lys  
1650 1655 1660

Gly His Ala Tyr Thr Asp Gln Leu Gly His Met Arg Phe Tyr Asp Pro  
1665 1670 1675 1680

Asp Ser Gly Asp Met Leu Ser Asn Arg Phe Glu Gln Ile Thr Pro Gly  
1685 1690 1695

Val Trp Ala Tyr Phe Gly Ala Asp Gly Val Ala Ile Thr Gly Gln His  
1700 1705 1710

Asp Ile Asn Gly Gln Lys Leu Phe Phe Asp Glu Thr Gly Tyr Gln Val  
1715 1720 1725

Lys Gly Ser Gln Arg Thr Ile Asp Gly Thr Leu Tyr Ser Phe Asp Ser  
1730 1735 1740

Gln Thr Gly Asn Gln Lys Arg Val Gln Thr Thr Leu Leu Pro Gln Ala  
1745 1750 1755 1760

Gly His Tyr Ile Thr Lys Asn Gly Asn Asp Trp Gln Tyr Asp Thr Asn  
1765 1770 1775

Gly Glu Leu Ala Lys Gly Leu Arg Gln Asp Ser Asn Gly Lys Leu Arg  
1780 1785 1790

Tyr Phe Asp Leu Thr Thr Gly Ile Gln Ala Lys Gly Gln Phe Val Thr  
1795 1800 1805

Ile Gly Gln Glu Thr Tyr Tyr Phe Ser Lys Asp His Gly Asp Ala Gln  
1810 1815 1820

ES 2 726 941 T3

Leu Leu Pro Met Val Thr Glu Gly His Tyr Gly Thr Ile Thr Leu Lys  
 1825 1830 1835 1840  
 Gln Gly Gln Asp Thr Lys Thr Ala Trp Val Tyr Arg Asp Gln Asn Asn  
 1845 1850 1855  
 Thr Ile Leu Lys Gly Leu Gln Asn Ile Asn Gly Thr Leu Gln Phe Phe  
 1860 1865 1870  
 Asp Pro Tyr Thr Gly Glu Gln Leu Lys Gly Gly Val Ala Lys Tyr Asp  
 1875 1880 1885  
 Asp Lys Leu Phe Tyr Phe Glu Ser Gly Lys Gly Asn Leu Val Ser Thr  
 1890 1895 1900  
 Val Ala Gly Asp Tyr Gln Asp Gly His Tyr Ile Ser Gln Asp Gly Gln  
 1905 1910 1915 1920  
 Thr Arg Tyr Ala Asp Lys Gln Asn Gln Leu Val Lys Gly Leu Val Thr  
 1925 1930 1935  
 Val Asn Gly Ala Leu Gln Tyr Phe Asp Asn Ala Thr Gly Asn Gln Ile  
 1940 1945 1950  
 Lys Asn Gln Gln Val Ile Val Asp Gly Lys Thr Tyr Tyr Phe Asp Asp  
 1955 1960 1965  
 Lys Gly Asn Gly Glu Tyr Leu Phe Thr Asn Thr Leu Asp Met Ser Thr  
 1970 1975 1980  
 Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn His Asp Ser Ser Ser  
 1985 1990 1995 2000  
 Phe Asp His Thr Val Asp Gly Phe Leu Thr Ala Asp Thr Trp Tyr Arg  
 2005 2010 2015  
 Pro Lys Ser Ile Leu Ala Asn Gly Thr Thr Trp Arg Asp Ser Thr Asp  
 2020 2025 2030  
 Lys Asp Met Arg Pro Leu Ile Thr Val Trp Trp Pro Asn Lys Asn Val  
 2035 2040 2045  
 Gln Val Asn Tyr Leu Asn Phe Met Lys Ala Asn Gly Leu Leu Thr Thr  
 2050 2055 2060  
 Ala Ala Gln Tyr Thr Leu His Ser Asp Gln Tyr Asp Leu Asn Gln Ala  
 2065 2070 2075 2080  
 Ala Gln Asp Val Gln Val Ala Ile Glu Arg Arg Ile Ala Ser Glu His  
 2085 2090 2095  
 Gly Thr Asp Trp Leu Gln Lys Leu Leu Phe Glu Ser Gln Asn Asn Asn  
 2100 2105 2110  
 Pro Ser Phe Val Lys Gln Gln Phe Ile Trp Asn Lys Asp Ser Glu Tyr  
 2115 2120 2125  
 His Gly Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr Leu Lys Tyr Gly  
 2130 2135 2140

ES 2 726 941 T3

Asn Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr Arg Gln Pro Gly  
 2145 2150 2155 2160

Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro  
 2165 2170 2175

Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr Leu Met Asn Phe  
 2180 2185 2190

Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe Asp Ser Ile Arg  
 2195 2200 2205

Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile Gln Arg Thr Tyr  
 2210 2215 2220

Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln Ser Glu Ala Lys Ala  
 2225 2230 2235 2240

Asn Gln His Ile Ser Leu Val Glu Ala Gly Leu Asp Ala Gly Thr Ser  
 2245 2250 2255

Thr Ile His Asn Asp Ala Leu Ile Glu Ser Asn Leu Arg Glu Ala Ala  
 2260 2265 2270

Thr Leu Ser Leu Thr Asn Glu Pro Gly Lys Asn Lys Pro Leu Thr Asn  
 2275 2280 2285

Met Leu Gln Asp Val Asp Gly Gly Thr Leu Ile Thr Asp His Thr Gln  
 2290 2295 2300

Asn Ser Thr Glu Asn Gln Ala Thr Pro Asn Tyr Ser Ile Ile His Ala  
 2305 2310 2315 2320

His Asp Lys Gly Val Gln Glu Lys Val Gly Ala Ala Ile Thr Asp Ala  
 2325 2330 2335

Thr Gly Ala Asp Trp Thr Asn Phe Thr Asp Glu Gln Leu Lys Ala Gly  
 2340 2345 2350

Leu Glu Leu Phe Tyr Lys Asp Gln Arg Ala Thr Asn Lys Lys Tyr Asn  
 2355 2360 2365

Ser Tyr Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu Thr Asn Lys Asp  
 2370 2375 2380

Thr Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln Asp Asp Gly Gln  
 2385 2390 2395 2400

Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu Val Ser Leu Met  
 2405 2410 2415

Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr Met Ser Val Asp  
 2420 2425 2430

Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys Asp Ala Met Thr  
 2435 2440 2445

Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu Gly Leu Gly Val  
 2450 2455 2460

ES 2 726 941 T3

Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn Asp Ser Asp Lys Val  
 2465 2470 2475 2480

Thr Leu Asp Met Gly Ala Ala His Lys Asn Gln Lys Tyr Arg Ala Val  
 2485 2490 2495

Ile Leu Thr Thr Arg Asp Gly Leu Ala Thr Phe Asn Ser Asp Gln Ala  
 2500 2505 2510

Pro Thr Ala Trp Thr Asn Asp Gln Gly Thr Leu Thr Phe Ser Asn Gln  
 2515 2520 2525

Glu Ile Asn Gly Gln Asp Asn Thr Gln Ile Arg Gly Val Ala Asn Pro  
 2530 2535 2540

Gln Val Ser Gly Tyr Leu Ala Val Trp Val Pro Val Gly Ala Ser Asp  
 2545 2550 2555 2560

Asn Gln Asp Ala Arg Thr Ala Ala Thr Thr Thr Glu Asn His Asp Gly  
 2565 2570 2575

Lys Val Leu His Ser Asn Ala Ala Leu Asp Ser Asn Leu Ile Tyr Glu  
 2580 2585 2590

Gly Phe Ser Asn Phe Gln Pro Lys Ala Thr Thr His Asp Glu Leu Thr  
 2595 2600 2605

Asn Val Val Ile Ala Lys Asn Ala Asp Val Phe Asn Asn Trp Gly Ile  
 2610 2615 2620

Thr Ser Phe Glu Met Ala Pro Gln Tyr Arg Ser Ser Gly Asp His Thr  
 2625 2630 2635 2640

Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe Thr Asp Arg Tyr  
 2645 2650 2655

Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr Asp Gly Asp Leu  
 2660 2665 2670

Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met Gln Val Met Ala  
 2675 2680 2685

Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly Lys Glu Val Val  
 2690 2695 2700

Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp Asp Ala Thr Gly Phe  
 2705 2710 2715 2720

Gly Thr Gln Leu Tyr Val Thr Asn Ser Val Gly Gly Gly Gln Tyr Gln  
 2725 2730 2735

Glu Lys Tyr Ala Gly Gln Tyr Leu Glu Ala Leu Lys Ala Lys Tyr Pro  
 2740 2745 2750

Asp Leu Phe Glu Gly Lys Ala Tyr Asp Tyr Trp Tyr Lys Asn Tyr Ala  
 2755 2760 2765

Asn Asp Gly Ser Asn Pro Tyr Tyr Thr Leu Ser His Gly Asp Arg Glu  
 2770 2775 2780



ES 2 726 941 T3

gatatgtcta	ctaagtcttt	ttctaccaa	aatggtgcat	tcaatcatga	cagtagcagt	60
ttcgaccata	ctggtgatgg	cttcttgacg	gcagatactt	ggtatcgacc	aaagtcaatt	120
ttggctaacg	ggacaacttg	gcgtgattcg	actgataagg	atatgcgacc	attaatcact	180
gtttggtggc	caaataagaa	tgttcaagtc	aactacctca	acttcatgaa	agcaaatggc	240
ttggtgacaa	cagcagcaca	atacacacta	cattcagatc	aatatgattt	gaaccaagct	300
gcacaagatg	ttcaagtggc	cattgaaagg	cgcattgcgt	cagagcatgg	cacagactgg	360
ttacagaaat	tggtggttga	atcacaaaat	aataacccat	catittgtgaa	gcaacaattc	420
at ttggaaca	aggattctga	atatcatggg	ggtggtgatg	cttggttcca	aggtggttat	480
ctgaagtatg	gcaataaccc	actcacacca	acaactaatt	ctgattatcg	tcaacctggt	540
aatgcatttg	at t t c t t g c t	agccaacgac	gtggataatt	ctaactcctgt	tgtgcaagct	600
gaaaacttaa	actggttaca	ttacttaatg	aactttggca	ccatcactgc	gggtcaagat	660
gacgctaatt	ttgatagtat	tcgtattgac	gctgtcgact	ttattcataa	tgatacaatc	720
caacgtactt	atgattatct	tcgtgatgct	tatcaagtcg	aacaaagtga	agccaaagca	780
aaccagcaca	tttcattggg	tgaagctggc	ttagacgcag	gtacatcaac	gattcataat	840
gatgctgtaa	ttgagtcaaa	cctccgtgaa	gcagcgacat	tgtcgttaac	aaatgaacct	900
ggtaaaaata	aaccattgac	gaatatgcta	caagacgttg	acggcgggtac	gcttatcacc	960
gaccataaag	agaatagtag	agaaaatcag	gcgacaccaa	actattcaat	tattcacgcg	1020
cacgataaag	gtgtgcaaga	aaaagttagg	gcagccatta	ctgatgctac	tggtgctgat	1080
tggacgaact	ttacagatga	acagttaaaa	gccggattag	agctattcta	taaggatcag	1140
cgcgcaacaa	acaaaaagta	taatagttat	aacataccea	gtatttatgc	cctgatggtg	1200
acaaacaaag	atactgttcc	tcgtatgtat	tatggggata	tgtatcaaga	tgacggacag	1260
tatatggcaa	acaagagtat	ctactatgat	gccttagtgt	cattaatgac	ggctcgtaaa	1320
agctatgtca	gcggtggtca	aactatgagt	ggtgacaatc	atgggttggt	gaagagtgtc	1380
cgttttggaa	aagatgcatg	gacagctaat	gatttaggta	catcagctac	gcgtactgag	1440
ggtcttggtg	tcattattgg	taatgatcca	aagttgcaac	ttaatgattc	ggataaagtg	1500
acactggata	tgggtgcagc	acataaaaaat	caaaagtatc	gcgcagttat	cttaacaaca	1560
cgtgatgggt	tggcaacctt	taattcagat	caagcaccaa	cagcttggac	aaacgatcaa	1620
ggaacgttaa	cattctcaaa	tcaagagatt	aacgggcaag	acaatacaca	aattcgtggt	1680
ggtgctaatc	cgcaagtttc	tggttatcta	gctggttggg	tgccctggtg	tgcatcagac	1740
aatcaagatg	cccgtacagc	agcaacgaca	acagaaaatc	atgatggtaa	agtattcac	1800
tcgaatgctg	cattagattc	taaccttatt	tatgaagggt	tctctaactt	ccaacctaa	1860
gcaacaacgc	atgatgaact	tacgaacggt	gtaattgcta	aaaatgccga	tgtcttcaat	1920
aattggggta	ttacgagttt	tgaaatggca	ccacagtacc	ggtcaagtgg	ggaccataca	1980
ttcttggtat	caacgattga	taatggttat	gccttcactg	atcgctatga	cttaggtttc	2040
aatacaccaa	caaagtatgg	cactgatggg	gatttgcgtg	caacgattca	agcgctacat	2100
catgctaata	tgcaagttat	ggctgacggt	ggtgataacc	aggtctataa	cttacctggt	2160
aaagaagttg	tttcagcaac	acgagcaggt	gtttatggta	atgacgacgc	cacgggcttt	2220
ggaacgcaac	tctatgtgac	taactccggt	ggtggtggtc	aataccaaga	gaaatatgct	2280
ggacaatact	tagaagctct	gaaagcaaag	tatccagacc	tctttgaggg	taaggcctat	2340
gattattggt	ataagaacta	tgcaaatgat	gggtcaaate	cttactatac	attgtcacac	2400
ggtgaccgtg	aatctatccc	agcagatggt	gctattaagc	aatggtcagc	taagtatatg	2460
aacggcacga	acgttttggg	caatggtatg	ggttatgtat	tgaaggattg	gcataatggt	2520
caatatttca	agcttgatgg	tgataaatca	acattacctc	aaatttaa		2568

<210> 4

<211> 8506

5

<212> ADN

<213> Secuencia que codifica DSR-E

<400> 4

ES 2 726 941 T3

atgagagaca tgagggtaat ttgtgaccgt aaaaaattgt acaaatcggg caaagtacta 60  
 gtaacagccg gtatttttgc tttgatgatg tttggcgtca caactgctag tgttagtgca 120  
 aatacgattg cagttgacac gaatcatagc cgtacttcag cacagattaa taagagtgcc 180  
 gttgataagg ttaatgatga caagactact ttaggagcgg caaaagtagt ggcagtagcc 240  
 acaacgccag cgacaacggt agcagataaa acagtaagtg caccgcgagc agataaggca 300  
 gtagatacaa cgtcatcaac gacacctgca acggataagg cagtagatac aacgccaacg 360  
 acacctgcag cagataaggc agtagataca acgccaacga cacctgcagc agataaggca 420  
 gtagatacaa cgccaacgac acctgcagca aataaagcag tagatacaac gccagcgacc 480  
 gctgcaacag ataaggcggg agccacgcca gccacacctg cagcagataa gctagcaaat 540  
 acgacgcctg caacggacaa ggtagtagcc acaacgccag cgacgccggg agcaataaaa 600  
 gcagcagaca cgagtagtat tcatgatcaa ccattagata caaatgtgcc aactgataaa 660  
 tcagcaaaacc tcgtctcgac aacacaaaaa agtacggata atcaacaagt taagtctaca 720  
 gaaacatctc atcttcaaga aatcaacggg aaaacctatt ttcttgacga caatgggtaa 780  
 gttaaaaaga acttcaccgc tattattgac ggtaaagtgc tatactttga taaaacatcc 840  
 ggccaattga ccgcaaatgc accgcaagtt actaagggat tagtaaataat tgataatgca 900  
 cataacgcgg ctcgatgatc cacagctgat aacttcacaa atgtcgtatg ttacttaaca 960  
 gctaacagtt ggtatcgtcc taaggacatc ttaaaaaacg gaacgacctg gacaccaaca 1020  
 acagcagaag attttcgacc attgctcatg tcttgggtgc cggataagaa tacgcaggta 1080  
 gcttatctac aatatatgca atcagttggg atgctacctg acgatgttaa agtatcaaat 1140  
 gatgataata tgagcacatt gactgatgct gctatgactg ttcaaaagaa tatcgaatcg 1200  
 cgaattgggtg tatctggaaa aactgattgg ctcaagcaag atatgaacaa actgattgat 1260  
 tcacaggcaa attggaatat tgatagttaa tcaaagggta atgatcattt acagggtggg 1320  
 gcattgttat atgtgaatga tgacaaaaca cctaacgcga actcagatta ccgtctgtta 1380  
 aaccgtacac caaccaacca aaccggccaa attactgatc caagtaacaa aggtggatat 1440  
 gagatgttat tagctaataa tggttgataa tctaaccctg ttgtacaagc tgagcaattg 1500  
 aactggcttc actacatgat gaacattggg actatagctc agaacgacc aacagctaat 1560  
 tttgacgggt atcgtgttga tgcggttgat aacgttgatg ccgatctctt acaaatgtct 1620  
 ggtgattact ttaaagctgc atacggtagt ggtaaaactg aggcaaacgc aaacaatcat 1680  
 atttcgatct tggaaagatt ggataaataat gattctgcgt acattaaagc ccacgggaat 1740  
 aaccaattga caatggattt tccagcacac ttggctttga aatacgcctt gaacatgcct 1800  
 ctggccgcac aaagtggcct agaaccgcta ataatataca gtcttgtaa gcgtgggaaa 1860  
 gatgccacag aaaatgaagc acaaccaaac tatgctttaa tccgtgccca tgatagttaa 1920  
 gtgcagaccg ttattgcaca aattattaag gataaaatta acacaaaatc agacggctta 1980  
 actgtaacac cagatgagat taagcaagct ttactattt acaacgccga tgaattaaaa 2040  
 gcagataagg aatatacagc atacaatatt cctgcttctt acgctgtatt gttgacaaac 2100  
 aaggatactg tgccacgtgt ttattatggg gatctatttt ctgatgatgg acagtatatg 2160  
 tcacagaagt caccatacta tgacgccatt acgtcacttt tgaaaagccg tatcaaatat 2220  
 gttgctgggtg gtcaaagtat gaatatgacg tacttgcatg agtgctttga tccagcaaaa 2280  
 aatgagacaa agccacaagg tgtcttaaca tcagtagctt acggtaaagg tgcgatgacg 2340  
 gctgacgatt tgggtaatag tgacacacgt caacaaggta ttggtttggg gattaataat 2400  
 aagccattct tgaattttaa tgatgatgaa caaattgtgc tcaatatggg tgctgctcac 2460  
 aaaaatcaag cttaccgacc acttatgttg acaacaaaat ctggcttcca aatttacgat 2520  
 aaggatgccg gagcgcagc tgtttatact aacgatgctg gtcaacttat ttttaagtca 2580  
 gatatggtct atggtgtcag caatccacag gtatctgggt attttgctgc atgggtacca 2640  
 gtcgggtcga gtgatagtca agatgctaga acacaaagca gccagtcaga aactaaggat 2700  
 ggcgatgtct atcattcaaa tgctgcgctt gattctaata tgatttatga aggcttctcg 2760  
 aatttccaag caatgcctga aaagaatgat gacttcacca acgtaaaaaat tgctcaaaaat 2820  
 gctaaattgt ttaaagattt agggattaca agctttgaat tagcaccgca atatcgttca 2880  
 agtacagata atagtttttt ggattcgggt atccaaaacg gctatgcctt tactgatcga 2940  
 tatgatgttg gctataatac gccacaaaaa tatggtacag ttgatcaact tctagatagt 3000

ES 2 726 941 T3

ctaagagcat tacacgcaca aggtattcag gctattaatg actgggtacc tgatcaaatt 3060  
 tataatttac ctggcgaaca aatcgtcacc gcagttcgta caaatggttc aggtaagtac 3120  
 gattatgatt cagtgattaa taacacgctc tatgattcac gaacagttgg gggcggcgaa 3180  
 taccaagaaa agtttgggtg cctgttctta gaccagttga aaaaagatta tcctagcttg 3240  
 ittgaacta agcagatate aacgaatcag ccgatgaate cggatgtaa aattaaagaa 3300  
 tggctcgcaa agtactttaa tggttcaaac attcaaggtc gtggcgcttg gtatgtactt 3360  
 aaagactggg caacaaatca atatttcaat gtgtctagtg ataatggatt cttgcctaaa 3420  
 cagttactgg gtgaaaaaac aagcaccggc tttataacag aaaatggtaa gacttctttc 3480  
 tactcaacaa gtggttatca agctaaagat acctttattc aagatggaac aaattgggat 3540  
 tactttgata atgcaggcta tatgttgaca ggtaaaccaaa atatccacga taaaaattat 3600  
 tatttcttac ctaatgggtg ggaacttcaa gatgcttacc tttttgatgg taatcaagaa 3660  
 ttttactata ataaagctgg ggaacaagtt atgaaccagt attatcaaga tagtcaaaaat 3720  
 caatggcatt atttctttga aaatggctgc atggcaattg gcctgacaga agttccgaac 3780  
 gctgatggca cccatgttac acaatatttt gatgctaatt gtgtccaaat taaaggcaca 3840  
 gctataaaaag atcagaataa tcaattacgc tttttgatg aggccacagg taatatgggtg 3900  
 gttaattcat ggggaacagtt agcagataag tcttggcttt acctaatgc acaaggcgtt 3960  
 gctgtgactg gtaacccaaa aattgatggg gaagagtact acttcaatgc tgatggtaag 4020  
 caagttaaag gcaatgcaat catcgataat aatgggtgat aacgttatta tgatgggtgat 4080  
 aagggtgtca tggtagttaa ttcattgggt gagttgccag atggctcatg gttatatattg 4140  
 aatgacaaag gtattgctgt aacaggccgt caagtcatta ataatcaagt taatttcttt 4200  
 ggtaaatgag gtaagcaaat caaagatgcc tttaaattat tatccgatgg taattgggtg 4260  
 ttttgggatg ataagggcct gataacaact ggagccaaaag ttatcaatgg tctaataatg 4320  
 ttttttgata aagacggtca tcaaatcaaaa ggtgatgcca gcacggatgc caatggtaag 4380  
 cgccattatt atgacaaaaa tgatgggtcat cttgtcacia attcattgggg tgagttgcca 4440  
 gatggttcat ggttatatct agaagaacaa ggtgatgctg ttactgggca acgtgtgatt 4500  
 gatggcaaga cacgctatatt tgatgaagat ggcaaacaaa ttaaaaatag ctaaaaaacg 4560  
 ctggccaatg gcgataagat ttatcttgat ggtgatgggg ttgctgcaac aggcttacia 4620  
 catgtggggc ataaaaatcat gtatcttgat gaagatggca aacaagttgt tggcaagttt 4680  
 gtatcagcaa aagatggttc atgggtattac ttaaatcagg atgggtgtgc cgcggttggg 4740  
 ccaagcagca ttaatggaca atcactttac tttgatcaag atggtaaaa agttaaatat 4800  
 aatgaagttc gtaaatagtga tggaaacaacc aactattaca caggattaac ggggtaaaaag 4860  
 ttaacgcaag acttcggtga actaccagat ggttcatgga tttatcttga ttcgcaagggt 4920  
 catacagtaa ctgggtgcaca aatcattaac ggtcaaaaatc tttactttaa ggctgacggc 4980  
 cagcaagtta aagggtcatgc ttatactgac caattagggtc atatgcgttt ttatgatcct 5040  
 gattcagggtg atatgttgag taatcgcttt gaacaaatca cacctgggtg atgggcttac 5100  
 tttgggtgctg atgggtgtggc cataactgga caacatgaca taaatgggca gaagctattc 5160  
 tttgatgaga caggatatca agttaaagggt tgcgaacgta caatagatgg tacgttatac 5220  
 agcttcgatt ctcaaactgg taacccaaaaa cgcgtacaga caacattgtt gccacaagca 5280  
 ggtcactata tcacgaaaaa tggtaacgat tggcagtatg ataccaatgg tgaactagcg 5340  
 aagggtctgc gtaagatag caatggtgag ttgcgttact ttgatttgac aaccggcata 5400  
 caagcgaaaag gccaatttgt tacaattggc caagaaactt attactttag taaagatcac 5460  
 ggggatgccc agttattgccc aatgggtcact gaagggcatt acggtacaat aacactcaag 5520  
 caaggtcaag acacccaaaac agcctgggtt tacctgtgatc aaaataatac tttttgaaag 5580  
 ggattgcaaa atatcaatgg cacgttgcaa ttctttgatc catatacagg tgaacaactt 5640  
 aagggtggcg tagcaaaagta tgacgacaag ctcttttact ttgaatcagg taaaggtaat 5700  
 cttgttagca ccgtagcagg tgactatcag gatgggtcatt atatttcca agatggccaa 5760  
 acacgttacg cagataagca aatcagctt gtaaagggtc ttgttactgt taatggggca 5820  
 ttacaatact ttgataacgc tactggtaac caaataaaaa atcaacaagt tattgttgat 5880  
 ggcaagacgt actattttga cgataaaggc aatgggtgat acttattcac taatacatta 5940  
 gatatgtcta ctaatgcctt ttctaccaa aatggtgcat tcaatcatga cagtgcagt 6000  
 ttcgaaccata ctggtgatgg ctcttgacg ccagataact ggtatcgacc aaagtcaatt 6060  
 ttggctaacg ggacaacttg gcgtgattcg actgataagg atatgcgacc attaatact 6120  
 gtttgggtggc caaataagaa tgttcaagtc aactacctca acttcatgaa agcaaatggc 6180  
 ttgttgacaa cagcagcaca atacacacta cattcagatc aatatgattt gaaccaagct 6240  
 gcacaagatg ttcaagtggc cattgaaagg cgcattgctg cagagcatgg cacagactgg 6300  
 ttacagaaat tgttgtttga atcacaataa aataaccat catttgtgaa gcaacaattc 6360  
 atttggaaaca aggattctga atatcatggg ggtgggtgat ctgggtcca aggtgggttat 6420  
 ctgaagtatg gcaataaccc actcacacca acaactaatt ctgattatcg tcaacctggg 6480  
 aatgcatttg atttcttget agccaacgac gtggataatt ctaatcctgt tgtgcaagct 6540  
 gaaaacttaa actgggtaca ttacttaatg aactttggca ccatcactgc ggggtcaagat 6600  
 gacgctaatt ttgatagtat tcgtattgac gctgtcgact ttattcataa tgatacaatc 6660

ES 2 726 941 T3

```

caacgtactt atgattatct tcgtgatgct tatcaagtgc aacaaagtga agccaaagca 6720
aaccagcaca tttcattggg tgaagctggc ttagacgcag gtacatcaac gattcataat 6780
gatgcgttaa ttgagtcaaa cctccgtgaa gcagcgacat tgcggttaac aaatgaacct 6840
ggtaaaaata aaccattgac gaatatgcta caagacgttg acggcggtag cttatcacc 6900
gaccatacgc agaatagtac agaaaatcag gcgacaccaa actattcaat tattcacgcy 6960
cacgataaag gtgtgcaaga aaaagtaggt gcagccatta ctgatgctac tggtgctgat 7020
tggacgaact ttacagatga acagttaaaa gccggattag agctattcta taaggatcag 7080
cgcgcaacaa acaaaaagta taatagttat aacataccaa gtattttatgc cctgatgttg 7140
acaaacaaag atactgttcc tcgtatgtat tatggggata tgtatcaaga tgacggacag 7200
tatatggcaa acaagagtat ctactatgat gccttagtgt cattaatgac ggctcgtaaa 7260
agctatgtca gcggtgggca aactatgagt gttgacaatc atggtttggt gaagagtgtc 7320
cgttttggaa aagatgcgat gacagcta at gatttaggta catcagctac gcgtactgag 7380
. ggtccttggg t cattattgg taatgatcca aagttgcaac ttaatgattc ggataaagtg 7440
acactggata tgggtgcaag acataaaaaat caaaagtatc gcgcagttat cttacaaca 7500
cgtgatgggt tggcaacctt taattcagat caagcaccaa cagcttgagc aaacgatcaa 7560
ggaacgttaa cttctcctaaa tcaagagatt aacgggcaag acaatacaca aattcgtggg 7620
gttgctaatc cgcaagtttc tggttatcta gctgtttggg tgcctgtggg tgcatcagac 7680
aatcaagatg cccgtacagc agcaacgaca acagaaaatc atgatggtaa agtattacac 7740
tcgaatgcgg cattagattc taaccttatt tatgaagggt tctctaactt ccaacctaa 7800
gcaacaacgc atgatgaact tacgaacggt gtaattgcta aaaatgccga tgccttcaat 7860
aattggggta ttacgagttt tgaaatggca ccacagtacc gttcaagtgg ggaccataca 7920
ttcttggatt caacgattga taatggttat gccttcaactg atcgctatga cttaggtttc 7980
aatacaccaa caaagtatgg cactgatggg gatttgctg caacgattca agcgtacat 8040
catgctaata tgcaagttat ggctgacgtt gttgataacc aggtctataa cttacctggg 8100
aaagaagttg tttcagcaac acgagcaggt gtttatggta atgacgacgc cacgggcttt 8160
ggaacgcaac tctatgtgac taactccgtt ggtgggtggtc aataccaaga gaaatagct 8220
ggacaatact tagaagctct gaaagcaaag tatccagacc tctttgaggg taaggcctat 8280
gattattggg ataagaacta tgcaaatgat gggtcaaact cttactatac attgtcacac 8340
ggtgaccgtg aatctatccc agcagatggt gctattaagc aatggtcagc taagtatatg 8400
aacggcacga acgttttggg caatggtatg ggttatgtat tgaaggattg gcataatggt 8460
caatatttca agcttgatgg tgataaatca acattacctc aaattt 8506

```

<210> 5  
 <211> 8931  
 <212> ADN  
 <213> Gen dsr-E  
  
 <400> 5

5

ES 2 726 941 T3

aataatctgt ctccattgct ttcaaaaataa taatagttaa ttattatcat ggaacaatca 60  
atattttatt tatattcact attgaatatic cttttttgca taaatctcta gagccgattt 120  
tttgggttat acaatgaatt ggtaaagggtt aatcattttt acaaaaacat ggtgggtttt 180  
tattttttct aaaattaccg aactagagga agagaaaagg agcaatagtt gtatgagaga 240  
catgagggta atttgtgacc gtaaaaaatt gtacaaatcg ggcaaagtac tagtaacagc 300  
cggatatttt gctttgatga tgtttggcgt cacaactgct agtgtagtg caaatagat 360  
tgcagttgac acgaatcata gccgtacttc agcacagatt aataagagtg ccgttgataa 420  
ggttaatgat gacaagacta ctttaggagc ggcaaaagta gtggcagtag ccacaacgcc 480  
agcgacaccg gtagcagata aaacagtaag tgcacccgca gcagataagg cagtagatac 540  
aacgtcatca acgacacctg caacggataa ggtagtagat acaacgcca cgcacctgc 600  
agcagataag gcagtagata caacgccaac gacacctgca gcagataagg cagtagatac 660  
aacgccaacg acacctgcag caaataaagc agtagataca acgcccagcga ccgctgcaac 720  
agataaggcg gtagccacgc cagccacacc tgcagcagat aagctagcaa atacgacgcc 780  
tgcaacggac aaggcagtag ccacaacgcc agcgacgccg gtagcaaata aagcagcaga 840  
cacgagtagt attcatgac aaccattaga tacaatgtg ccaactgata aatcagcaaa 900  
cctcgtctcg acaacacaaa aaagtacgga taatcaacaa gtaagtcta cagaaacatc 960  
tcatcttcaa gaaatcaacg gtaaaaccta ttttcttgac gacaatggtc aagttaaaaa 1020  
gaacttcacc gctattattg acggtaaagt tctatacttt gataaaacat ccggcgaatt 1080  
gaccgcaaat gcaccgcaag ttactaaggg attagtaaatt attgataatg cacataacgc 1140  
ggctcatgat ctcacagctg ataacttcac aaatgtcgat ggttacttaa cagctaacag 1200  
ttggtatcgt cctaaggaca tcttaaaaaa cggaacgacc tggacaccaa caacagcaga 1260

ES 2 726 941 T3

agattttcga ccattgctca tgtcttggtg gccggataag aatacgcagg tagcttatct 1320  
 acaatataatg caatcagttg gtatgclacc tgacgatggt aaagtatcaa atgatgataa 1380  
 tatgagcaca ttgactgatg ctgctatgac tgttcaaaaag aatatcgaat cgcggaattgg 1440  
 tgtatctgga aaaactgatt ggctcaagca agatatgaac aaactgattg attcacaggc 1500  
 aaattggaat attgatagtg aatcaaaggg taatgatcat ttacaggggtg gggcattggt 1560  
 atatgtgaat gatgacaaaa cacctaacgc gaactcagat taccgtctgt taaaccgtac 1620  
 accaaccaac caaacccggcc aaattactga tccaagtaaa caaggtggat atgagatggt 1680  
 attagctaata gatggtgata attctaaccct tgttgataca gctgagcaat tgaactgggt 1740  
 tcaactacatg atgaacattg gtactatagc tcagaacgac ccaacagcta attttgacgg 1800  
 ttatcgtggt gatgcgggtg ataacggtga tgccgatctc ttacaaatg ctggtgatta 1860  
 ctttaagct gcatacggta ctggtaaaaac tgaggcaaac gcaacaatc atatttcgat 1920  
 cttggaagat tgggataata atgattctgc gtacattaaa gccacggga ataaccaatt 1980  
 gacaatggat tttccagcac acttggtctt gaaatacggc ttgaacatgc ctcttgccgc 2040  
 acaaagtggc ctagaaccgc taattaatac aagtcttggt aagcgtggga aagatgccac 2100  
 agaaaatgaa gcacaaccaa actatgcctt tatccgtgcc catgatagtg aagtgcagac 2160  
 cgttattgca caaattatta aggataaaat taacacaaaa tcagacggct taactgtaac 2220  
 accagatgag attaagcaag ctttcaactat ttacaacgcc gatgaattaa aagcagataa 2280  
 ggaatataca gcatacaata ttccigtctc ttacgctgta ttggtgacaa acaaggataa 2340  
 tggccacgtg gtttattatg gtgatctatt ttctgatgat ggacagtata tgtcacagaa 2400  
 gtcaccatac tatgacgcca ttacgtcact tttgaaaagc cgtatcaaat atggtgctgg 2460  
 tggtaaaagt atgaatatga cgtacttgca tgagtgtctt gatccagcaa aaaatgagac 2520  
 aaagccacaa ggtgtcttaa catcagtagc ttacggtaaa ggtgcgatga cggctgacga 2580  
 tttgggtaat agtgacacac gtcaacaagg tattggtttg gtgattaata ataagccatt 2640  
 cttgaattta aatgatgatg aacaaattgt gctcaatatg ggtgctgctc acaaaaatca 2700  
 agcttaccga ccacttatgt tgacaacaaa atctggtctt caaatttaccg ataaggatgc 2760  
 cggagcgcca gttgtttata ctaacgatgc tggccaactt atttttaagt cagatatggt 2820  
 ctatggtgct agcaatccac aggtatctgg ttattttgct gcatgggtac cagtcgggtc 2880  
 gatgatagtg caagatgcta gaacacaaaag cagccagtca gaaactaag atggcgtgt 2940  
 ctatcattca aatgctgcgc ttgattctaa tgtgatttat gaaggcttct cgaatttcca 3000  
 agcaatgcct gaaaagaatg atgacttcac caacgtaaaa attgctcaaa atgctaaatt 3060  
 gtttaaaagt ttagggatta caagctttga attagcaccg caatatcgtt caagtacaga 3120  
 taatagtttt ttggattcgg ttatccaaaa cggctatgcc tttactgatc gatatgatgt 3180  
 tggctataat acgccaacaa aatatggtac agttgatcaa cttctagata gtctaagagc 3240  
 attacacgca caaggtatctc aggtattaa tgactgggta cctgatcaaa tttataatct 3300  
 acctggcgaa caaatcgtca ccgcagttcg tacaatgggt tcaggtaatg acgattatga 3360  
 ttcagtgatt aataacacgc tctatgattc acgaacagtt gggggcggcg aataccaaga 3420  
 aaagtttggg ggcctgttct tagaccagtt gaaaaaagat tatcctagct tggttgaaac 3480  
 taagcagata tcaacgaatc agccgatgaa tccggatggt aaaattaaag aatggtctgc 3540  
 aaagtacttt aatggttcaa acattcaagg tctggtgctc tggtatgtac ttaaagactg 3600  
 ggcaacaaat caatatttca atgtgtctag tgataatgga ttcttgctta aacagttact 3660  
 gggtgaaaaa caaagcaccg gctttataac agaaaatgggt aagacttctt tctactcaac 3720  
 aagtgttat caagctaaag atacccttat tcaagatgga acaaattgggt attactttga 3780  
 taatgcaggc tatatgttga caggtaaaca aaatatccac gataaaaatt attatttctt 3840  
 acctaatggt gtggaacttc aagatgctta cctttttgat ggtaatacaag aattttacta 3900  
 taataaagct ggggaacaag ttatgaacca gtattatcaa gatagtcaaa atcaatggca 3960  
 ttatttcttt gaaaatggtc gcatggcaat tggcctgaca gaagttccga acgctgatgg 4020  
 cacccatggt acacaatatt ttgatgctaa tgggtgccaa attaaaggca cagctataaa 4080  
 agatcagaat aatcaattac gctattttga tgaggccaca ggtaatatgg tggttaattc 4140  
 atggggacag ttagcagata agtcttggtt ttaccttaat gcacaaggcg ttgctgtgac 4200  
 tggtaaccba aaaattgatg gtgaagagta ctacttcaat gctgatggta agcaagttaa 4260  
 aggcaatgca atcatcgata ataatggtga tcaacgttat tatgatgggtg ataagggtgt 4320  
 catggtagtt aattcatggg gtgagttgcc agatggctca tggttatatt tgaatgacaa 4380  
 aggtattgct gtaacaggcc gtcaagtcac taataatcaa gtttaattct ttggtaatga 4440  
 tggtaagcaa atcaaagatg cctttaaatt attatccgat ggttcatggg tgtatttga 4500  
 tgataaagggc ctgataacaa ctggagccaa agttatcaat ggtctaaata tgtttttga 4560  
 taaagacggg catcaaatca aaggtgatgc cagcacggat gccaatggta agcgcatta 4620  
 ttatgacaaa aatgatggtc atctgtcac aaattcatgg ggtgagttgc cagatggttc 4680  
 atggttatat ctagaagaac aaggtgatgc tgttactgggt caacgtgtga ttgatggcaa 4740  
 gacacgctat tttgatgaag atggcaacaa aattaaat agcctaaaaa cgctggccaa 4800  
 tggcgataag atttatcttg atggtgatgg ggttgctgca acaggcttac aacatgtggg 4860  
 cgataaaaatc atgtattttg atgaagatgg caacaagtt gttggcaagt ttgtatcagc 4920

ES 2 726 941 T3

aaaagatggt tcatggtatt actttaatca ggatgggtgtt gccgcggttg gtccaagcag 4980  
 cattaatgga caatcacttt actttgatca agatgggtaaa caagttaaat ataatgaagt 5040  
 tcgtaatagt gatggaacaa ccaactatta cacaggatta acgggtgaaa agttaacgca 5100  
 agacttcggt gaactaccag atggttcatg gatttatctt gatgcgcaag gtcatacagt 5160  
 aactgggtgca caaatcatta acgggtcaaaa tctttacttt aaggctgacg gccagcaagt 5220  
 taaaggtcat gcttatactg accaattagg tcatatgctt ttttatgatc ctgattcagg 5280  
 tgatatgttg agtaatcgct ttgaacaaat cacacctggg gtatgggctt actttgggtc 5340  
 tgatgggtgtg gccataactg gacaacatga cataaatggg cagaagctat tctttgatga 5400  
 gacaggatat caagttaaag gttcgcaacg tacaatagat ggtacgttat acagcttcga 5460  
 ttctcaaaact ggtaaccaaa aacgcgtaca gacaacattg ttgccacaag caggctacta 5520  
 taccacgaaa aatggtaacg attggcagta tgaaccaat ggtgaactag cgaagggtct 5580  
 gcgtcaagat agcaatggta agttgctgta ctttgatttg acaaccggca tacaagcgaa 5640  
 aggccaattt gttacaattg gccaaagaac ttattacttt agtaaagatc acggggatgc 5700  
 gcagttattg ccaatggtca ctgaagggca ttacggtaca ataactctca agcaaggtca 5760  
 agacacccaaa acagcctggg tttaccgtga tcaaaataat actattttga agggattgca 5820  
 aaatatcaat ggcacggtgc aattctttga tccatataca ggtgaacaac ttaaggggtg 5880  
 cgtagcaaaag tatgacgaca agctctttta ctttgaatca ggtaaaggta atcttgttag 5940  
 caccgtagca ggtgactatc aggatgggtca ttatatttcc caagatggcc aaacacgtta 6000  
 cgagataaag caaaatcagc ttgtaaaggg acttgttact gttaatgggg cattacaata 6060  
 ctttgataac gctactggta accaaataaa aaactcaaca gttattgttg atggcaagac 6120  
 gtactatttt gacgataaag gcaatgggtga atacttattc actaatacat tagatagtct 6180  
 tactaatgct ttttctacca aaaatggtgc attcaatcat gacagtagca gtttcgacca 6240  
 tactgttgat ggcttcttga cggcagatac ttggtatcga ccaaagtcaa ttttggtctaa 6300  
 cgggacaact tggcgtgatt cgactgataa ggatatgcga ccattaatca ctgtttggtg 6360  
 gccaaaataag aatgttcaag tcaactacct caacttcatg aaagcaaatg gcttgttgac 6420  
 aacagcagca caatacacac tacattcaga tcaatatgat ttgaaccaag ctgcacaaga 6480  
 tgttcaagtg gccattgaaa ggcgcattgc gtcagagcat ggcacagact ggttacagaa 6540  
 attgttgttt gaatcacaaa ataataacc c atcattttgtg aagcaacaat tcatttggaa 6600  
 caaggattct gaatatcatg gtggtggtga tgcttggttc caagggtggtt atctgaagta 6660  
 tggcaataac ccactcacac caacaactaa ttctgattat cgtcaacctg gtaatgcatt 6720  
 tgatttcttg ctagccaacg acgtggataa ttctaactct gttgtgcaag ctgaaaactt 6780  
 aaactggtta cattacttaa tgaactttgg caccatcact gcgggtcaag atgacgctaa 6840  
 ttttgatagt attcgtattg acgctgtcga ctttatcoat aatgatacaa tccaacgtac 6900  
 ttatgattat ctctgtgatg cttatcaagt gcaacaaagt gaagccaaag caaacagca 6960  
 catttcattg gttgaagctg gcttagacgc aggtacatca acgattcata atgatgcgtt 7020  
 aattgagtca aacctcctg aagcagcgac attgtcgtta acaaatgaac ctggtaaaaa 7080  
 taaaccattg acgaatatgc tacaagacgt tgacggcggg acgcttatca ccgaccatac 7140  
 gcagaatagt acagaaaatc aggcgacacc aaactattca attattcacg cgcacgataa 7200  
 aggtgtgcaa gaaaaagtag gtgcagccat tactgatgct actgggtgctg attggacgaa 7260  
 ctttacagat gaacagttaa aagccggatt agagctattc tataaggatc agcgcgcaac 7320  
 aaacaaaaag tataatagtt ataacatacc aagtatttat gccctgatgt tgacaaaaca 7380  
 agatactgtt cctcgtatgt attatgggga tatgtatcaa gatgacggac agtatatggc 7440  
 aaacaagagt atctactatg atgccttagt gtcattaatg acggctcgtg aaagctatgt 7500  
 cagcgggtgt caaactatga gtgttgacaa tcatggtttg ttgaagagtg tccgttttgg 7560  
 aaaagatgct atgacagcta atgatttagg tacatcagct acgcgtactg agggctttgg 7620  
 tgtcattatt ggtaatgatc caaagttgca acttaatgat tcggataaag tgacactgga 7680  
 tatgggtgca gcacataaaa atcaaaagta tcgcgcagtt atcttaacaa cacgtgatgg 7740  
 tttggcaacc ttttaattcag atcaagcacc aacagcttgg acaaacgatc aaggaacggt 7800  
 aacattctca aatcaagaga ttaacgggca agacaataca caaattcgtg gtgttgctaa 7860  
 tccgcaagtt tctggttatc tagctggttg ggtgcctcag ggtgcatcag acaatcaaga 7920  
 tgcccgtaca gcagcaacga caacagaaaa tcatgatggg aaagtattac actcgaatgc 7980  
 ggcattagat tctaacctta tttatgaagg tttctctaac ttccaaccta aggcaacaac 8040  
 gcatgatgaa cttacgaaoc ttgtaattgc taaaaatgcc gatgtcttca ataattgggg 8100  
 tattacgagt tttgaaatgg caccacagta ccggtcaagt ggggaccata cattcttggg 8160  
 ttcaacgatt gataatgggt atgccttcac tgatcgctat gacttaggtt tcaatacacc 8220  
 aacaaaagtat ggcactgatg gtgatttgcg tgcaacgatt caagcgctac atcatgctaa 8280  
 tatgcaagtt atggctgacg ttggtgataa ccaggtctat aacttacctg gtaaagaagt 8340  
 tgtttcagca acacgagcag gtggttatgg taatgacgac gccacgggct ttggaacgca 8400  
 actctatgct actaaactccg ttggtggtgg tcaataccaa gagaaatag ctggacaata 8460  
 cttagaagct ctgaaagcaa agtatccaga cctctttgag ggtaaaggct atgattattg 8520  
 gtataagaac tatgcaaatg atgggtcaaa tcttactat acattgtcac acgggtgaccg 8580

ES 2 726 941 T3

```
tgaatctatc ccagcagatg ttgctattaa gcaatggcca gctaagtata tgaacggcac 8640
gaacgttttg ggcaatggta tgggttatgt attgaaggat tggcataatg gtcaatattt 8700
caagcttgat ggtgataaat caacattacc tcaaatttaa tttattttga tagggaacga 8760
ttatcttata aaattgtagt gacaaaagtc gcagatattg aatccaatat ctgcgacttt 8820
tcgtctgtaa agctatgcta taataacggt atgacaaaag aaaattattt taaagttggc 8880
acaattgtca acaccacggt tattcgtggc gaagtgaaga ttatggatat c 8931
```

5 <210> 6  
<211> 17  
<212> PRT  
<213> Péptidos

<400> 6

```
Ala Asn Trp Asn Ile Asp Ser Glu Ser Lys Gly Asn Asp His Leu Gln
  1           5           10           15
```

10 Gly

15 <210> 7  
<211> 24  
<212> PRT  
<213> Péptidos

<400> 7

```
Gly Gly Tyr Glu Met Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro
  1           5           10           15
```

20 Val Val Gln Ala Glu Gln Leu Asn  
20

25 <210> 8  
<211> 21  
<212> PRT  
<213> Péptidos

<400> 8

```
Ala Asn Phe Asp Gly Tyr Arg Val Asp Ala Val Asp Asn Val Asp Ala
  1           5           10           15
```

```
Asp Leu Leu Gln Ile
  20
```

30 <210> 9  
<211> 12  
<212> PRT  
<213> Péptidos

35 <400> 9

```
His Ile Ser Ile Leu Glu Asp Trp Asp Asn Asn Asp
  1           5           10
```

40 <210> 10  
<211> 15  
<212> PRT  
<213> Péptidos

45 <400> 10

ES 2 726 941 T3

Tyr Ala Phe Ile Arg Ala His Asp Ser Glu Val Gln Thr Val Ile  
 1 5 10 15

5 <210> 11  
 <211> 8  
 <212> PRT  
 <213> Péptidos  
 <400> 11

10 Asp Trp Val Pro Asp Gln Ile Tyr  
 1 5

15 <210> 12  
 <211> 19  
 <212> PRT  
 <213> Péptidos  
 <400> 12

Phe Ile Trp Asn Lys Asp Ser Glu Tyr His Gly Gly Gly Asp Ala Trp  
 1 5 10 15

20 Phe Gln Gly

25 <210> 13  
 <211> 24  
 <212> PRT  
 <213> Péptidos  
 <400> 13

Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro  
 1 5 10 15

Val Val Gln Ala Glu Asn Leu Asn  
 20

30 <210> 14  
 <211> 13  
 <212> PRT  
 <213> Péptidos  
 35 <400> 14

Ala Asn Phe Asp Ser Ile Arg Ile Asp Ala Val Asp Phe  
 1 5 10

40 <210> 15  
 <211> 8  
 <212> PRT  
 <213> Péptidos  
 45 <400> 15

His Ile Ser Leu Val Glu Ala Gly  
 1 5

<210> 16  
 <211> 8

ES 2 726 941 T3

<212> PRT  
<213> Péptidos

5

<400> 16

Tyr Ser Ile Ile His Ala His Asp  
1 5

<210> 17  
<211> 8  
<212> PRT  
<213> Péptidos

10

<400> 17

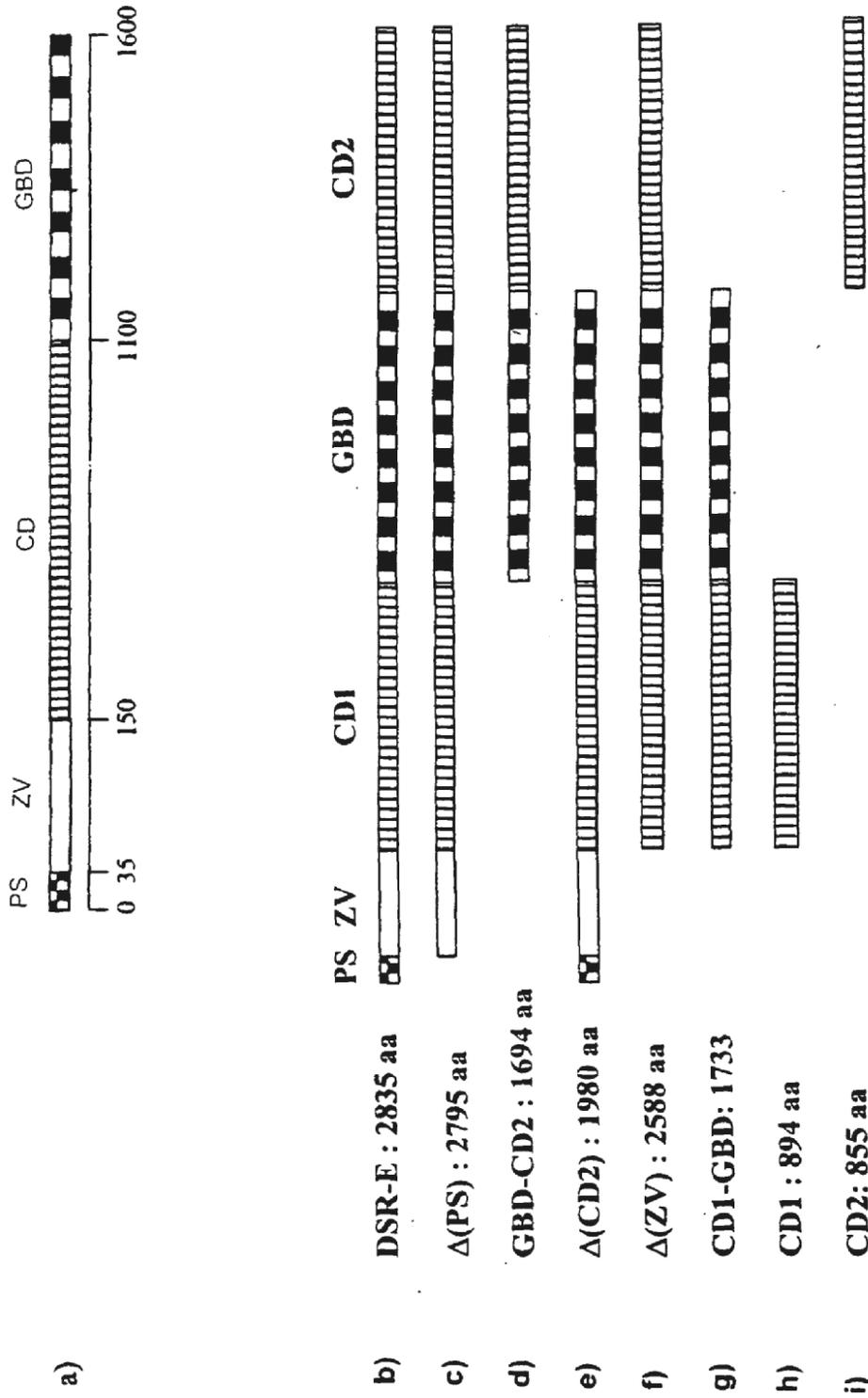
Asp Val Val Asp Asn Gln Val Tyr  
1 5

15

## REIVINDICACIONES

- 5 1. Polipéptido aislado que tiene una actividad enzimática de glucosiltransferasa capaz de formar dextranos que presentan ramificaciones  $\alpha(1 \rightarrow 2)$  a partir de sacarosa, de  $\alpha$ -D-fluoro-glucosa, de para-nitrofenil- $\alpha$ -D-glucopiranosido, de  $\alpha$ -D-glucopiranosido- $\alpha$ -D-sorbofuranósido o de 4-O- $\alpha$ -D-galactopiranosilsacarosa, estando caracterizado dicho polipéptido:
- 10 a. por que comprende la SEQ ID NO: 2, o  
b. por que comprende una estructura constituida, partiendo de la parte aminada del polipéptido, por una secuencia señal, una región variable, un primer dominio catalítico, un dominio de unión a glucano y un segundo dominio catalítico, en cuyo marco:
- 15 i) las secuencias de los dominios catalíticos primero y segundo de dicho polipéptido tienen un porcentaje de similitud con la SEQ ID NO: 1 comprendido entre el 65 % y el 100 %, conservándose la tríada catalítica Asp/Glu/Asp en las posiciones respectivas 230/268/342 de la SEQ ID NO: 1, y  
ii) el dominio de unión a glucano tiene un tamaño superior a 500 aminoácidos.
- 20 2. Polipéptido según la reivindicación 1, en el que los aminoácidos siguientes no varían con respecto a la secuencia de SEQ ID NO: 2:
- 25 W en las posiciones 425 y 2122,  
E en las posiciones 430, 565 y 2127, 2248,  
D en las posiciones 487, 489, 527, 638, 2170, 2172, 2210 y 2322,  
H en las posiciones 637 y 2321,  
Q en las posiciones 1019 y 2694.
- 30 3. Ácido nucleico aislado que codifica una enzima con actividad glucosiltransferasa capaz de formar dextranos que presentan ramificaciones  $\alpha(1 \rightarrow 2)$  a partir de sacarosa, de  $\alpha$ -D-fluoro-glucosa, de para-nitrofenil- $\alpha$ -D-glucopiranosido, de  $\alpha$ -D-glucopiranosido- $\alpha$ -D-sorbofuranósido o de 4-O- $\alpha$ -D-galactopiranosilsacarosa, estando caracterizado dicho ácido nucleico:
- 35 a) por que comprende o consiste en la SEQ ID NO: 4, o  
b) por que comprende, de la parte 5' a la parte 3', una secuencia que codifica un péptido señal, una secuencia variable, una secuencia que codifica un primer dominio catalítico, una secuencia que codifica un dominio de unión a glucano y una secuencia que codifica un segundo dominio catalítico, y:
- 40 i. las dos secuencias de ácido nucleico que codifican los dominios catalíticos tienen al menos un 80 % de identidad con la secuencia SEQ ID NO: 3, conservándose la codificación de la tríada catalítica Asp/Glu/Asp codificada por los tripletes GAT, GAA y GAT encontrados respectivamente en las posiciones 6628/6742/6964 de la SEQ ID NO: 4, y  
ii. la secuencia de ácido nucleico que codifica un dominio de unión a glucano codifica un dominio de unión a glucano que tiene un tamaño superior a 500 aminoácidos.
- 45 4. Ácido nucleico aislado según la reivindicación 3, que comprende una secuencia que tiene al menos un 80 % de identidad con la secuencia que codifica un dextrano sacarasa expresada por el plásmido pCR-7-*dsrD* depositado en la CNCM el 15 de marzo de 2001 con el número I-2649, en la que se conserva la codificación de la tríada catalítica Asp/Glu/Asp codificada por los tripletes GAT, GAA y GAT encontrados respectivamente en las posiciones 6628/6742/6964 de la SEQ ID NO: 4.
- 50 5. Vector de expresión que comprende un ácido nucleico según una cualquiera de las reivindicaciones 3 o 4.
6. Vector según la reivindicación 5, en el que el ácido nucleico está bajo el control de la secuencia que permite su expresión en células procariotas o eucariotas.
- 55 7. Célula huésped que comprende un ácido nucleico según la reivindicación 3 o 4 o un vector según una de las reivindicaciones 5 o 6 elegida de un grupo que comprende *E. coli*, los vegetales, los *Lactococcus*, los *Bacillus* y las levaduras.
- 60 8. Célula según la reivindicación 7, caracterizada por que se trata de una cepa de *E. coli* depositada en la CNCM el 15 de marzo de 2001 bajo el n.º I-2649.
9. Proceso de producción de un dextrano sacarasa capaz de formar las uniones  $\alpha 1 \rightarrow 2$  que comprende:
- 65 a) la inserción de un ácido nucleico según una de las reivindicaciones 3 o 4 o un vector según una de las reivindicaciones 5 o 6 en una célula huésped según la reivindicación 7;  
b) la purificación de la enzima a partir de un extracto celular.

10. Proceso según la reivindicación 9, en el que la célula huésped es una eucariota elegida de un grupo que comprende las levaduras, los hongos y los vegetales.
- 5 11. Proceso de obtención de una glucosiltransferasa aislada capaz de formar oligósidos o dextranos que presentan una tasa de uniones  $\alpha(1 \rightarrow 2)$  superior al 30 % que comprende:
- 10 - una etapa de modificación aleatoria de la secuencia SEQ ID NO: 4 y de establecimiento de un banco de variantes,  
- una etapa de expresión de esas secuencias modificadas en una célula huésped apropiada, un huésped que contiene una variante,  
- una etapa de cribado de los huéspedes que expresan una enzima capaz de formar más del 30 % de uniones  $\alpha(1 \rightarrow 2)$  en un sustrato apropiado,  
- una etapa de aislamiento del o de los genes mejorados.
- 15 12. Utilización de una glucosiltransferasa obtenida por un proceso según una de las reivindicaciones 9 o 10 en la fabricación de una composición farmacéutica o cosmética.



**Fig. 1**

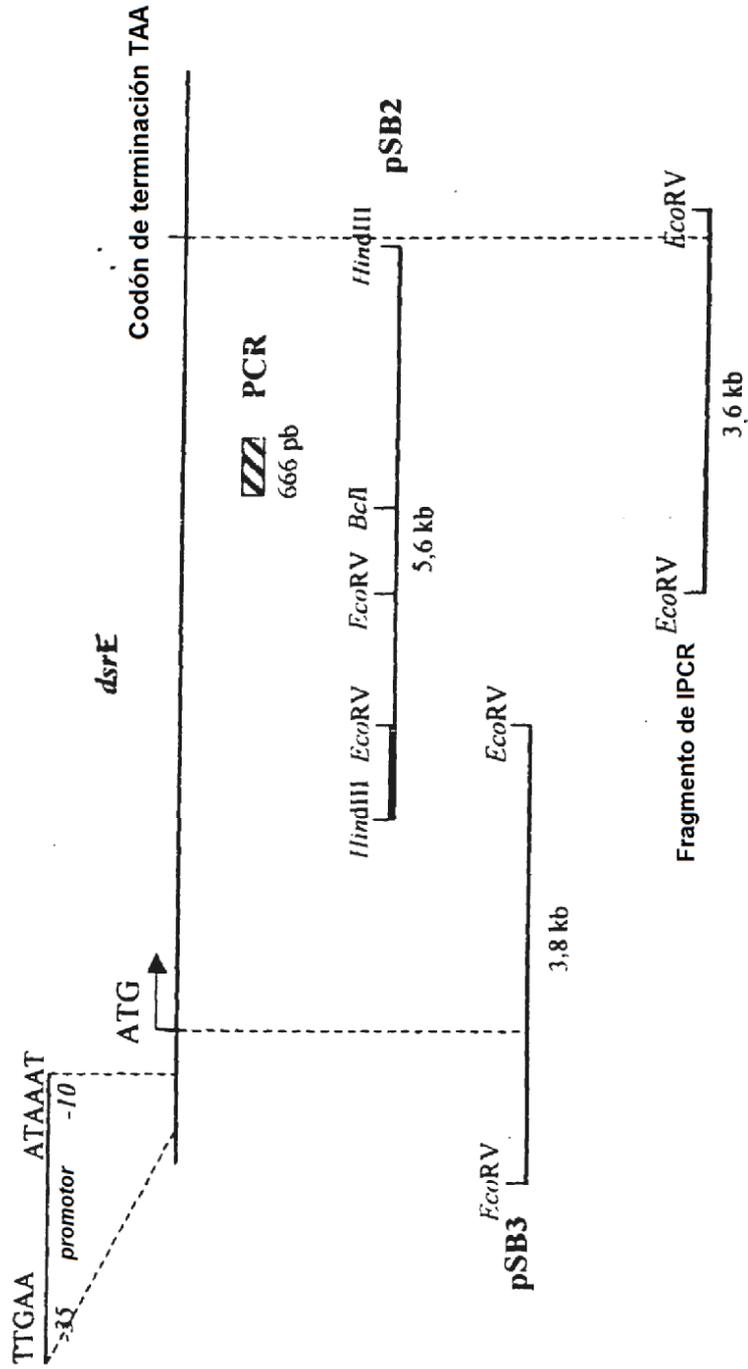


Fig. 2

DSR-E	MRDMRVICDRRKKLLYKSGKVLVTAG-IFALMMFGVTTASVSA
DSR-B	MEMIKERNVRRKLLYKSGKSWVIGGLIILSTIMLSMTATS----
DSR-S	-MPFTEKVMRKKLLYKVGKSWVGG-----VCAFALTAS-----
ASR	-MKQDETVTTRKK-YKSGKVVVAAATAFAVLGVSTVTTVHA-

Fig. 3

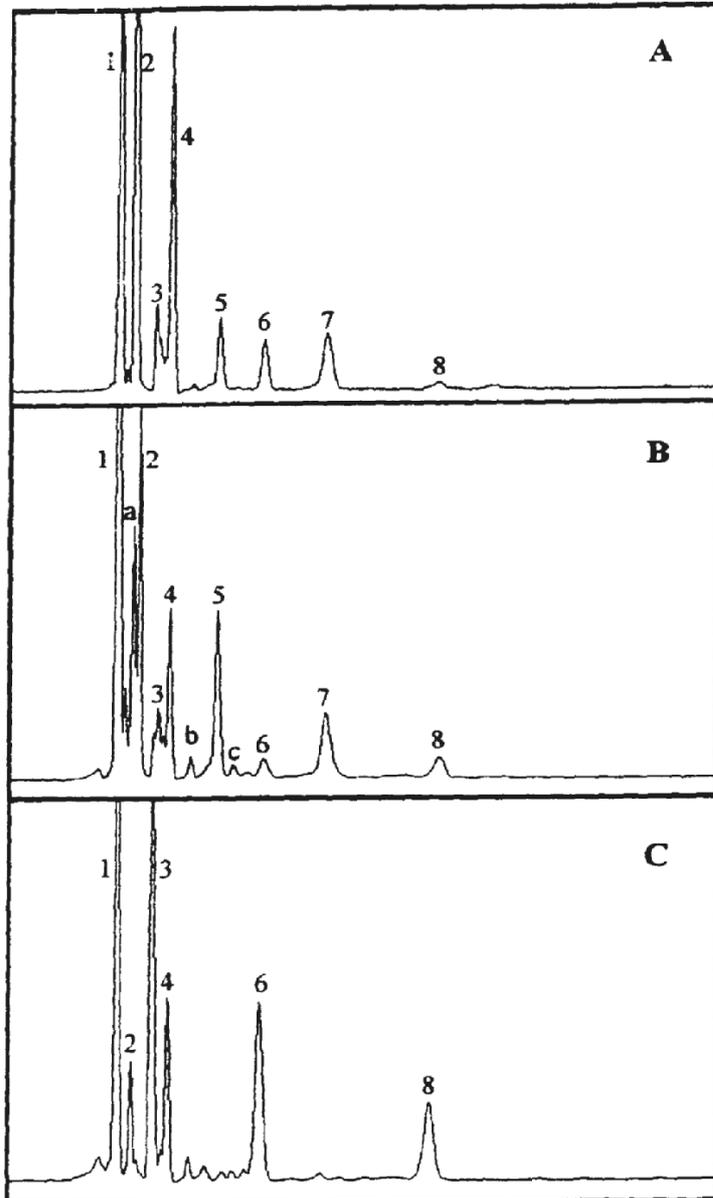
<sup>73</sup>AAKVVAVATTP-AT  
<sup>86</sup>FVADKTVSA  
<sup>95</sup>PAADKAVDTTSSIT  
<sup>109</sup>PATDKAVDTTP-TT  
<sup>122</sup>PAADKAVDTTP-TT  
<sup>135</sup>PAADKAVDTTP-TT  
<sup>148</sup>PAANKAVDTTP-AT  
<sup>161</sup>AATDKAV-ATP-AT  
<sup>173</sup>PAADKLANTT--AT  
<sup>185</sup>---DKAVATTP-AT  
<sup>196</sup>FVANKAA  
PAADKAVDTTP-~~A~~T ← Secuencia consenso propuesta para repetición S

Fig. 4

Fig. 5

A		B		C		D		E		F	
GTFB	341	SAWNSDSEK-----PFDDHLQN	402	GGYELLANDVDNSNPVVQAEQLN	484	HLSILEAWSND	555	YSFIRAHNSEVQDLI	928	DWVPDQMY	
GTFI	341	PQWNGESEK-----PYDDHLQN	404	GGYELLANDVDNSNPVVQAEQLN	486	HVSIVEAWSND	557	YSFARAHNSEVQDLI	932	DWVPDQMY	
GTFE	327	NQWSIASENETVYPNQDHMQG	388	AGYELLANDVDNSNPVVQAEQLN	470	HLSILEAWSGND	540	YVFIRAHNSEVQTRI	915	DLVFNQLY	
dsrS	444	PQWNETSED-----MSNDHLQN	502	GGFELLANDVDNSNPVVQAEQLN	584	HLSILEDSHND	655	YSFVRAHNSEVQTVI	1024	DWVPDQIY	
dsrA	181	PNWNIDSEA-----KGDDHLQG	237	GGFELLANDVDNSNPVVQAEQLN	319	IYQFWKTGEMKI	390	YSFIRAHNSEVQTVI	765	DWVPDQIY	
dsrB	426	PQWNMSSED-----PKNDHLQN	484	GGFELLANDVDNSNPVVQAEQLN	566	HLSILEDSHND	637	YSFVRAHNSEVQTVI	1005	DWVPDQIY	
asr	525	ANWNKQTEDEAF-DGLOWLQG	585	KGSEFLLANDIDNSNPVVQAEQLN	667	HLSILEDWNGKD	759	YSFVRAHDYDAQDPI	1168	DWVPDQIY	
CD1	423	ANWNIDSES-----KGNDHLQG	478	GGYEMLLANDVDNSNPVVQAEQLN	560	HISILEDWNDND	631	YAFIRAHNSEVQTVI	1014	DWVPDQIY	
CD2	2120	FIWNKDSEYHG--GGDAWFQG	2161	NAFDPELLANDVDNSNPVVQAEQLN	2243	HISLVEAG----	2315	YSIIHAHDKGVQEKV	2689	DVVNDQVY	
		*. :*		. :*****:*****:***:***		. : . :		* : :**** . *		* * :*:*	

**Fig. 6**



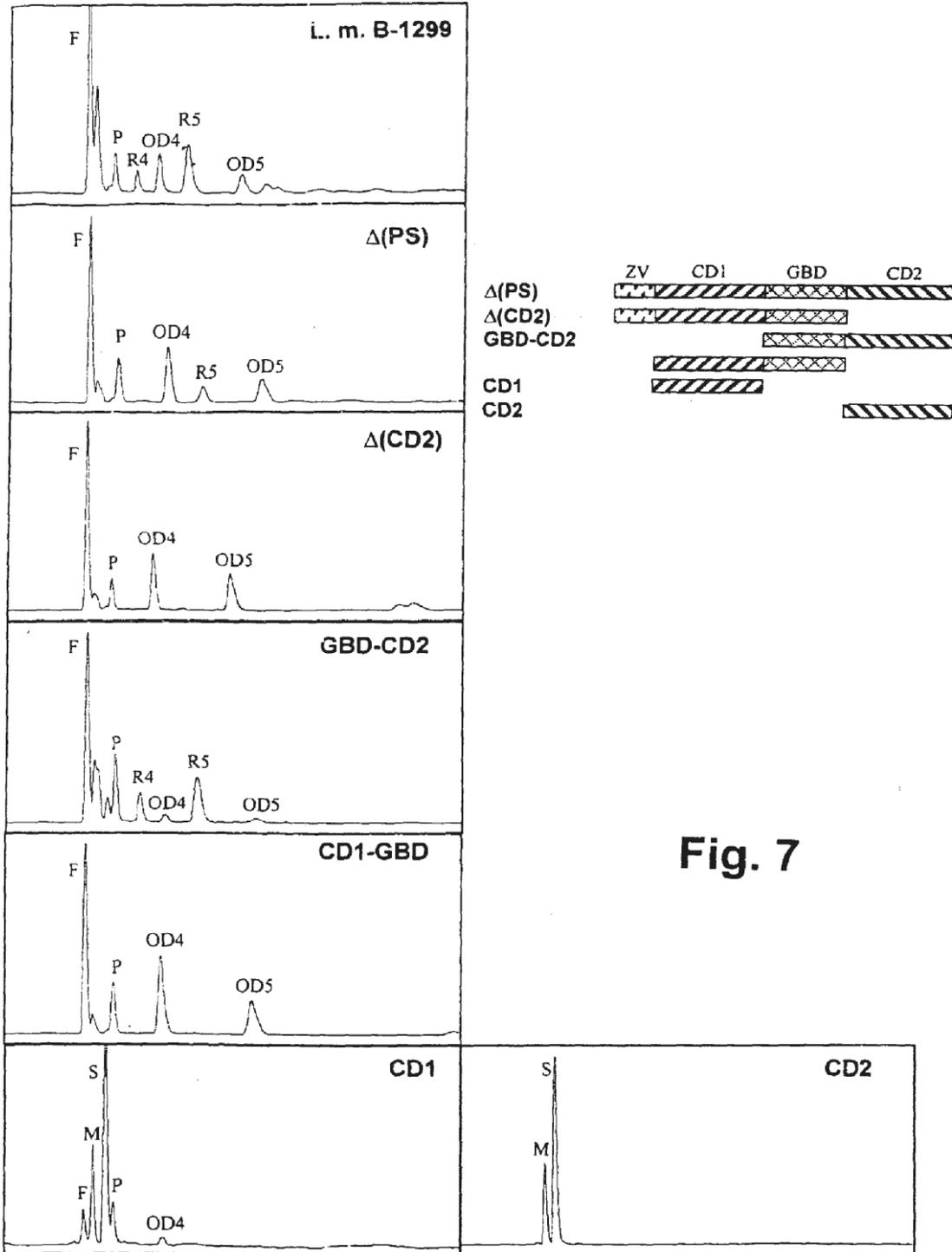


Fig. 7