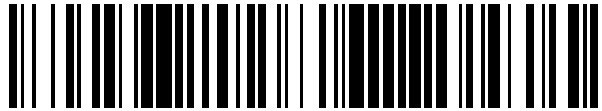


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 640 747**

51 Int. Cl.:

**C07K 1/00** (2006.01)

**C07K 14/415** (2006.01)

**C12N 15/64** (2006.01)

**C07H 21/04** (2006.01)

**C12N 15/29** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.10.2007 PCT/AU2007/001542**

87 Fecha y número de publicación internacional: **17.04.2008 WO08043147**

96 Fecha de presentación y número de la solicitud europea: **10.10.2007 E 07815347 (5)**

97 Fecha y número de publicación de la concesión europea: **30.08.2017 EP 2079755**

54 Título: **Procedimiento para la generación de proteínas y usos del mismo**

30 Prioridad:

**10.10.2006 AU 2006905622**

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

**06.11.2017**

73 Titular/es:

**THE AUSTRALIAN NATIONAL UNIVERSITY  
(100.0%)**

**Acton ACT 0200 AU**

72 Inventor/es:

**GREADY, JILL, E. y  
KANNAPPAN, BABU**

74 Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

**ES 2 640 747 T3**

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

## DESCRIPCIÓN

**Procedimiento para la generación de proteínas y usos del mismo****5 Referencia cruzada a la solicitud relacionada**

Esta solicitud reclama el beneficio de la Solicitud de Patente Provisional de Australia Núm. 2006905622 presentada el 10 de octubre de 2006.

**10 Campo técnico**

La presente invención se refiere a un procedimiento para la generación de proteínas con propiedades mejoradas, en particular proteínas Rubisco, y usos del mismo.

**15 Antecedentes**

La ribulosa-1,5-bisfosfato (RuBP) carboxilasa/oxigenasa es más comúnmente conocida por la abreviatura Rubisco. Rubisco es una enzima que participa en la fijación de carbono en el Ciclo de Calvin, donde el dióxido de carbono atmosférico se fija y se pone a disposición de los sistemas biológicos en forma de moléculas ricas en energía.

20 En las plantas, algas, cianobacterias y proteobacterias fotótrofas y quimioautótrofas, Rubisco comprende cadenas de subunidades grandes (LSU) y cadenas de subunidades pequeñas (SSU). Los sitios de unión al sustrato están situados en las cadenas grandes. Las cadenas grandes forman dímeros en los que los aminoácidos de cada cadena grande contribuyen a los sitios de unión. Un total de cuatro dímeros de cadena grande y ocho cadenas pequeñas se ensamblan en un complejo más grande de aproximadamente 540.000 Da.

Rubisco cataliza el primer paso en la asimilación de CO<sub>2</sub> fotosintético (reducción de carbono), así como la fijación competitiva de O<sub>2</sub> que produce el producto de desecho que se recicla en oxidación fotorrespiratoria de carbono. La importancia de Rubisco es subrayada por el hecho de que es el principal catalizador en la reacción química clave por la cual el carbono inorgánico entra en los sistemas biológicos. Además, Rubisco es una proteína muy abundante. Parry *et al.* (2003) sugieren que Rubisco representa 30-50% de la proteína soluble total en los cloroplastos.

35 Sin embargo, la abundancia relativa de Rubisco puede atribuirse al hecho de que es una enzima de acción muy lenta, que sólo fija unas pocas moléculas de CO<sub>2</sub> por segundo, en contraste con las miles de reacciones químicas por segundo que caracterizan a muchas enzimas. La enzima es ineficiente como catalizador para la carboxilación de RuBP y está sujeta a inhibición competitiva por O<sub>2</sub>, a inactivación por pérdida de carbamilación y a una inhibición de punto muerto por unión a RuBP antes de la activación de la enzima por carbamilación por CO<sub>2</sub>. Este comportamiento no óptimo convierte a Rubisco en limitante de la velocidad de fotosíntesis. En consecuencia, en la mayoría de las condiciones y cuando la luz no limita la fotosíntesis, Rubisco es la principal enzima limitante del ciclo de Calvin.

Puesto que Rubisco limita frecuentemente la tasa de fotosíntesis en las plantas, las formas mejoradas de Rubisco tendrían un impacto considerable en el aumento de la productividad agrícola. Se han realizado varios intentos para aumentar la eficacia de las reacciones mediadas por Rubisco. Los enfoques anteriores incluyen introducir constructos que expresan Rubisco de un organismo en otro organismo, aumentar el nivel de expresión de subunidades de Rubisco, expresar las subunidades pequeñas de Rubisco del ADN del cloroplasto y alterar los genes de Rubisco mediante mutagénesis para intentar aumentar la especificidad por el dióxido de carbono sobre el oxígeno) o aumentar de otro modo la tasa de fijación de carbono.

50 Se han realizado intentos para introducir Rubisco foráneas, por ejemplo las de algas rojas tales como *Galdieria partita* que tienen una alta especificidad por CO<sub>2</sub>/O<sub>2</sub>, en plantas verdes con flores. Se esperaría que esto mejorara la eficiencia fotosintética de las plantas de cultivo, pero estos intentos fracasaron debido a problemas con la producción, el ensamblaje y la regulación de la Rubisco foránea en la planta anfitriona (Spreitzer y Salvucci, 2002, Parry *et al.*, 2003). Por otro lado, la subunidad grande de Rubisco de tabaco ha sido sustituida con éxito por la subunidad homóloga grande de la más simple bacteria fotosintética púrpura *Rhodospirillum rubrum* que no requiere una pequeña subunidad para plegarse y ensamblarse para proporcionar una enzima activa (Andrews y Whitney, 2003). Si bien se demostró que se podía lograr el reemplazo de Rubisco, la planta transgénica exhibía las muy inferiores propiedades de especificidad y catalíticas de Rubisco de *R. rubrum*.

60 Se han realizado numerosos intentos de definir las funciones de los residuos del sitio activo en etapas específicas de la reacción o de modificar y mejorar las propiedades catalíticas de Rubisco utilizando la mutagénesis dirigida al sitio acoplada al conocimiento de estructuras cristalográficas de rayos X de complejos de Rubisco de varias especies. Sin

embargo, estos estudios no han conseguido proporcionar una definición detallada y autoconsistente de las diversas funciones de los residuos del sitio activo durante todo el curso del tiempo de reacción. Estas técnicas han sido uniformemente infructuosas en la modificación de una "mejor" Rubisco. Si bien se puede deducir un mecanismo por el cual Rubisco opera mediante estos estudios, puede no ser único debido a diferentes interpretaciones posibles de los datos experimentales incompletos, y, por lo tanto, puede no ser el mecanismo que existe en la realidad. Por ejemplo, los mecanismos propuestos en el mecanismo de consenso de Cleland para Rubisco suponen que una molécula de agua era desplazada del magnesio en el sitio activo antes de la formación del complejo reactivo para la carboxilación y que consecuentemente todas las etapas subsiguientes de la reacción también partirían del supuesto de que este desplazamiento tiene lugar. Sin embargo, no hay evidencia experimental de que el agua haya sido desplazada realmente.

Esto contrasta con los programas de reestructuración para muchas otras enzimas en las que mutaciones únicas, o en algunos casos mutaciones múltiples, que se dedujeron directamente de los datos estructurales y mecánicos obtenidos experimentalmente han demostrado ser exitosas en modificar la especificidad del sustrato o la eficiencia catalítica en direcciones deseables predecibles.

Una dificultad importante en la implementación de un enfoque de reestructuración racional para Rubisco es que ninguno de los estudios experimentales referidos ha proporcionado evidencia directa de la estructura para todos los intermedios involucrados, ni los papeles exactos de todos los residuos participantes en el sitio activo debido a la incompleta de los datos experimentales. Los enfoques experimentales son inherentemente incapaces de definir funciones precisas para protones y moléculas de agua involucradas en procesos catalíticos, ya que son "invisibles" a las sondas experimentales. Esta dificultad se ve agravada por la complejidad tanto del sitio activo de Rubisco como por el hecho de que está implicada una secuencia de reacciones. Se propone que hay una serie de "elementos" del sitio activo que comprenden diferentes combinaciones de residuos del sitio activo que participan en las diferentes etapas de las reacciones, siendo los grupos de residuos a menudo "reutilizados".

Parece que las Rubisco actuales representan sólo "soluciones evolutivas parciales" para optimizar la eficiencia enzimática, es decir, que los procesos evolutivos han sido incapaces de muestrear eficazmente el espacio de secuencia de LSU, y que estas soluciones actuales representan soluciones que distan de lo óptimo. Por lo tanto, existe la oportunidad de crear soluciones más ideales mediante una ruta diferente, o combinación de rutas, de lo que la evolución biológica ha sido capaz de proporcionar hasta ahora.

La creación o identificación e introducción de formas más eficientes de Rubisco en organismos fotosintéticos por transformación, selección selectiva u otras manipulaciones puede permitir un crecimiento más eficiente de estos organismos, incluidas las plantas verdes y en particular las plantas con flores, ya que éstas harían un uso más eficiente del agua y el nitrógeno, y puede crecer más eficientemente a temperaturas más altas. Esto a su vez ofrece perspectivas para cultivos con mayor rendimiento, revegetación de tierras degradadas o propensas a la sequía, mejores opciones para el secuestro de carbono y mejoras en la producción de biocombustible o energía de biomasa, etc.

Read et al. (1994, Archives of Biochemistry and Biophysics, 312, 210-218) describen ribulosa bifosfato carboxilasa/oxigenasa de alto factor de especificidad de sustrato de algas marinas eucarióticas y propiedades de rubisco cianobacterianas recombinantes que contienen modificaciones de residuos de algas. Yu et al (2005, Protein Engineering, Design and Selection, 18, 589-596) describen un esquema de análisis basado en la evolución para identificar los factores determinantes de la especificidad de CO<sub>2</sub>/O<sub>2</sub> para ribulosa 1,5-bisfosfato carboxilasa/oxigenasa. Pearce et al. (2003, The Journal of Biological Chemistry, 278, 32526-32527) describen la relación entre las reacciones secundarias y la inhibición lenta de la ribulosa-bisfosfato carboxilasa revelada por un mutante del lazo 6 de la enzima del tabaco. Spreitzer et al. (2005, PNAS, 102, 17225-17230) describen la modificación filogenética en una interfaz entre subunidades grandes y pequeñas que confiere propiedades cinéticas tierra-planta a rubisco de algas. Wei - Wen et al. (1997, Chinese Journal of Chemistry, 15, 353-360) describen el modelado molecular de rubisco de arroz. Gutteridge et al. (1988, Plant Physiol. Biochem., 26, 675-682) describen detalles de las reacciones catalizadas por formas mutantes de rubisco. Las páginas 263 a 283 del libro "Artificial Photosynthesis: from basic biology to industrial application" (2006, ISBN: 978-3-527-60674-0) definen las ineficiencias en el mecanismo químico de la enzima fotosintética rubisco mediante estimulación computacional.

En resumen, existe la necesidad de generar proteínas, tales como una Rubisco, que tenga propiedades funcionales mejoradas en las que, por ejemplo, tales proteínas tengan una eficacia mejorada y estén adaptadas específicamente para condiciones ambientales particulares.

## 60 Compendio

Los autores de la presente invención han teorizado que las Rubisco existentes han muestreado sólo una fracción del

espacio mutacional disponible teóricamente para mejorar su eficiencia y que pueden haber sido muestreados diferentes espacios por diferentes grupos de Rubisco. Por lo tanto, proponen el acoplamiento de estas "soluciones" evolutivas parciales mediante el injerto de características de más de un grupo filogenético, o de más de una especie ambientalmente adaptada, sobre una Rubisco anfitriona para obtener los beneficios de un muestreo mutacional evolutivo más amplio de lo que ha sido posible naturalmente. Proponen que este procedimiento identificará las Rubisco con mayor eficiencia u otras propiedades funcionales.

De acuerdo con un primer aspecto de la invención, se proporciona un método para generar una proteína Rubisco con una propiedad funcional mejorada seleccionada del grupo consistente en una mejora de la eficiencia cinética de la Rubisco, un incremento de la especificidad para dióxido de carbono sobre oxígeno de la Rubisco, una alteración de la especificidad de la Rubisco para uno o más sustratos, una alteración de la especificidad para uno o más productos de la Rubisco y una alteración del intervalo de temperatura eficaz para la catálisis de Rubisco, comprendiendo el método:

(a) identificar al menos un Residuo de aminoácido Diana en una primera proteína Rubisco, en donde dicho Residuo de aminoácido Diana está asociado con dicha propiedad funcional, en donde los Residuos de aminoácidos Diana de la proteína Rubisco son los residuos de la primera envoltura que se coordinan directamente con el centro de reacción o residuos de la segunda envoltura que se coordinan directamente con uno o más de los residuos de la primera envoltura, en donde dichos residuos de la primera envoltura de Rubisco se seleccionan del grupo que consiste en GLU60, ASN123, LYS175, LYS177, KCX201, ASP203, GLU204 y HIS294, donde la numeración de residuos es de Rubisco de espinaca, y KCX201 denota LYS201 carbamilada;

(b) comparar al menos una segunda proteína Rubisco homóloga de la misma o de una rama filogenética diferente como la primera proteína Rubisco con la primera proteína Rubisco e identificar al menos un Residuo de aminoácido Variante entre la primera proteína Rubisco y la segunda proteína Rubisco, en donde los residuos de la segunda proteína Rubisco utilizada para la comparación con la primera proteína Rubisco comprenden todos los residuos directamente coordinados con el sitio activo de una proteína Rubisco, interactuando todos los residuos con el centro reactivo del sustrato o especies de reacción intermedias de una proteína Rubisco, y otros residuos dentro de una distancia próxima entre 3 y 28 Å de cualquier átomo de sustrato o especie de reacción intermedia del sitio activo de una proteína Rubisco o un subconjunto de tales residuos;

(c) seleccionar al menos un Residuo de aminoácido Candidato de los Residuo de aminoácido Variantes identificados en (b) basándose en la evaluación de la proximidad espacial de los Residuos Variantes a los Residuos Diana y estimar y clasificar su capacidad para influir en la electrostática y la orientación de los Residuos Diana y modular la propiedad funcional de la proteína formada;

(d) producir al menos una proteína Rubisco Mutante Candidata *in vitro* en donde dicho al menos un Residuo de aminoácido Candidato de la segunda proteína Rubisco sustituye un residuo correspondiente en la primera proteína Rubisco; y

(e) escrutar dicha al menos una proteína Rubisco Mutante Candidata producida en (d) para identificar una proteína Rubisco que tiene dicha propiedad funcional mejorada.

En un segundo aspecto, la etapa (a) y la etapa (b) se realizan simultáneamente.

En un tercer aspecto, la etapa (b) del primer aspecto se realiza antes de la etapa (a).

En un cuarto aspecto, la etapa (d) de uno cualquiera de los aspectos primero a tercero, comprende producir al menos una proteína Rubisco Mutante Candidata utilizando al menos un Residuo de aminoácido Candidato de la segunda proteína Rubisco para sustituir un residuo correspondiente en una proteína Rubisco homóloga o en proteínas Rubisco homólogas distintas o además de la primera proteína Rubisco.

En un quinto aspecto, la etapa (d) de uno cualquiera de los aspectos primero a cuarto comprende producir al menos una proteína Rubisco Mutante Candidata en la que al menos dos Residuos de aminoácido Candidato de la segunda proteína Rubisco sustituyen residuos correspondientes en la primera proteína Rubisco y/o en una proteína Rubisco homóloga distinta de la primera proteína Rubisco.

En un sexto aspecto, el al menos un Residuo de aminoácido Diana en el método de cualquiera de los aspectos primero a quinto, está contenido en un conjunto de proteínas que contienen la primera proteína Rubisco.

En un séptimo aspecto, el al menos un Residuo de aminoácido Diana en el método de cualquiera de los aspectos primero a sexto es al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 12, al menos 15, al menos 20, al menos 30 o al menos 50 Residuos de aminoácido Diana.

En un octavo aspecto, los Residuos de aminoácido Diana de la proteína Rubisco en el método del primer aspecto, están en el dominio N-terminal de la subunidad Grande de Rubisco.

5 En el noveno aspecto, la etapa (c) de cualquiera de los aspectos primero a octavo, comprende seleccionar al menos un Residuo de aminoácido Candidato Divergente y/o al menos un Residuo de aminoácido Candidato Alternativo, en lugar de al menos un Residuo de aminoácido Candidato, a partir de los Residuos de aminoácidos Variantes identificados en (b), en donde dicho Residuo de aminoácido Candidato Divergente o Residuo Candidato Divergente (DCR) se refieren a un residuo de aminoácido que se selecciona de entre una pluralidad de Residuos Variantes y que se sospecha que es capaz para influir estéricamente y/o electrostática sobre uno o más Residuos Diana y, por tanto, influir en la función de la proteína Rubisco mediada por los uno o más Residuos Diana; y en donde dicho Residuo de aminoácido Candidato alterno o Residuo Candidato Alternativo (ACR) se refiere a un residuo de aminoácido alternativo en la posición de un Residuo Candidato que se expresa en una segunda proteína Rubisco, pero que no es el aminoácido que se expresa en la secuencia consenso de la segunda proteína Rubisco y que se sospecha que es capaz de influir estéricamente y/o electrostática sobre uno o más Residuos Diana y, por tanto, influir en la función de la proteína Rubisco mediada por los uno o más Residuos Diana.

En un décimo aspecto, el método de acuerdo con cualquiera de los aspectos primero a noveno, comprende la etapa adicional de realizar una evolución dirigida sobre dicha proteína Rubisco que tiene dicha propiedad funcional mejorada y escrutar sus productos.

## 20 Abreviaturas

	<i>aadA</i>	plásmido que confiere resistencia a la estreptomycin y espectinomycin
	ACR	Residuo Candidato Alternativo
25	2CABP	2-carboxiarabinitol 1,5-bisfosfato
	2C3KABP	2-carboxi-3-cetoarabinitol 1,5-bisfosfato
	3-PGA	3-fosfo-D-glicerato
	CM	Mutante Candidato
30	Modelo CPK	Representación molecular de llenado espacial de Corey-Pauling-Koltun de átomos esféricos con radios proporcionales al radio de van der Waals del átomo
	CR	Residuo Candidato
	CvR	Residuo co-variante
	DCR	Residuo Candidato Divergente
	DFT	teoría funcional de densidad
35	ESP	potencial electrostático
	FM	modelo de fragmento
	H <sub>6</sub>	6x etiqueta de afinidad de histidina
	K <sub>c</sub>	Constante de Michaelis (K <sub>m</sub> ) para CO <sub>2</sub> a 0% de oxígeno
	K <sub>c</sub> <sup>aire</sup>	Constante de Michaelis (K <sub>m</sub> ) para CO <sub>2</sub> al 21% de oxígeno (ambiente)
40	K <sub>cat</sub> <sup>c</sup>	número de recambio de carboxilación
	K <sub>o</sub> o K <sub>io</sub>	constante de Michaelis (K <sub>m</sub> ) para O <sub>2</sub>
	LSU	Subunidad grande
	MD	dinámica molecular
	Código PDB	Número de identidad del banco de datos de proteínas
45	ONIOM	El Método de Orbital molecular integrado de n-capas + mecánica molecular de los autores de la presente invención
	QM	mecánica cuántica
	QM/MM	mecánica cuántica híbrida/mecánica molecular
	QM/QM	mecánica cuántica híbrida/mecánica cuántica
50	<i>rbcL</i>	polinucleótido que codifica una subunidad grande de Rubisco
	<i>rbcLS</i>	polinucleótido que codifica una subunidad grande de Rubisco de <i>Synechococcus</i> sp.
	<i>rbcLS-rbcSS</i>	(o <i>rbcL-S</i> ) que codifica una subunidad grande de Rubisco y una subunidad pequeña de Rubisco de <i>Synechococcus</i> sp.
	<i>rbcM</i>	polinucleótido que codifica la Forma II de Rubisco de <i>Rhodospirillum rubrum</i>
55	<i>rbcS</i>	polinucleótido que codifica una subunidad pequeña de Rubisco
	<i>rbcSS</i>	polinucleótido que codifica una subunidad pequeña de Rubisco de <i>Synechococcus</i> sp.
	RuBP	Ribulosa-1,5-bisfosfato
	Rubisco	Ribulosa-1,5-bisfosfato carboxilasa/oxigenasa
	S <sub>co</sub>	Especificidad de Rubisco para CO <sub>2</sub> comparada con O <sub>2</sub> definida como (k <sub>cat</sub> <sup>c</sup> /K <sub>c</sub> )/(k <sub>cat</sub> <sup>o</sup> /K <sub>o</sub> )
60	SSU	Subunidad pequeña
	SsVR	Residuo Variante específico de la especie
	Barril TIM	una estructura de dominio proteico definida en primer lugar en la triosa fosfato isomerasa

TR	Residuo Diana
Ts	estado de transición
Ub	ubicuitina
$V_c^{m\acute{a}ximo}$	actividad carboxilasa máxima extrapolada
5 VR	Residuo Variante

### Definiciones

10 Según se utiliza en esta solicitud, la forma singular "un", "uno", "una", "el" y "la" incluyen referencias plurales a menos que el contexto indique claramente lo contrario. Por ejemplo, el término "una célula vegetal" también incluye una pluralidad de células vegetales.

15 Según se utiliza en la presente memoria, el término "que comprende" significa "que incluye". Las variaciones de la expresión "que comprende", tales como "comprenden" y "comprende", tienen significados correspondientemente variados. Así, por ejemplo, un polinucleótido "que comprende" una secuencia que codifica una proteína puede consistir exclusivamente en esa secuencia o puede incluir una o más secuencias adicionales.

20 Por "célula anfitriona" se entiende una célula que contiene un constructo de ácido nucleico introducido y que soporta la replicación y/o expresión del constructo. Las células anfitrionas pueden ser células procariontas tales como *E. coli*, o células eucariotas tales como células de algas, hongos, levaduras, insectos, anfibios, nematodos, plantas o mamíferos. Las células anfitrionas pueden ser células vegetales, tales como células vegetales monocotiledóneas o células vegetales dicotiledóneas. Un ejemplo de una célula anfitriona es un célula anfitriona de *E. coli*.

25 Se entiende que el término "planta verde" según se utiliza en la presente memoria abarca organismos que incluyen, pero no necesariamente se limitan a, organismos unicelulares o multicelulares de las Divisiones *Pteridophyta* (helechos), *Bryophyta* (musgos), *Charophyta* y *Chlorophyta* (algas verdes acuáticas), *Magnoliophyta* (plantas con flores o angiospermas), y *Pinofita* (coníferas).

30 Según se utilizan en la presente memoria, las proteínas "homólogas" son proteínas que comparten un origen evolutivo. Las proteínas homólogas pueden compartir la misma función esencial (proteínas ortólogas) o pueden exhibir funciones significativamente diferentes, evolutivamente divergentes (proteínas parálogas).

35 Según se utiliza en la presente memoria, "ácido nucleico" significa un polinucleótido e incluye un polímero de una o dos cadenas de bases de desoxirribonucleótido o ribonucleótido. Los ácidos nucleicos también pueden incluir fragmentos y nucleótidos modificados.

40 Según se utiliza en la presente memoria, "conectado operablemente" incluye la referencia a una conexión funcional de al menos dos secuencias. Conectado operablemente incluye una conexión entre un promotor y una segunda secuencia, en donde la secuencia promotora inicia y media la transcripción de la secuencia de ADN correspondiente a la segunda secuencia. El *rbcl-S* es un ejemplo de una secuencia conectada operablemente.

45 Según se utiliza en la presente memoria, la "fotosíntesis" se refiere al proceso en plantas verdes y ciertos otros organismos mediante el cual se sintetizan hidratos de carbono a partir de dióxido de carbono y agua utilizando luz como fuente de energía. La mayoría de las formas de fotosíntesis liberan oxígeno como subproducto.

50 Según se utiliza en la presente memoria, "rama filogenética" se refiere a un grupo de organismos conectados por linaje. En el contexto de la presente invención, la filogenética constituye un medio para clasificar grupos de organismos de acuerdo con el grado de relación evolutiva. Una rama filogenética puede contener organismos de diferentes divisiones, clase, orden, familia, género o especie taxonómicas.

55 "Injerto filogenético" se refiere al procedimiento de introducción de al menos un residuo de aminoácido de una proteína donadora de un organismo de una rama filogenética en una proteína receptora de un organismo de una rama filogenética diferente con el fin de mejorar las propiedades funcionales de la proteína receptora. El injerto filogenético puede llevarse a cabo sustituyendo por lo menos un residuo de aminoácido en una posición dada en la secuencia de una proteína receptora, estando el al menos un residuo de aminoácido en la misma posición en la segunda proteína donadora seleccionada sobre la base del análisis filogenético.

60 Según se utiliza en la presente memoria, "planta" incluye plantas y partes de plantas que incluyen pero no se limitan a células vegetales, tejidos vegetales tales como hojas, tallos, raíces, flores y semillas.

Según se utiliza en la presente memoria, "promotor" incluye la referencia a una región de ADN que está implicada en el reconocimiento y la unión de una ARN polimerasa y otras proteínas para iniciar la transcripción.

- Según se utiliza en la presente memoria, "proteína" se refiere a cualquier polímero de aminoácidos conectados mediante enlaces peptídicos o enlaces peptídicos modificados, ya sea producidos naturalmente o sintéticamente. La proteína de la invención puede comprender componentes no peptídicos, tales como grupos carbohidrato. Los carbohidratos y otros sustituyentes no peptídicos pueden ser añadidos a una proteína por la célula en la que se produce la proteína y varían con el tipo de célula. Las proteínas se definen en la presente memoria, en términos de sus estructuras de la cadena principal de aminoácidos; no se especifican generalmente sustituyentes tales como grupos carbohidrato, pero pueden estar presentes no obstante.
- Según se define en la presente memoria, la "función" de una proteína se refiere al papel normal de la proteína en la célula, o un papel para el cual se puede modificar genéticamente una proteína para que lo lleve a cabo. En una realización la proteína es una enzima. Cuando la proteína es una enzima, la función puede ser la catálisis de al menos una reacción química. En otras realizaciones, la función puede ser estructural (p.ej., servir como una proteína del citoesqueleto). La función puede implicar el transporte activo o pasivo de una sustancia dentro de la célula o entre el interior y exterior de la célula, o entre diferentes compartimientos dentro de la célula, o entre diferentes regiones del organismo, por ejemplo cuando la proteína está implicada en un canal o un poro de membrana, o la proteína está implicada en el tráfico de materiales a compartimientos celulares específicos o la proteína actúa como una chaperona o un transportador. La función puede estar implicada con interacciones ligando/receptor, por ejemplo cuando la proteína es un factor de crecimiento, una citocina, un neurotransmisor o un ligando intracelular o extracelular, o la proteína es un receptor para el factor de crecimiento, la citocina, el neurotransmisor o el ligando intracelular o extracelular.
- Cuando la proteína es una enzima, la enzima puede estar implicada en catabolismo o metabolismo. La enzima puede estar implicada en la síntesis de al menos un producto. La enzima puede estar implicada en la rotura de al menos un sustrato. La enzima puede estar implicada en la modificación química de al menos un sustrato, por ejemplo la adición o deleción de uno o más grupos fosfato de una molécula. Las enzimas pueden estar implicadas en la degradación de al menos un sustrato.
- Así, una "propiedad funcional" de una proteína es una propiedad que contribuye a la función de la proteína. Por ejemplo, cuando una proteína es una enzima, la propiedad funcional puede ser la especificidad de la enzima para un sustrato concreto, la eficiencia cinética de la enzima, su intervalo eficaz de temperatura para la catálisis o su especificidad para catalizar su reacción normal y minimizar las reacciones secundarias a subproductos no deseados y/o potencialmente tóxicos.
- El término residuo en el contexto de un polipéptido se refiere a una unidad de aminoácidos en la cadena polipéptida lineal. Es lo que queda de cada aminoácido, es decir,  $-NH-CHR-C-$ , después de que el agua se elimina en la formación del polipéptido de los  $\alpha$ -aminoácidos, es decir,  $NH_2-CHR-COOH$ .
- Los términos "Residuo de aminoácido Diana" o "Residuo Diana" (TR) se refieren a un residuo de aminoácido que se identifica y/o pronostica que contribuye directamente a la función de la proteína. Cuando la proteína es una enzima, la contribución de un Residuo Diana a la propiedad funcional puede ser una influencia directa del Residuo Diana sobre la reacción o reacciones catalíticas llevadas a cabo por la enzima en uno o más de los sitios activos de la enzima, a través de la interacción directa o la participación en el sitio activo. Un Residuo Diana no estará distante de un sitio activo enzimático. Cuando la proteína es un receptor, el Residuo Diana estará directamente implicado en el sitio receptor. En los métodos descritos en la presente memoria, se puede identificar un Residuo Diana mediante métodos de simulación computacional y/o molecular, ya que estos métodos son capaces de considerar las posiciones de las moléculas de agua, protones, estados de ionización y enlaces de hidrógeno asociados con los Residuos Diana en el sitio activo enzimático, o sitio receptor, que no son definibles sin ambigüedad por el experimento. En los métodos de la presente invención, se pronostica que la mutación de los Residuos Diana conduciría en general a una interrupción o reducción de la función. Por consiguiente, en los métodos de la presente invención, los Residuos Diana no se alteran directamente mediante sustitución por otro aminoácido, sino que las propiedades de los Residuos Diana tales como su posición y carga se "sintonizan" mediante la manipulación de uno o más residuos que interactúan con los Residuos Diana.
- Los términos "Residuo de aminoácido Variante" o "Residuo Variante" (VR) se refieren a un residuo de aminoácido específico de una segunda proteína o un residuo de aminoácido específico identificado en una secuencia consenso de una pluralidad de segundas proteínas que se identifica como diferente del residuo de aminoácido correspondiente encontrado en una primera proteína que es homóloga a la segunda proteína. Los Residuos Variantes pueden identificarse, por ejemplo, utilizando un alineamiento de las secuencias de aminoácidos de la primera y segunda proteínas. Las secuencias de la primera y/o segunda proteínas pueden ser secuencias consenso. Las secuencias pueden derivarse de organismos de la misma o de una rama filogenética diferente. Para la subunidad grande de Rubisco, es raro que haya adiciones o deleciones de secuencias entre las secuencias de diferentes organismos,

aparte de en las regiones N o C-terminales que no son importantes para la función catalítica de Rubisco. Sin embargo, para los propósitos de la presente invención, en una realización, un residuo de aminoácido variante puede ser un residuo que está presente en una primera proteína pero ausente en una segunda proteína, o que está ausente en una primera proteína pero presente en una segunda proteína .

5 Los términos "Residuo de aminoácido Candidato" o "Residuo Candidato" (CR) se refieren a un residuo de aminoácido de una segunda proteína que está seleccionada de entre una pluralidad de Residuos Variantes y que se sospecha que puede influir en uno o más Residuos Diana estéricamente y/o electrostáticamente, e influyen de este modo en la función de la proteína mediada por uno o más Residuos Diana. Un Residuo Candidato es un residuo que  
10 puede ser transplantado selectivamente a una primera proteína de un anfitrión con el fin de intentar modular la actividad funcional de la primera proteína hacia la actividad funcional deseada de la segunda proteína.

En algunas realizaciones, el Residuo Candidato se puede seleccionar basándose en la uniformidad y/o diferencia de  
15 residuos de aminoácidos entre dos o más ramas filogenéticas. En el contexto de la selección de Residuos Candidato para Rubisco, un Residuo Candidato es uno que se sospecha que es capaz de influir estéricamente y/o electrostáticamente en uno o más Residuos Diana de Rubisco, por ejemplo modificando la distribución de carga sobre el Residuo Diana y/o modificando la posición espacial del Residuo Diana y/o modificando la capacidad del Residuo Diana para moverse.

20 Cuando la proteína es Rubisco, el Residuo Candidato puede estar presente en la secuencia consenso de algas rojas pero diferente en la posición de residuo correspondiente en las secuencias consenso de plantas con flores y cianobacterias y en una posición en la que el aminoácido es el mismo en plantas con flores y cianobacterias. En una descripción concreta, el Residuo de aminoácido Candidato puede elegirse entre el residuo presente en algas rojas.

25 Cuando la proteína es Rubisco, la identificación de un Residuo Candidato como uno que se sospecha que afecta a la etapa de adición de gas permite diferenciarlo de otros Residuos Variantes que muestran cambios no conservados entre ramas o sub-ramas, que pueden representar una deriva filogenética neutra o que pueden tener un papel fisiológico específico de rama, tal como en el plegamiento, el ensamblaje, incluyendo interacciones con la subunidad  
30 pequeña, o estabilidad.

Los términos "Residuo de aminoácido Candidato Divergente" o Residuo candidato Divergente (DCR) se refieren a un  
35 residuo de aminoácido que se selecciona de entre una pluralidad de Residuos Variantes y que se sospecha que es capaz de influir estéricamente y/o electrostáticamente, y de este modo influye en la función de la proteína mediada por uno o más Residuos Diana. El Residuo Candidato Divergente se selecciona basándose en la diferencia de residuos de aminoácidos en una posición dada en la secuencia consenso de la proteína entre al menos tres ramas filogenéticas. Por ejemplo, en el contexto de la selección de Residuos Candidato Divergentes para Rubisco, un Residuo Candidato Divergente es uno que se sospecha que es capaz de influir estéricamente y/o electrostáticamente en uno o más Residuos de aminoácidos Diana de Rubisco y que puede estar presente en la  
40 secuencia consenso de algas rojas, pero diferente en el residuo correspondiente de la secuencia consenso de plantas con flores y la secuencia consenso de cianobacterias, y también diferente en el residuo correspondiente de la secuencia consenso de plantas con flores y la secuencia consenso de cianobacterias. El Residuo Candidato Divergente puede seleccionarse de la secuencia o secuencia consenso de la proteína de cualquiera de las ramas filogenéticas que se compararon.

45 Los términos "Residuo de aminoácido Candidato Alternativo" o "Residuo Candidato Alternativo" (ACR) se refieren a un aminoácido alternativo en la posición de un Residuo Candidato que se expresa en una segunda proteína, pero que no es el aminoácido que se expresa en la secuencia consenso de la segunda proteína. De este modo, un Residuo Candidato Alternativo proporciona un residuo que se expresa en la posición dada en al menos una especie de la segunda proteína, pero que no se expresa en la mayoría de las secuencias de la misma rama filogenética.  
50 Cuando la proteína es Rubisco, se puede seleccionar un ACR, por ejemplo, a partir de la secuencia de Rubisco de *Griffithsia monolis* que exhibe una tasa catalítica significativamente más alta que una Rubisco de una especie típica de algas rojas.

Los términos "Residuo de aminoácido Co-Variante" o "Residuo Co-Variante" (CvR) se refieren a un residuo que se  
55 identifica en la secuencia de una segunda proteína de una especie concreta, al estar cerca de un Residuo Alternativo Candidato y mostrar una variación complementaria al Residuo Candidato Alternativo. Se puede sospechar que esta variación del Residuo Co-Variante, que no está presente en la secuencia consenso para la segunda proteína, refleja cambios complementarios en las propiedades estructurales y/o electrostáticas del Residuo Candidato Alternativo y del Residuo Co-Variante. La identificación de los residuos co-variantes proporciona un medio  
60 para identificar posiciones de residuos auxiliares que pueden ser mutadas en la primera proteína para acomodar mejor los cambios realizados a partir de la transferencia de Residuos Candidato Alternativos.

Los términos "Residuo de aminoácidos variante específico de especie" o "Residuo Variante específico de especie" (SsVR) se refieren a residuos que varían entre especies estrechamente relacionadas de una proteína. Conjuntamente con datos funcionales asociados para cada proteína variante, pueden utilizarse SsVR para mapear las relaciones secuencia-estructura-función. Estas correlaciones pueden usarse para pronosticar qué residuos variables podrían estar contribuyendo más a la mejora de la propiedad deseada para la proteína. Esta información de SsVR puede usarse conjuntamente con el método general para identificar CR. Las predicciones de una proteína mutante candidata que contiene grupos de residuos pueden incluir SsVR además de los CR, ACR, DCR o CvR que afectan más directamente a los Residuos Diana.

Los términos "proteína Mutante Candidata" o "Mutante Candidato" (CM) se refieren a una proteína mutante en la que se combinan al menos uno o grupos de dos o más CR y/o ACR y/o DCR, con CvR o SsVR adicionales opcionales en una sola proteína.

El término "Región" se refiere a una división de la estructura proteica que rodea a los Residuos Diana que puede hacerse basándose en la proximidad de un Residuo Diana o Residuos Diana concretos en partes concretas del centro reactivo o sitio de unión. Una región comprende un volumen espacialmente contiguo de estructura proteica que contiene subconjuntos de CR, DCR, ACR, CvR y SsVR que pueden influir preferentemente en las interacciones de un Residuo Diana o Residuos Diana concretos con partes concretas del centro reactivo o sitio de unión. Los límites de las Regiones no están definidos con precisión. Las Regiones pueden tener segmentos solapantes de la estructura de la proteína. El propósito de la definición de Regiones es facilitar la aplicación del método de injerto filogenético mediante la identificación de los subgrupos de CR, DCR, ACR, CvR y SsVR que pueden agruparse para formar Mutantes Candidato. Por ejemplo, cuando la proteína es Rubisco, se puede identificar una región marcada como Región 1, que comprende la mayor parte del dominio N-terminal de la subunidad grande y varios segmentos pequeños de estructura del dominio C-terminal de la subunidad grande adyacente para influir en los Residuos Diana Asn123, Glu60 y Tyr20, que interaccionan específicamente, directa o indirectamente, con el grupo carboxilato naciente de la especie reactiva durante la etapa de adición de gas.

El término "subregión" se refiere a una división de una región que puede hacerse basándose en la proximidad de un Residuo Diana o Residuos Diana concretos en un componente de la región. Una subregión comprende un volumen espacialmente contiguo de estructura proteica que contiene un subconjunto de los CR, DCR, ACR, CvR y SsVR de la Región que se pronostica que influyan preferentemente en las propiedades del Residuo Diana o de los Residuos Diana concretos vinculados a la región. Los límites de las Subregiones no están definidos con precisión. Las Subregiones pueden tener segmentos solapantes de estructura proteica. La definición de Subregiones puede facilitar la aplicación del método de injerto filogenético mediante la identificación somera de los subconjuntos de los CR, DCR, ACR, CvR y SsVR de la Región, los cuales pueden agruparse preferentemente para formar Mutantes Candidato. Por ejemplo, cuando la proteína es Rubisco, se pueden identificar las regiones componentes de la Región 1, marcadas como 1A, 1B y 1C, que se pronostica que influyan preferentemente en las propiedades de los Residuos Diana Tyr20, Asn123 y Glu60, respectivamente.

El término "espacio de secuencia" se refiere en sentido amplio al conjunto de todas las secuencias posibles de residuos para un polímero de longitud específica. Por ejemplo, el espacio de secuencia completo para una proteína o polipéptido de 100 residuos de aminoácidos de longitud es el conjunto de todas las secuencias que consisten en todas las posibles variaciones de los 100 aminoácidos. Si se consideran únicamente las mutaciones individuales, el espacio de secuencia para una proteína o polipéptido de 100 residuos de aminoácidos de longitud es 100 veces 19 (para el conjunto de residuos de 20 aminoácidos). Una reducción del espacio de secuencia de una proteína o polipéptido de una longitud dada, por lo tanto, reducirá el número de secuencias posibles de residuos de aminoácidos en un conjunto de mutantes de proteínas o polipéptidos. En el contexto de la presente invención, la identificación independiente de Residuos de aminoácido Diana y Residuos de aminoácidos Variantes produce dos conjuntos reducidos diferentes de espacio de secuencia a examinar. La subsiguiente intersección de estos conjuntos efectuada cuando los Residuos de aminoácido Diana se utilizan para identificar Residuos de aminoácido Candidato de Residuos Variantes reduce aún más el espacio de secuencia a considerar. Por consiguiente, el número de residuos de la proteína y sus combinaciones que necesitan considerarse para la mutación se reducen por lo tanto enormemente.

Según se utiliza en la presente memoria, el término "evolución dirigida" se refiere a procedimientos que tienen como objetivo modificar la función y/o la estructura de la proteína o proteínas diana. En general, la evolución dirigida es un procedimiento mediante el cual las proteínas son "adaptadas" para actuar en entornos químicos o biológicos naturales o artificiales diferentes o existentes y/o para obtener nuevas funciones y/o aumentar o disminuir una actividad dada, y/o para modular una característica dada.

## Breve descripción de las figuras

La presente invención se describirá ahora a modo de ejemplo con referencia a las figuras adjuntas.

El **Diagrama de flujo 1** proporciona un diagrama de flujo de la estrategia de reestructuración de Rubisco, destacando los roles de la química computacional y la bioinformática al proporcionar la información mecanicista, secuencial, estructural y filogenética que está integrada por el procedimiento de injerto filogenético para producir las predicciones *in silico* de proteínas Mutantes Candidatas. Esta integración proporciona un medio por el cual los Residuos Diana pueden usarse para seleccionar Residuos Candidato a partir de Residuos Variantes, reduciendo así enormemente el número de residuos de la proteína y sus combinaciones, los cuales necesitan ser considerados para la mutación hacia la mejora de la propiedad funcional. Los Residuos Candidato, tal como se muestra en la columna central, pueden comprender opcionalmente ACR y/o DCR. Estos procedimientos se muestran con más detalle en el Diagrama de flujo 2. Las etapas de predicción están seguidas de escrutinio experimental y evaluación de la mejora de la propiedad funcional. Los resultados pueden ser retroalimentados en el procedimiento de predicción para refinar las predicciones de las proteínas Mutantes Candidato, seguido de ciclos adicionales de escrutinio computacional y experimental y evaluación de la mejora de la propiedad funcional, para optimizar la actividad de la proteína. Los recuadros de líneas discontinuas representan extensiones opcionales al procedimiento de injerto filogenético básico que se muestran en forma ampliada en el Diagrama de flujo 3.

El **Diagrama de flujo 2** proporciona un diagrama de flujo que detalla la estrategia de predicción para el método de injerto filogenético, que muestra la integración de la química computacional con el análisis bioinformático para producir la lista clasificada de proteínas Mutantes Candidatas. Este procedimiento implica etapas de selección de Residuos Variantes frente a Residuos Diana para producir Residuos Candidato (que opcionalmente pueden comprender ACR y/o DCR), agrupamiento y combinación de grupos de Residuos Candidato (y/o Residuos Candidato Divergentes o Residuos Candidato Alternativos), y clasificación de grupos y grupos combinados para producir una lista clasificada de proteínas Mutantes Candidatas para su pre-escrutinio computacional opcional y escrutinio experimental.

El **Diagrama de flujo 3** proporciona un diagrama de flujo que detalla una extensión opcional del método para injerto filogenético, en donde se evalúa una lista de Residuos Candidato y se ensamblan en Mutantes Candidato, y a continuación se pueden reclutar y agrupar otros Residuos Candidato, Residuos Candidato Alternativos, Residuos Candidato Divergentes, Residuos Co-variantes y especies, basándose en su influencia en los Residuos Diana, y formar en nuevos Mutantes Candidato refinados.

La **Figura 1** proporciona una ilustración de las estructuras moleculares implicadas en el mecanismo de reacción propuesto para la conversión de RuBP en dos moléculas de 3-PGA en el sitio activo de Rubisco. Se muestran las cinco etapas de enolización, carboxilación, hidratación, escisión de enlaces C2-C3 y protonación de C2. Los números romanos corresponden a las especies intermedias y de producto que se muestran en las Figuras 3 y 4. Grupo R = -CH (OH) -CH<sub>2</sub>-O-PO<sub>3</sub><sup>2-</sup>.

La **Figura 2** ilustra la estructura molecular del modelo de fragmento de 77 átomos del sitio activo de Rubisco designado FM20 que se usó para estudios de química computacional QM *ab initio* de la ruta de reacción. Muestra las especies de fragmentos moleculares que se usaron para representar los residuos de aminoácidos LYS175, LYS177, ASP203, GLU204, KCX201 (LYS201 carbamilada), HIS294 y LYS334, la molécula de agua, la molécula de dióxido de carbono y el fragmento de 4 carbonos de la forma enodiolato del sustrato RuBP. Además, muestra los estados de carga de las especies componentes y sus interacciones, a saber, los seis átomos coordinados con el átomo de Mg, los enlaces de hidrógeno y la interacción de van der Waals entre CO<sub>2</sub> y C2.

Las **Figuras 3A a 3E** proporcionan una serie de estructuras moleculares calculadas utilizando métodos QM *ab initio* para el modelo de fragmento FM20 mostrado en la Figura 2. Éstas muestran geometrías de mínimos locales a lo largo de la ruta de reacción de enodiolato a los productos finales en la reacción carboxilasa de Rubisco. Los átomos de hidrógeno de la molécula de agua coordinada con Mg se destacan mediante círculos de color negro. Las distancias relevantes para cada estructura que demuestran las interacciones cambiantes entre las especies componentes a medida que la reacción avanza se muestran a través de etiquetas. Las etiquetas en algunas estructuras pueden ser oscurecidas. La etiqueta "d" se refiere a R<sub>KCX201-H...O2</sub>, "i" se refiere a R<sub>Ow...C3</sub>, "j" se refiere a R<sub>H2O[Mg]-Hw1...O-GLU204</sub> y "k" se refiere a R<sub>H2O[Mg]...Hw1</sub>. Los números Romanos corresponden a las especies intermedias, del estado de transición y producto que se muestran en la Figura 4.

La **Figura 4** proporciona un gráfico que ilustra la superficie de energía potencial para la carboxilación y las reacciones subsiguientes en la vía de la carboxilasa calculada utilizando cálculos de química computacional QM *ab initio* en el modelo de fragmentos con 77 átomos (FM20). Las diferentes etapas del progreso de la reacción de la carboxilasa se distribuyen a lo largo del eje X. Los números Romanos a lo largo de la ruta de reacción designan los estados de partida (I), intermedios (III, V, VII) y de producto (IX) de la reacción de la carboxilasa mostrada en la Figura 1, y sus estados de transición de conexión (II, IV, VI, VIII). Las estructuras

de todos estos estados se muestran en la Figura 3. Las energías (en kcal/mol) de todos los estados se muestran con respecto a la del estado de enediolato de partida I, y las energías del estado de transición con respecto al estado intermedio inicial o relevante se muestran por  $E_a$  y una flecha.

5 Las **Figuras 5A a 5D** proporcionan un alineamiento de secuencia de las secuencias de aminoácidos de la LSU de Rubisco de organismos fotosintéticos pertenecientes a trece phyla diferentes que cubren algas rojas, cianobacterias, glaucophyta y plantas (10 phyla), utilizando una simbología de aminoácidos de una sola letra. Estas secuencias también se proporcionan como SEQ ID NO: 1-13. Cuando está disponible más de una secuencia de Rubisco en un phylum, se usó una secuencia consenso del 50% para representar ese phylum.

10 Para las secuencias consenso, el número entre paréntesis representa el número de géneros cuyas secuencias se usaron para generar la secuencia consenso. Los números de acceso a la base de datos para todas las secuencias utilizadas en el alineamiento se proporcionan en la Tabla 1. La figura muestra que la secuencia de la LSU de Rubisco está altamente conservada, incluyendo la ausencia casi completa de lagunas a excepción de diferencias mínimas en los extremos N y C, y códigos para una cadena polipeptídica larga de 475 residuos (plantas y cianobacterias). El símbolo "~" denota huecos de secuencia. Las letras mayúsculas de una letra indican el alfabeto de residuos de aminoácidos, mientras que las letras minúsculas y otros símbolos se muestran en posiciones en secuencias consenso donde sólo se conserva el tipo de residuo: "h", hidrófobo (A, C, F, G, H), "i, k, l, m, r, t, v, w, y"; "s", pequeño (A, C, D, G, N, P, S, T, V); "u", diminuto (A, G, S); "a", aromático (F, H, W, Y); "c", cargado (D, E, H, K, R); "1", alifático (I, L, V); "p", polar (C, D, E, H, K, N, Q, R, S, T); "o", alcohol (S, T); "t", con tendencia a la formación de giros (A, C, D, E, G, H, K, N, Q, R, S, T); "-", cargado negativamente (D, E); y "+" cargado positivamente (H, K, R). Las secuencias consenso se obtuvieron utilizando el servidor en <http://coot.embl.de/Alignment/consensus.html>. El alineamiento se corrigió alrededor de la posición de separación cerca de CR T271 (posición de alineamiento 273) basándose en comparaciones estructurales entre las estructuras de rayos X para complejos de espinaca (pdb 8ruc), *Synechococcus* (pdb 1rb1) y *Galdieri partita* (pdb 1bwv) con  $Mg^{2+}$  y 2CABP. Los SEQ ID NO para las estructuras son 17, 16 y 14, respectivamente, como se muestra en la Tabla 1. Los números en los extremos de las líneas denotan el número de secuencia. Los números uniformemente espaciados en la parte superior son marcadores de alineamiento, mientras que los números en negrita indican el número de secuencia para la espinaca.

20

25

30 Las **Figuras 6A y 6B** proporcionan secuencias consenso del 50% de la LSU de Rubisco de algas rojas (*Rhodophyta*; 9 especies), cianobacterias (11 especies) y plantas con flores (*Magnoliophyta*; 134 especies). Estas secuencias se presentan en la lista de secuencias como SEQ ID NO: 2, 3 y 11, respectivamente, como se muestra en la Tabla 1. El sombreado gris claro indica los 134 residuos que son los mismos en plantas con flores y cianobacterias pero diferentes en algas rojas, es decir, los Residuos Variantes. Los huecos en la secuencia se muestran con "~". Las definiciones de letras mayúsculas y minúsculas y los símbolos "-" y "+" son las mismas que en la Figura 5. Los números en los extremos de las líneas denotan el número de secuencia. Los números uniformemente espaciados en la parte superior son marcadores de alineamiento, mientras que los números en negrita indican el número de secuencia para la espinaca.

35

40 Las **Figuras 7A y 7B** proporcionan secuencias consenso del 50% de las LSU de Rubisco de algas rojas (*Rhodophyta*; 9 especies), cianobacterias (11 especies) y plantas con flores (*Magnoliophyta*; 134 especies), las mismas secuencias mostradas en la Figura 6. Las posiciones en las que se conservan los residuos de aminoácidos entre las tres secuencias consenso se muestran mediante el símbolo de una sola letra para las algas rojas (línea superior) y puntos en las líneas para las cianobacterias y las plantas. Las posiciones donde el residuo de aminoácido es el mismo para las cianobacterias y las plantas como para las algas rojas, se muestran por el símbolo de una sola letra para las algas rojas (línea superior) y puntos en las líneas de cianobacterias o plantas. Las posiciones en las que el residuo de aminoácido es el mismo en las cianobacterias y las plantas pero diferentes en las algas rojas, es decir, los 134 Residuos Variantes mostrados en la Figura 6, se muestran mediante el símbolo de una sola letra para las algas rojas (línea superior) y el símbolo de una letra para las cianobacterias (segunda línea), mientras que la posición del residuo tiene un espacio en blanco para las plantas (tercera línea). El sombreado inverso resalta aquellas posiciones de VR que se seleccionaron como Residuos Candidato para la Región 1, como se muestra en la Tabla 2, y que se agruparon en Mutantes Candidato pronosticados, como se muestra en las Tablas 3 y 4. Las posiciones en las que los residuos de aminoácidos son diferentes en las tres secuencias consenso se muestran con los símbolos de una sola letra en las tres líneas, resaltando el sombreado de color gris aquellas posiciones seleccionadas como Residuos Candidato Divergentes para la Región 1, como se muestra en la Tabla 2, y que fueron agrupadas en los Mutantes Candidato pronosticados, como se muestra en las Tablas 3 y 4. El símbolo "~" denota huecos de secuencia. Las definiciones de letras mayúsculas y minúsculas y de símbolos "-" y "+" son las mismas que en la Figura 5. Los números en los extremos de las líneas denotan el número de secuencia. Los números uniformemente espaciados en la parte superior son marcadores de alineamiento, mientras que los números en negrita indican el número de secuencia para la espinaca.

45

50

55

60

La **Figura 8** proporciona una vista lateral de la estructura del barril TIM C-terminal (residuos 151-475) de un polipéptido LSU y el dominio N-terminal (residuos 1-150) de una LSU adyacente que comprende la unidad que alberga un sitio activo de Rubisco. Muestra las posiciones de los Residuos Diana, E60, N123 y Y20 para la Región 1, H294 para la Región 2 y K334 para la Región 3. Esta estructura se dibujó utilizando coordenadas de átomos de la estructura de rayos x hexadecamérica  $L_8S_8$  completa de Rubisco de espinaca de tipo salvaje en el complejo con  $Mg^{2+}$  y 2CABP (pdb 8ruc). 2CABP se muestra como un modelo CPK. Los residuos E60 y N123 se localizan en la hélice  $\alpha B$  y cerca del extremo C-terminal de la hélice  $\alpha C$ , respectivamente, del dominio N-terminal. Las cadenas laterales de estos dos residuos están situadas a ambos lados del grupo carboxilato de 2CABP. Los hélices  $\alpha B$  y  $\alpha C$  tienen interacciones hidrófobas con las hebras  $\beta$  del dominio N-terminal. Y20 se encuentra en una estructura de una sola bobina cerca del extremo N-terminal de  $\beta A$ . Tanto K334 como H294 se localizan en el dominio C-terminal que contiene la mayoría de los residuos de la primera envoltura del sitio activo. K334 está en el bucle 6, mientras que H294 está en  $\beta 5$ .

La **Figura 9** proporciona una vista de la estructura del dominio N-terminal del polipéptido LSU de Rubisco (utilizando las coordenadas de la estructura de rayos X de espinaca 8ruc) con las cadenas laterales de los residuos mutados (G23, R81, W25 e I51) para el Mutante Núm. 6 (T23G, Y25W, D51<sup>Gp</sup>, K81R), y los Mutantes componente Núm. 4 (T23G, K81R) y Núm. 1a (Y25W, DS1<sup>Gp</sup>), modelado como modelos de barras. Véase la Tabla 3 para las definiciones de mutantes. Muestra las posiciones relativas de estos residuos en la estructura, y las interacciones entre los residuos 25 (W25) y 51 (I51), y entre los residuos 23 (G23) y 81 (R81). Las mutaciones añaden una interacción hidrófoba entre los residuos W25 e I51, en comparación con Y25 y D51 de tipo salvaje, y eliminan una interacción de unión de hidrógeno entre los residuos 23 y 81 (T23 de tipo salvaje, K81 a mutante G23, R81). Individualmente o combinadas, se pronostica que estas dos mutaciones dobles alteran la orientación del residuo Y20. Las posiciones de los Residuos Diana Y20, E60 y N123 se muestran como referencia como modelos de esferas y barras, mientras que el mimético del intermedio de reacción 2CABP se muestra en forma de modelo de alambre.

La **Figura 10** proporciona una vista de la estructura del dominio N-terminal del polipéptido LSU de Rubisco (utilizando las coordenadas de la estructura de rayos X de espinaca 8ruc) con las cadenas laterales de los residuos mutados (W25, I51, A54, A84, V87) para el Mutante Núm. 7a (Y25W, D51<sup>Gp</sup>, G54A<sup>Gp</sup>, C84A, 187V), y los Mutantes Componente Núm. 1a (Y25W, D51<sup>Gp</sup>) y Núm. 5a (G54A<sup>Gp</sup>, C84A, 187V), modelados como modelos de barras. La figura ilustra la formación de una región hidrófoba por mutaciones en los sitios 54 (G  $\rightarrow$  A), 84 (C  $\rightarrow$  A) y 87 (I  $\rightarrow$  V). Adicionalmente, si G54 de tipo salvaje fue mutado a S54<sup>Gm</sup> en lugar de A54<sup>Gp</sup> (Mutante Núm. 7b, Núm. 5b), se pronostica que se introduzca un enlace de hidrógeno adicional con la cadena principal del residuo 51. A su vez, I51 tiene una interacción hidrófoba con el residuo W25. Se pronostica que todos estos cambios afectan al posicionamiento del Residuo Diana Y20, mostrado en forma de esferas y barras. Las posiciones de los otros residuos de la Zona 1 de destino E60 y N123 se muestran como referencia como modelos de esferas y barras, mientras que el mimético intermedio de reacción 2CABP se muestra en forma de modelo de alambre. Como referencia, también se muestran las posiciones de las mutaciones para el Mutante Núm. 4 (G23, R81).

**Figura 11** proporciona una vista del dominio N-terminal del polipéptido LSU de Rubisco (utilizando las coordenadas de la estructura de rayos X de espinaca 8ruc) con las cadenas laterales de los residuos mutados (V<sup>Gp</sup> o I<sup>Gm</sup>36, L116, I121, L<sup>Gp</sup> o I<sup>Gm</sup>140, G297, T300) para el Mutante Núm. 10a (L36V, I116L, V121I, F140L, M297G, V300T) y otras variantes del Mutante Núm. 10 Gp/Gm (Núm. 10b, c, d) y los Mutantes componente 8 (V121I, M297G, V300T) y Núm. 9a (L36V, I116L, F140L) y otras variantes del Mutante Núm. 9 Gp/Gm (Núm. 9b, c, d), modeladas como modelos de barras. Los residuos A296 y V271 que no están mutados también se muestran como modelos de barras. Para facilitar la comprensión de los cambios en las interacciones pronosticadas que resultan de las mutaciones, el recuadro remarcado RHS muestra los residuos correspondientes en el *Synechococcus* de tipo salvaje como modelo de barras. La figura ilustra la interrupción pronosticada de una interacción hidrófoba entre los residuos V300, M297 y V121 en el tipo salvaje para formar una nueva interacción hidrófoba entre A296, V271 e I121 en el mutante (Núm. 8 o Núm. 10a). La predicción se basa en comparaciones estructurales entre las estructuras de rayos X para complejos de espinaca (pdb 8ruc), *Synechococcus* (pdb 1rb1) y *Galdieri partita* (pdb 1bwv) con  $Mg^{2+}$  y 2CABP. Una nueva interacción hidrófoba entre los residuos 36, 140 y 116 también está pronosticada por las mutaciones L  $\rightarrow$  V, F  $\rightarrow$  L e I  $\rightarrow$  L. Individualmente o combinadas, se pronostica que estas nuevas interacciones hidrófobas mediadas por estos grupos de mutaciones afectan a la posición y la orientación del Residuo Diana N123, que se muestra en el modelo de esferas y barras. Las posiciones de los otros Residuos Diana de la Región 1 E60 y Y20 se muestran como referencia como modelos de esferas y barras.

**Figura 12** proporciona una vista estéreooscópica del dominio N-terminal de la LSU de Rubisco de espinaca (utilizando las coordenadas de la estructura de rayos X de espinaca 8ruc). Las cadenas laterales de todos los residuos N-terminales se muestran en modelo de estructura de alambre. Los Residuos Diana (Y20, E60 y

N123) se muestran en el modelo de barras. El análogo carboxilado intermedio 2CABP y  $Mg^{2+}$  se muestran como modelos de esferas y barras. La figura muestra una representación 3-D completa de todas las cadenas laterales de residuos del dominio N-terminal de las que se seleccionan los CR que se pronostica que influyan en los TR Y20, E60 y N123, como se muestra en la Figura 13. Como referencia, la orientación de la estructura es la misma que en la Figura 15.

La **Figura 13** proporciona una vista estereoscópica del dominio N-terminal de la LSU de Rubisco de espinaca (utilizando las coordenadas de la estructura de rayos x de espinaca 8ruc), con los 20 Residuos Candidato y 6 Residuos Candidato Divergentes enumerados listados en la Tabla 2 mostrados como modelos de barras y los Residuos Diana marcados N123, E60 y Y20. La figura muestra una representación tridimensional completa de las cadenas laterales de CR y DCR en el dominio N-terminal que se pronostica que influyan en los TR Y20, E60 y N123, así como un segmento pequeño del dominio C-terminal de la LSU asociada mostrando los CR M297 y V300 que también se pronostica que influyan en estos TR. Los residuos marcados T23 y K81 son aquellos que están mutados en el Mutante Núm. 4 (T23G, K81R), mientras que los residuos marcados con Y25 y E51 se mutan adicionalmente en el Mutante Núm. 6 (T23G, Y25W, D511<sup>GP</sup>, K81R).

Las **Figuras 14A y 14B** proporcionan un alineamiento de secuencia de aminoácidos de secuencias consenso del 50% del dominio N-terminal de la LSU de Rubisco de plantas con flores (*Magnoliophyta*; 134 especies), cianobacterias (11 especies) y algas rojas (*Rhodophyta*; 9 especies), así como secuencias de varias plantas (espinaca, tabaco, arroz, soja y caña de azúcar), *Synechococcus* sp. PCC6301 y especies de algas rojas (*Galdieri partita* y *Griffithsia monolis*) de interés especial. Las secuencias presentadas en esta figura también se exponen en las SEQ ID NO 2, 3 y 11, utilizadas anteriormente en las Figuras 5-7 y 14-21. Éstas se enumeran con los números de acceso de la base de datos en la Tabla 1. La secuencia completa se muestra en una sola letra y forma de símbolo para el consenso de cianobacterias (véase la Figura 5 para una explicación de los símbolos), con residuos de aminoácidos en otras secuencias mostradas por puntos si se conservan o por el formato de una letra/símbolo si no se conservan. Los números en los extremos de las líneas denotan el número de secuencia. Los números uniformemente espaciados en la parte superior son marcadores de alineamiento, mientras que los números en negrita indican el número de secuencia para la espinaca. Las franjas verticales sombreadas de forma invertida en color negro ponen de relieve los 17 Residuos Candidato actuales (es decir, aquellos residuos comunes entre secuencias consenso del 50% de las plantas con flores y las cianobacterias pero diferentes en algas rojas) en el dominio N-terminal: residuos 18, 19, 23, 25, 54, 59, 64, 68, 81, 84, 87, 88, 104, 114, 118, 121. Las franjas verticales sombreadas en color gris ponen de relieve los 6 residuos Candidato Divergentes actuales (es decir, aquellos residuos que difieren en plantas verdes, cianobacterias y algas rojas): residuos 36, 86, 116, 117, 138, 140.

La **Figura 15** proporciona una vista del dominio N-terminal y segmentos del dominio C-terminal de la subunidad adyacente de la LSU de Rubisco de espinaca (utilizando las coordenadas de la estructura de rayos X de espinaca 8ruc). En la parte inferior el intermedio de la reacción  $\beta$ -ceto (2-carboxi-3-cetoarabinitol 1,5-bisfosfato (2C3KABP), III, véase la Figura 1),  $Mg^{2+}$  y otros residuos del dominio C-terminal de la LSU asociada coordinados en la primera envoltura del sitio activo (véase la Figura 2) y su disposición a los Residuos Diana Y20, E60 y N123. Y20, E60, N123 y el intermedio de reacción  $\beta$ -ceto se muestran como modelos de barras gruesas, mientras que los residuos del sitio activo se muestran en un modelo de barras finas. La figura muestra la Subregión 1A en el dominio N-terminal y la Subregión 1B en el dominio N-terminal y segmentos del dominio C-terminal de la subunidad adyacente. La Subregión 1A incluye las posiciones de CR18 y nuevos CR 19, 68 y 104 que se espera que mejoren el Mutante Núm. 4 (T23G/K81R) si se mutan en combinaciones (Núm. 17-1A (Núm. 4 + K18I/T68V), Núm. 18-1A (Núm. 4 + P104E (D)), Núm. 19 (Núm. 4 + D19P)) y otras combinaciones de los mismos. La Subregión 1B incluye posiciones en el nuevo Mutante Candidato Núm. 20-1B (V116L/L117T/V121I/I138M/F140L) que contiene los DCR 116, 117, 138 y 140, CR 121 y el residuo 135 totalmente conservado.

Las **Figuras 16A y 16B** proporcionan un gráfico con una vista ampliada de las interacciones en el Mutante Candidato Núm. 20-1B (V116L/L117T/V121I/I138M/F140L) en la Subregión 1B. El panel A muestra como referencia el dominio N-terminal completo, las posiciones de los TR N123, E60 e Y20 y el centro de sitio activo (2CABP,  $Mg^{2+}$ ) y la hélice C y la hebra E. El panel B muestra cómo interactúan los residuos 116, 117, 120, 121 y 135, 138 y 140 del componente situados en la hélice C y la hebra E, y cómo se pronostica que los cambios en estas interacciones influirán en TR N123 cercano; otra vista se encuentra en la Figura 17. 2CABP y  $Mg^{2+}$  se muestran en el modelo CPK. N123, E60 y Y20 se muestran en el modelo de barras gruesas.

La **Figura 17** proporciona un gráfico del dominio N-terminal de la LSU de Rubisco y segmentos del dominio C-terminal de la subunidad adyacente como un modelo de cinta que muestra las posiciones relativas de las tres Subregiones 1A, 1B y 1C y algunos de los CR y DCR en estas Subregiones. Algunas de las interacciones pronosticadas clave utilizadas en el diseño de Mutantes Candidato de la Región 1 se muestran mediante

líneas que conectan los CR y DCR.

La **Figura 18** proporciona un modelo estructural utilizado en los cálculos QM/QM de ONIOM realizados con DFT como capa alta y QM semi-empírica (PM3)) como capa baja. La capa DFT (mostrada como modelo de esferas y barras grandes) comprende 93 átomos incluyendo fragmentos de todos los residuos en la primera capa de coordinación del sitio activo y el sustrato: el átomo de magnesio ( $Mg^{2+}$ ), enodiolato de RuBP, GLU60, ASN123, LYS175, LYS177, LYS201 carbamida, ASP203, GLU204, HIS294, LYS334. Esta capa central está rodeada por otros 711 átomos en la capa de PM3 (QM semi-empírica), que comprende residuos de aminoácidos de hasta  $\sim 12\text{\AA}$  del  $Mg^{2+}$ . Las coordenadas de partida se tomaron de la estructura de rayos X del complejo de Rubisco-2CABP de espinaca (pdb 8ruc). El tamaño de todo el sistema es de 804 átomos.

La **Figura 19** proporciona un gráfico que muestra la estructura central del sitio activo de Rubisco y la división de la estructura que rodea los Residuos Diana en tres regiones espacialmente contiguas. La Región 1 comprende residuos de aminoácidos que pueden influir en los TR GLU60, ASN123 y TYR20 y comprende el dominio N-terminal y segmentos pequeños del dominio C-terminal de la subunidad adyacente. Las regiones 2 y 3 están en el dominio C-terminal y comprenden residuos de aminoácidos que pueden influir en los TR HIS294 y LYS334, respectivamente.

**Figura 20** proporciona un gel que muestra la separación de Rubisco  $L_2$  de *R. rubrum* y Rubisco  $L_8S_8$  de tabaco mediante electroforesis de poliacrilamida no desnaturante. La proteína soluble de hoja de tabaco de las líneas de tipo salvaje, del Mutante Núm. 4 de tabaco y del Mutante Núm. 23-1B de tabaco utilizadas para las mediciones de cinética se separó mediante electroforesis de poliacrilamida no desnaturante y se visualizó mediante tinción con Coomassie. La Rubisco  $L_2$  que está presente en la línea tabaco-*rubrum*  $\Delta aad$  (tr) y en la línea heteroplásmica Núm. 4 transformante del Mutante Núm. 4 de tabaco (es decir, produce Rubiscos  $L_2$  y  $L_8S_8$  mutadas) no está presente en las otras líneas mutantes de tabaco o transformantes de tabaco de tipo salvaje. Se indican las proteínas marcadoras con tamaños (m). Se muestra la cantidad de proteína soluble de hojas cargada por calle.

### Descripción detallada

La presente invención proporciona un método para generar una proteína con una propiedad funcional mejorada. La invención comprende un procedimiento para reducir el espacio de secuencia para la prueba mutacional convencional mediante la definición de mutantes con una o, preferiblemente, múltiples mutaciones utilizando información mecanicista (computacional) y bioinformática y de bases de datos (cambios de secuencia específicos de la filogenia/específicos del entorno, datos cinéticos, estructura 3-D y modelado), en primera instancia. Este procedimiento busca maximizar el uso de toda la información disponible, particularmente adaptaciones evolutivas parciales codificadas en secuencias de Rubisco, de modo que el procedimiento de mejora de proteínas para la propiedad funcional requerida comienza a un nivel funcional más cercano al nivel requerido de eficiencia funcional.

Una primera etapa del método comprende el procedimiento de identificación de Residuos de aminoácido Diana en una primera proteína. Como se describe con más detalle a continuación, el procedimiento de identificación de Residuos de aminoácido Diana puede comprender cálculos QM de fragmentos del sitio activo (etapa (i)) utilizando, por ejemplo, métodos DFT) y cálculos QM/QM o QM/MM híbridos (etapa ii) utilizando, por ejemplo, métodos ONIOM), junto con el uso de datos empíricos (por ejemplo, datos cinéticos y estructuras de cristales de rayos X). Se pueden usar simulaciones de dinámica molecular (MD) (etapa (iii)), y cálculos utilizando combinaciones de métodos QM, QM/QM o QM/MM y MD (etapas (iv) y (v)) para evaluar el estabilidad de las proteínas Mutantes Candidatas pronosticadas, o para una comprensión más detallada de las funciones de los Residuos Diana en la reacción enzimática. Estos Residuos de aminoácidos Diana serán normalmente residuos fundamentales para el correcto funcionamiento de la proteína con respecto a la propiedad funcional bajo investigación. Es una característica preferida de la invención que la propiedad funcional mejorada se consiga modificando las propiedades químicas de estos residuos (por mutación de otros residuos) para mejorar, por ejemplo, su actividad cinética.

Convenio para la numeración de residuos de aminoácidos de Rubisco

A lo largo de esta memoria descriptiva, cuando se identificó un residuo de una LSU de Rubisco mediante un número, la numeración de los residuos se basó en la numeración del residuo de la secuencia de aminoácidos de la LSU de Rubisco de espinaca (SEQ ID NO: 17). Esta convención de numeración se utilizó para todos los residuos identificados en la química computacional, en las estructuras de proteínas y en las mutaciones. Esto no crea ambigüedad ya que estos números pueden ser asignados a números de secuencia en alineamientos y comparaciones estructurales y, en consecuencia, un determinado número de residuo de espinaca puede ser mapeado con total confianza con respecto a un número de residuo de cianobacterias o de algas rojas estructuralmente equivalente.

**Listas de secuencias**

5 La Tabla 1 proporciona un resumen de las secuencias de aminoácidos de Rubisco de origen natural y consenso del 50% comentadas en la presente memoria y que se proporcionan en la lista de secuencias legibles por ordenador, con los SEQ ID NO que se muestran en la columna "SEQ ID". Cuando se usaron secuencias múltiples para producir una secuencia consenso del 50%, el número total de secuencias implicadas en la creación de la secuencia consenso se enumera entre paréntesis en la columna "Descripción". Los números de acceso a la base de datos proporcionan identificadores únicos para cada una de las secuencias que se utilizaron, incluidas las secuencias que se consideraron en la creación de secuencias consenso. Los números de las figuras en las que se utilizan las secuencias en los alineamientos también se proporcionan en la columna de "Núm. de Figura".

**Tabla 1:** Números de acceso a la base de datos para las Secuencias LSU de Rubisco utilizadas en las Figuras 5, 6, 7 y 14. Los números de ID de secuencia corresponden a los de la lista de secuencias legibles por ordenador.

SEQ ID NO:	Descripción	Núm. de Figura	Números de acceso a la base de datos			
1	Protista glaucophyta	5	P24312			
2	Protista rhodophyta (9)	5,6,7,14	BAA75676	AAB17222	BAA75796	AAR13681
			ABU53651	BAE78417	CAB58236	AAD04746
			BAE78409			
3	Eubacteria cyanobacteria (11)	5,6,7,14	P00879	400856410	4797620	402041420
			5199140	P27568	403124030	Q8DIS5
			4296100	403238160	P00880	
4	Plantae anthocerotophyta	5	Q31795			
5	Plantae bryophyta (28)	5	Q95G53	Q76GQ2	Q5TM96	Q75W61
			Q76GQ0	Q50L57	Q5TMB1	Q75W60
			Q76HL3	Q5TM93	Q9BB41	Q95G63
			Q94N80	Q5TMA1	Q9TM58	Q75VP7
			Q8HW62	Q8SN97	Q95G62	Q9TM63
			Q5TM99	Q5TMA5	Q9GIF5	Q5TMA3
			Q5TM97	Q9GGM2	Q5TM95	NP_904194
6	Plantae charophyta (4)	5	Q8SN66	P48716	Q32RY7	Q32RQ1
7	Plantae chlorophyta (4)	5	NP_958405	BAE48225	AAD00447	BAC06367
8	Plantae coniferophyta	5	P41621			
9	Plantae equisetophyta	5	P48702			
10	Plantae gnetophyta	5	Q9THI3			
11	Plantae magnoliophyta (134)	5,6,7,14	Q3T5C7	Q5EKM0	Q95E10	Q06021
			Q31857	Q9MRW9	Q3V6P6	Q9GDM8
			Q3V6M3	Q31670	063085	Q3T5G3
			Q5I221	Q06022	Q37167	AAF78948
			Q6R615	Q3T5C1	P48688	Q06023
			P48690	Q75VD8	Q42664	Q7YKF9
			Q9XQE3	Q9XPK2	Q8WJD8	P48693
			Q5EKL6	063123	Q8SLM3	Q7YKF8
			CAA57001	Q42674	P92255	098530

ES 2 640 747 T3

SEQ ID NO:	Descripción	Núm. de Figura	Números de acceso a la base de datos			
			Q9XQB9	Q6R613	Q32072	Q3T575
			Q9XQE7	Q5EKM4	Q8WLJ3	Q95F13
			Q3L237	Q5EKM2	P92287	Q37319
			Q33449	Q9ZT30	Q32188	Q9SB16
			Q9BBC7	P19161	P48703	CAB08877
			Q9GGC1	Q42828	Q4VWN7	Q5XLF7
			Q95F23	Q3T5G1	Q95BC5	Q95F12
			Q32488	Q5EKL5	AAK72524	Q01873
			Q75VD6	Q9BBU1	Q32518	Q95EH6
			Q9GHT6	Q8WIB0	Q32622	Q9MVF1
			Q42916	098611	Q5EKL8	Q3T5E7
			Q32685	NP_054507	Q5C9P7	BAA00147
			Q9GHS0	Q68RZ8	Q3T5F6	Q7YL87
			Q37257	Q36849	Q95F15	Q95F24
			Q95A48	Q32916	P04717	Q8WGU4
			Q9XK53	Q9XQA7	Q32820	Q8ME88
			Q5MB28	Q5EKL2	Q8M962	Q75VD3
			098612	Q8WKT8	Q6R614	Q9MTS7
			Q8WIA8	Q95F20	AAX44985	062943
			Q8WIC4	P48715	ABB90049	Q9XQ93
			Q9GHN0	Q3T5E2	Q33064	Q37281
			Q8WIC3	Q9XPK3	Q6R617	Q8LUX7
			Q8WKR2	AAP92166	Q6USP5	Q3T5E4
			P28459	Q95F10	P92364	CAA60294
			Q75VD7	NP_054944		
12	Plantae pinophyta (2)	5	P26961	P26962		
13	Plantae pteridophyta (2)	5	Q85WR7	Q33015		
14	Galdieria partita	14	BAA75796			
15	Griffithsia monolis	14	ABU53651			
16	Synechococcus elongatus PCC6301	14	P00880			
17	Espinaca (Spinacia oleracea)	14	NP_054944			
18	Tabaco (Nicotiana tabacum)	14	NP_054507			
19	Arroz (Oriza sativa)	14	BAA00147			
20	Soja (Glycine max)	14	YP_538747			
21	Caña de azúcar (Saccharum officinarum)	14	BAD27301			

Identificación de Residuos Diana - Mecanismo Computacional

Las etapas (i) y (ii) siguientes relacionadas con la identificación de Residuos de aminoácido Diana se realizaron utilizando el paquete de programas GAUSSIAN, para Cálculos de QM y ONIOM *ab initio* (Frisch *et al.*, 2004), pero hay varios otros programas patentados o de uso libre disponibles que se podrían utilizar alternativamente. La etapa (iii) utiliza el programa AMBER generalmente disponible (Case *et al.*, 2006) para realizar simulaciones MD de proteínas; esta capacidad también está disponible en muchos otros programas. Las etapas (iv) y (v) emplean teoría, protocolos y programas publicados para simulaciones de mecanismos enzimáticos (Gready *et al.*, 2006); los métodos semi-empíricos de simulación de QM/MM MD (Cummins y Gready, 1997, 1998, 1999, 2003, 2005, Cummins *et al.*, 2007) se implementan en el programa MOPS (Cummins, 1996).

#### 10 (i) Cálculos QM de fragmentos de sitio activo complejo

La siguiente descripción se refiere a cálculos con respecto a la LSU de Rubisco, en la que los residuos del sitio activo se conservan totalmente entre especies. Estos cálculos utilizan un método QM *ab initio* de alto nivel (B3LYP/6-31G (d, p)) para definir las energías y estructuras de las especies de reacción (sustrato, estado de transición (TS), intermedio y complejos de producto) en el mecanismo de reacción de Rubisco de múltiples etapas, como que se muestra en la Figura 1. Los cálculos se realizaron para las etapas de reacción a partir de la reacción de adición de gas, utilizando el conjunto de programas Gaussian 03 (Frisch *et al.*, 2004) y con las coordenadas iniciales tomadas de la estructura de rayos X del complejo de Rubisco-2CABP de espinaca (pdb 8ruc). El modelo de fragmento de sitio activo, FM20, como se muestra en la Figura 2, era suficientemente grande para contener todos los residuos y moléculas de agua que interactuaban inmediatamente con el centro de reacción y el átomo de Mg coordinado en el sitio activo y para permitir la definición de sus funciones en las diferentes etapas de reacción, incluyendo las etapas de adición de gas claves (CO<sub>2</sub> u O<sub>2</sub>). Las estructuras de la especie de reacción y la ruta de la energía de reacción para la carboxilación y las etapas de reacción subsiguientes se muestran en las Figuras 3 y 4, respectivamente.

#### 25 (ii) Cálculos QM/QM y QM/MM híbridos de ONIOM:

Estos cálculos definen las perturbaciones de las energías y estructuras de las especies de la ruta de reacción y se centran principalmente en la etapa de adición de gas de los vecinos próximos más próximos y más allá de los residuos del sitio activo. Esto se hizo utilizando métodos que utilizan un alto, pero computacionalmente costoso, modelo QM *ab initio* para el núcleo del sistema (es decir, como en (i)) y un modelo QM (QM semi-empírico) o MM menos costoso para una región ampliada.

Los cálculos (QM/QM y QM/MM) se realizaron en varias etapas utilizando el módulo ONIOM en GAUSSIAN 03. Los cálculos de QM/QM de ONIOM utilizaron un modelo de una capa central de QM *ab initio* de alto nivel de 93 átomos. Para el estudio del punto de partida en la etapa de adición de gas, la capa central de QM comprende el átomo de magnesio (Mg<sup>2+</sup>), el enodiolato de RuBP (para calcular la estructura y energías de las especies reactivas subsiguientes se usaron las correspondientes especies químicas derivadas de RuBP), GLU60, ASN123, LYS175, LYS177, LYS201 carbamilada, ASP203, GLU204, HIS294 y LYS334. La capa central está rodeada por otros 711 átomos en la capa externa calculada en el nivel de PM3 (QM semi-empírica), que comprende residuos de aminoácidos hasta ~ 12 Å del átomo de magnesio. Las coordenadas de partida se tomaron de la estructura de rayos X del complejo de Rubisco-2CABP de espinaca (pdb 8ruc). Este modelo se ilustra en la Figura 18. El papel de estos cálculos es comparar los efectos de los residuos injertados en la vecindad del sitio activo con los del tipo salvaje sobre las energías y estructuras de las especies de la ruta de reacción, enfocadas principalmente en la etapa de adición de gas. Estos cálculos permitieron que las estructuras y energías de los fragmentos del sitio activo de 93 átomos de las especies de reacción, análogas a las del modelo QM FM20 mostrado en las Figuras 3 y 4, fueran re-optimizadas en el entorno de los residuos enzimáticos circundantes (la capa externa de 711 átomos), y, por lo tanto, permitió las perturbaciones del perfil de energía de la reacción debido a los residuos injertados a estimar. Estos cálculos pueden proporcionar así una información valiosa sobre las perturbaciones del mecanismo básico de, en este caso, Rubisco de espinaca de tipo salvaje a determinar y, en particular, los detalles de la estructura de CO<sub>2</sub> y el grupo carboxilato naciente en la etapa de adición de gas. Por estos medios, se puede calcular la magnitud y dirección de las perturbaciones electrostáticas de las interacciones de Residuos Diana con especies de reacción debidas a residuos injertados en Mutantes Candidato. Esta información puede usarse para la pre-selección de Mutantes Candidato para el ensayo experimental y/o utilizarse en la interpretación de los resultados de las pruebas experimentales, como se muestra en el diagrama de flujo 1, columna de la izquierda.

#### 60 (iii) Simulaciones de dinámica molecular (MD) de complejos de proteínas

Estas simulaciones evaluaron si la estructura proteica de los Mutantes Candidato de Rubisco injertados podría acomodar los residuos cambiados, es decir, si la estructura de proteína mutante era conformacionalmente estable o si tendía a deshacerse. Las simulaciones MD son particularmente útiles para mutantes con injerto múltiple y proporcionan una prueba de escrutinio de la estabilidad global para complementar las pruebas electrónicas sobre el

mecanismo químico de (ii).

Estos cálculos se realizaron con el paquete de programas AMBER8 o AMBER9 (Case *et al.*, 2006), pero se podrían utilizar otros paquetes de simulación de MD de proteínas (p.ej. GROMACS) para obtener resultados similares.

5 (iv) Cálculos QM/QM o QM/MM híbridos ONIOM múltiples de diferentes estados conformacionales muestreados de complejos

10 En este método se lleva a cabo una serie de cálculos utilizando coordenadas para complejos de proteínas (p.ej., para diferentes etapas de reacción) tomadas de instantáneas de trayectorias de simulaciones MD QM/MM, como describen Gready *et al.* (2006). Estos cálculos permitieron un examen más detallado de las características de la ruta catalítica, a saber, los efectos de la flexibilidad conformacional de la proteína sobre las geometrías del complejo de la enzima y las energías de activación y reacción.

15 (v) Generación de las superficies de energía libre de reacción completa para las reacciones de adición de gas

Se puede generar un conjunto estadístico completo (promedio conformacional) de los estados enzimáticos sobre el curso completo de una etapa de reacción a partir de simulaciones de MD QM/MM semi-empírica para una cuadrícula de puntos definida por las coordenadas de reacción (una hipersuperficie de energía libre). A continuación se puede generar una superficie de energía libre más precisa a nivel QM *ab initio* mediante cálculos QM/MM ONIOM utilizando múltiples configuraciones, por ejemplo, hasta 120, muestreadas a partir de los puntos de la hipersuperficie de reacción QM/MM semi-empírica (Gready *et al.*, 2006; Cummins *et al.*, 2007). Estas superficies de energía libre enzimática proporcionan energías libres de reacción y activación que pueden compararse directamente con datos experimentales, tales como constantes cinéticas medidas experimentalmente, y también pueden usarse para calcular diferencias en las energías libres de reacción y activación entre el tipo salvaje y los mutantes.

vi) Definición del mecanismo de reacción y Residuos Diana

30 Basándose en los resultados de estos cálculos, los autores de la presente invención fueron capaces de deducir un mecanismo para toda la secuencia de reacciones en la catálisis de la carboxilasa, y de definir funciones precisas para los residuos del sitio activo, individualmente y en conjunto. A partir de los cálculos del fragmento QM, se identificaron un par de residuos de aminoácidos clave, uno que actuaba como una base y el otro que actuaba como un ácido, para cada etapa de reacción. En particular, se identificaron el par HIS294 y LYS334 para la etapa de adición de gas.

35 Para la reacción de carboxilasa de Rubisco, el punto de partida es el complejo de Rubisco con la forma enodiolato de RuBP unida al sitio activo y la molécula de CO<sub>2</sub> mantenida a una distancia de interacción de van der Waals con respecto al carbono C2 del enodiolato. Este estado se representa como I en las Figuras 1, 3 y 4. La reacción transcurre a través de un estado de transición caracterizado por la formación de un enlace covalente parcial entre el átomo de carbono de CO<sub>2</sub> y el carbono C2 del enodiolato, acompañada de dobles enlaces parciales entre los átomos de carbono C2 y C3 y entre los átomos C3 y O3 del enodiolato. Este estado se representa como II en las Figuras 3 y 4. La etapa de reacción termina con la formación completa de un enlace covalente entre la molécula de gas y el carbono C2 del intermedio β-ceto (2C3KABP). Este estado se representa como III en las Figuras 1, 3 y 4.

45 En la etapa de adición de gas, HIS294 actúa como una base para separar completamente un protón del átomo O3; esto transfiere una carga negativa parcial al carbono C2 y le permite formar un enlace covalente con el átomo de carbono de CO<sub>2</sub>, y también transfiere la carga negativa al grupo carboxilato naciente. LYS334, que está cargado positivamente, ayuda a estabilizar esta carga negativa que se desarrolla en el grupo carboxilato naciente. Estas características se pueden observar en las estructuras I - III en la Figura 3. Por lo tanto, la alcalinidad de HIS294 y la acidez (carga) de LYS334 son cruciales para la etapa de adición de gas.

55 La modificación de las propiedades de estos dos residuos, por ejemplo, alterando estéricamente la orientación/distancia de sus interacciones con el sustrato enodiolato o el intermedio β-ceto o alterando electrónicamente la carga sobre los átomos que interactúan con el sustrato enodiolato o el intermedio β-ceto puede afectar la energética de la etapa de adición de gas. HIS294 y LYS334 se identifican, por lo tanto, como "Residuos Diana", definidos ampliamente como residuos que se pronostica que tienen un efecto significativo sobre el mecanismo de reacción y la energética. HIS294 y LYS334 están en el dominio C-terminal, están espacialmente separados, y afectan a diferentes partes del sustrato enodiolato o el intermedio β-ceto. Por lo tanto, los aminoácidos que pueden afectar a sus propiedades se han clasificado en diferentes regiones; Región 2 para His294 y Región 3 para Lys334, como se muestra en la Figura 19.

Aunque el residuo ASN123 no se incluyó en el modelo de fragmento de sitio activo FM20 para los cálculos QM, el

examen de las estructuras de cristal y los cálculos QM/QM preliminares sugieren que también está involucrado en la estabilización de la carga en el grupo carboxilato naciente añadido en C2. Además, el examen de las estructuras cristalinas mostró que los residuos E60 e Y20 están situados para alterar directamente la carga sobre LYS334 (es decir, la carga/orientación de LYS334 puede alterarse manipulando E60 e Y20) y el grupo C2-carboxilato del intermedio 2C3KABP. Por lo tanto, E60, Y20 y N123 también se identificaron como Residuos Diana. Estos tres residuos están en el dominio N-terminal de la LSU y, por tanto, los aminoácidos que pueden afectar a sus propiedades se clasificaron como pertenecientes a una región diferente (Región 1) de las de HIS294 y LYS334 (Figura 19). Estos tres residuos se conservan estrictamente en todas las LSU catalíticamente activas de Rubisco y se pronostica que actúan de manera concertada en la etapa de adición de gas. Esto puede medirse mediante la inspección de las Figuras 8, 9, 10, 12 y 15, que proporcionan diferentes vistas de la disposición relativa de ASN123, GLU60 y Tyr20 con respecto al grupo carboxilato del análogo intermedio  $\beta$ -ceto (2C3KABP). Estas figuras muestran que ASN123, GLU60 y Tyr20 están unidos a tres regiones de estructura secundaria separadas del dominio N-terminal y que sus cadenas laterales se extienden hacia el sitio activo en una constelación tripartita.

En resumen, el método anterior comprende una serie completa de métodos computacionales para investigar cuestiones mecanísticas, energéticas y de estabilidad a niveles globales o más detallados para las etapas de carboxilación y oxigenación para el tipo salvaje y cualquier Mutante Candidato pronosticado de Rubisco. De esta manera, fue posible identificar uno o más Residuos Diana para actuar como el foco para el injerto filogenético.

#### 20 Comparaciones de proteínas - Injerto filogenético

En su forma más amplia, el método descrito en la presente memoria también comprende la comparación de al menos una segunda proteína con al menos la primera proteína. La segunda proteína puede originarse a partir de la misma o de una rama filogenética diferente de la de la primera proteína. El procedimiento de comparación implica la identificación de al menos un residuo de aminoácido Variante entre la primera proteína y la segunda proteína. Una pluralidad de Residuos Variantes de la segunda proteína actúan como un conjunto de diferentes identidades de residuos de aminoácidos específicos que pueden "injerterse" sobre la primera proteína en un intento de mejorar la propiedad funcional de la primera proteína mediada por los Residuos Diana.

Tomando como ejemplo a Rubisco, los cambios filogenéticos específicos de rama en la secuencia de aminoácidos de Rubisco, tales como cambios en las Rubisco de grupos filogenéticos de diferentes linajes evolutivos o en las Rubisco que expresan cambios específicos del entorno, representan posibles optimizaciones parciales de la eficiencia catalítica de las Rubisco. Se desarrolló una estrategia, denominada "injerto filogenético", para identificar los residuos clave que representan estas soluciones evolutivas parciales y para "transplantar" selectivamente estos residuos a un anfitrión Rubisco, tal como una Rubisco de *Synechococcus* sp., cambiando los residuos específicos del anfitrión por los de la Rubisco donadora o el grupo donador de las Rubisco con una o más características cinéticas mejoradas (o preferidas), con vistas a producir una Rubisco anfitriona con estas características cinéticas mejoradas.

Las soluciones evolutivas parciales descritas anteriormente se identificaron mediante un procedimiento de combinación de los resultados de los estudios computacionales (los Residuos Diana), como se muestra en el Diagrama de Flujo 1, columna izquierda, con los del análisis filogenético (los Residuos Variantes), como se muestra en el Diagrama de Flujo 1, columna derecha, para seleccionar los Residuos Candidato, como se muestra en el Diagrama de Flujo 1, columna central. Estas soluciones se distribuyen entre los Residuos Candidato en cambios de secuencia conservados característicos (consenso) de la LSU entre diferentes ramas filogenéticas de Rubiscos, o en cambios entre las Rubisco de la misma rama que están mejor adaptadas a ambientes específicos, p. seco/húmedo o caliente/frío.

La integración de los resultados de los estudios computacionales con los del análisis filogenético para identificar un subconjunto específico de Residuos Variantes (es decir, los Residuos Candidato) permite la diferenciación entre residuos que pueden afectar a una propiedad funcional, por ejemplo, la eficacia de la adición de gas de otros cambios de secuencia conservados característicos (consenso) entre ramas filogenéticas, que pueden representar, por ejemplo, una derivación filogenética neutra o un papel fisiológico específico de la rama. Tomando el ejemplo de una enzima Rubisco, una función fisiológica específica de la rama puede incluir el plegamiento y ensamblaje de la proteína, incluyendo interacciones con la subunidad pequeña, o estabilidad de proteína.

#### (i) Identificación de Residuos Variantes mediante análisis filogenéticos

El uso combinado de los mecanismos deducidos computacionalmente para identificar Residuos Diana con conservación de secuencias e información filogenética con el fin de identificar los Residuos Variantes se ilustra mediante la siguiente discusión de factores de especificidad de Rubisco. Los factores de muy alta especificidad de Rubiscos de algas rojas pueden atribuirse a residuos que son comunes entre las Rubisco de cianobacteria y de

plantas con flores, pero difieren en las Rubisco de algas rojas. Dichos residuos se definen en la presente memoria como "Residuos Variantes". Si se identifican Residuos Variantes individuales o una pluralidad de Residuos Variantes que actúan como factores determinantes de la especificidad en las Rubisco de algas rojas y se incorporan selectivamente a las Rubisco de plantas con flores/cianobacterias, se puede producir una Rubisco que es fisiológicamente activa en el organismo anfitrión que tiene mayor especificidad para las emisiones de CO<sub>2</sub> que la enzima nativa.

Los radicales de la primera envoltura, es decir, aquellos residuos que se coordinan directamente con el centro de reacción (Glu60, Asn123, Lys175, LYS177, KCX201, Asp203, Glu204, His294 y Lys334) se conservan totalmente entre las Rubisco. Esta conservación se ilustra en la Figura 5, que muestra un alineamiento de secuencias de LSU de Rubisco de los organismos fotosintéticos pertenecientes a trece diferentes phyla que abarcan las algas rojas, cianobacterias, Glaucophyta y plantas (10 phyla). Cuando estaba disponible más de una secuencia de Rubisco en un phylum, se usó una secuencia consenso del 50% para representar ese phylum. Las secuencias consenso se obtuvieron utilizando el servidor en <http://coot.embl.de/Alignment/consensus.html>. La Figura 5 también muestra que la secuencia de LSU de Rubisco de 475 residuos (plantas y cianobacterias) está en general altamente conservada, incluyendo la ausencia casi completa de espacios a excepción de diferencias mínimas en los extremos N y C terminales.

Sin embargo, los residuos en la segunda envoltura y subsiguientes que rodean el centro de reacción muestran variación entre las principales ramas de Rubisco de plantas con flores, algas rojas y cianobacterias. Las algas rojas muestran la mayor especificidad para el CO<sub>2</sub>, como se identifica por la proporción de CO<sub>2</sub>/O<sub>2</sub> de ~160 en comparación con ~80 para las plantas verdes y ~40 para las cianobacterias. La variación de la secuencia entre las plantas con flores, las algas rojas y las cianobacterias se ilustra más claramente en el alineamiento de la Figura 6 que comprende solamente las secuencias consenso del 50% de las LSU de Rubisco de algas rojas (*rhodophyta*; 9 especies), cianobacterias (11 especies) y plantas con flores (*Magnoliophyta*; 134 especies), ya mostradas en la Figura 5. La Figura 6 muestra que hay 134 residuos que son los mismos en plantas con flores y cianobacterias pero diferentes en algas rojas, es decir, los Residuos Variantes. Los números de acceso a la base de datos para cada una de las secuencias utilizadas en los alineamientos de las Figuras 5 y 6 se proporcionan en la Tabla 1, junto con los SEQ ID NO para las listas de secuencias legibles por ordenador.

En este ejemplo, se identificaron 134 Residuos Variantes a partir de la LSU de Rubisco, entre los cuales se distribuyen los residuos responsables de la solución evolutiva parcial para mayor especificidad exhibida por las Rubisco de algas rojas. Estos se muestran como residuos de sombreados en color gris en el alineamiento de la Figura 6. Como los factores determinantes de la especificidad pueden ser codificados por combinaciones de varios Residuos Variantes, son posibles muchos miles de tales combinaciones. Por lo tanto, para ser de cualquier uso práctico, el subconjunto que comprende los determinantes de la especificidad debe seleccionarse de la lista de Residuos Variantes.

#### (ii) (a) Identificación de Residuos Candidato

Usando los métodos descritos a continuación, se identificaron Residuos Variantes específicos que tienen el potencial de afectar a la etapa de adición de gas de la reacción catalizada por Rubisco, y se denominaron "Residuos Candidato". Esto permitió cambios conservados entre ramas filogenéticas o sub-ramas/subespecies, que pueden representar una deriva filogenética neutra o que pueden tener un rol fisiológico específico de rama, como en la estabilidad, el plegamiento o el ensamblaje de la LSU de Rubisco, que deben ser desestimados.

Muchos de los Residuos Variantes pueden no contribuir a la mejora de la propiedad de la Rubisco y, en consecuencia, pueden no formar parte de la solución evolutiva para, en este ejemplo, incrementar las emisiones de CO<sub>2</sub> pero son mutaciones silenciosas o mutaciones relevantes para otras propiedades enzimáticas, tales como el plegamiento y el ensamblaje de la proteína, o su estabilidad, en la célula. Con el fin de identificar los Residuos Variantes más probables de formar parte de la solución evolutiva para la propiedad mejorada, en este ejemplo se incrementó la especificidad de CO<sub>2</sub>, se emplearon los conocimientos mecánicos obtenidos de los cálculos QM para seleccionar los Residuos Candidato de la pluralidad de Residuos Variantes. Este procedimiento se basó en la hipótesis de que los Residuos Variantes que pueden influir en la funcionalidad de los Residuos Diana identificados por la etapa de química computacional implicados en la etapa de adición de gas en Rubisco forman parte de la solución evolutiva. Este fue el principal criterio utilizado para seleccionar de los Residuos Variantes para obtener un subconjunto de residuos aquí denominado "Residuos Candidato".

En general, el procedimiento de selección se basó en la evaluación de la proximidad espacial de los Residuos Variantes a los Residuos Diana y la estimación y clasificación de su capacidad para influir en la electrostática y la orientación de los Residuos Diana. La selección puede utilizar la exploración visual de estructuras cristalográficas utilizando un paquete de programa de modelado molecular y visualización tal como Accelrys Discovery Studio v2.0

(Accelrys Software Inc., San Diego, CA, 2007), aunque podrían usarse otros paquetes de modelado similares. En una exploración inicial de los residuos para la selección se pueden utilizar conceptos químicos convencionales para interacciones intermoleculares, tales como emparejamiento electrostático de carga-carga, distancias típicas de van der Waals y de enlace de hidrógeno y modelos de llenado de espacio para cadenas laterales de aminoácidos. El procedimiento también puede ser sistematizado, por ejemplo, cartografiando todas las interacciones electrostáticas e hidrófobas átomo a átomo de cada uno de los Residuos Variantes con todos los demás residuos de aminoácidos que están dentro de una distancia de 3 - 5 Å y excluyendo aquellas interacciones que son equivalentes en las secuencias de la primera y segunda proteínas.

Ejemplos de tales interacciones equivalentes incluyen interacciones de la cadena principal-cadena principal que son generalmente, pero no siempre, inalteradas por mutación. Las interacciones también se consideraron equivalentes, por ejemplo, si una interacción hidrófoba en la secuencia de una proteína implicaba un carbono alifático  $\alpha$  o  $\beta$  de la cadena lateral de un Residuo Variante con un átomo de un Residuo no Variante y en la secuencia de la segunda proteína, el Residuo Variante de aminoácidos, aunque diferente de la secuencia de la otra proteína, también tenía un carbono alifático  $\alpha$  o  $\beta$  en la cadena lateral que interactuaba con el mismo Residuo no Variante que en la primera secuencia de proteína.

Las interacciones de grupos metilo en las cadenas laterales de aminoácidos, tales como las de valina, leucina e isoleucina se consideraron equivalentes si sólo el grupo metilo concreto estuviera implicado en la interacción con un Residuo no Variante. Los enlaces de hidrógeno formados por los grupos carboxilato de residuos de aspartato y glutamato también se consideraron equivalentes si las distancias de enlace de hidrógeno correspondientes eran similares.

Después de escrutar los Residuos Variantes para determinar las diferencias en los patrones de interacción entre las secuencias de la primera y segunda proteína, sólo se conservaron los Residuos Variantes que tenían el potencial de afectar a un Residuo Diana identificado a través de patrones de interacción cambiados. El potencial de un Residuo Variante para afectar a un Residuo Diana se reconoció por la interacción de un Residuo Variante con el Residuo Diana o con residuos de aminoácidos adyacentes a los Residuos Diana o con residuos de aminoácidos en la unidad estructural secundaria que albergaba el Residuo Diana. Incluso los Residuos Variantes que son partes de bucles, espiras o hebras no estructuradas, pero que están conectados a las unidades estructurales secundarias que albergan el Residuo Diana, tienen el potencial de alterar la orientación de un Residuo Diana ayudando al reposicionamiento de las unidades estructurales secundarias a través de interacciones cambiadas. Los Residuos Variantes seleccionados, que tenían una o más interacciones variantes y el potencial para influir en los Residuos Diana constituían el conjunto de "Residuos Candidato". Los 20 Residuos Candidato identificados como capaces de influir en los Residuos Diana en la Región 1, es decir, ASN123, Glu60 y TYR20, se identifican mediante sombreado inverso en el alineamiento de la Figura 7 y se muestran gráficamente en la Figura 13. También se enumeran en la Tabla 2.

#### (ii) (b) Identificación de Residuos Candidato Alternativos y Residuos Candidato Divergentes (ACR y DVR)

El esquema descrito anteriormente para la selección de un Residuo Candidato se basa en el criterio de selección de que el Residuo Candidato está presente en una secuencia consenso de una segunda proteína que exhibe una propiedad mejorada o deseable, siendo diferente del residuo correspondiente en una pluralidad de secuencias consenso de primeras proteínas que no exhiben la propiedad mejorada o deseable, y donde las secuencias consenso de las primeras proteínas comparten el mismo residuo. Así, utilizando el ejemplo de especificidad de  $\text{CO}_2$  en Rubisco, el Residuo Candidato será un residuo que está presente en la secuencia consenso de algas rojas, y que es diferente del residuo común a ambas secuencias consenso de plantas con flores y cianobacterias.

Otras soluciones evolutivas parciales pueden expresarse también en los residuos encontrados en el grupo de Residuos Variantes. Estos han sido denominados Residuos Candidato Alternativos y Residuos Candidato Divergentes. Estas soluciones parciales encontradas en la naturaleza pueden usarse para ampliar el procedimiento de selección de Mutantes Candidato, para producir un conjunto expandido de residuos alternativos o suplementarios con los que influir en los Residuos Diana.

Por ejemplo, cuando el residuo seleccionado de la segunda proteína no es el residuo mayoritario encontrado en la secuencia consenso de la segunda proteína, sino que se encuentra en menor frecuencia en una pluralidad de las segundas proteínas, aunque sigue siendo diferente de los residuos presentes en una pluralidad de secuencias consenso de las primeras proteínas, y donde las secuencias consenso de las primeras proteínas comparten el mismo residuo, el residuo se denomina un Residuo Candidato Alternativo (ACR). El ACR representa una alternativa al residuo consenso en una posición de Residuo Candidato. En cuanto al Residuo Candidato, un ACR debe satisfacer también los criterios de selección relacionados con la influencia de un Residuo Diana relevante. Un propósito para la introducción de ACR puede ser proporcionar residuos para injerto que se sospecha están

asociados con una mayor mejora de la propiedad deseable que el residuo mayoritario de la secuencia consenso de la segunda proteína.

5 Así, utilizando el ejemplo de la especificidad de CO<sub>2</sub> en Rubisco, el Residuo Candidato Alternativo será un residuo que es expresado por al menos una de las algas rojas que contribuyen a la secuencia consenso de algas rojas pero que no es el residuo mayoritario que está contenido en la secuencia consenso y que es diferente al residuo común a ambas secuencias consenso de plantas verdes y cianobacterias. La especie de alga roja *Griffithsia monolis* tiene una tasa catalítica más alta en comparación con otras especies de algas rojas, manteniendo la alta especificidad típica de las algas rojas (Whitney *et al.*, 2001), y por lo tanto la secuencia de *Griffithsia monolis* puede utilizarse como fuente de ACR. La variación de la secuencia del dominio N-terminal de Rubisco entre *G. monolis* y el de la secuencia consenso de algas rojas y una especie de alga roja típica (de referencia) *Galdieri partita*, se ilustra en la Figura 14. Esto muestra los ACR en los residuos 36, 51, 54, 88 y 104. Obsérvese que el residuo 36 es también un DCR.

15 Se puede identificar un Residuo Covariante (CvR) en la secuencia de una segunda proteína de una especie concreta, por estar en la proximidad de un Residuo Alternativo Candidato (ACR) y mostrar variación complementaria al ACR. Se puede sospechar que esta variación en la posición del CvR, que no está presente en la secuencia consenso para la segunda proteína, refleja cambios complementarios en las propiedades estructurales y/o electrostáticas del ACR y CvR. Estos cambios complementarios pueden constituir una solución evolutiva parcial. La identificación de los CvR proporciona un medio para identificar posiciones de residuo auxiliar que pueden ser mutadas en la primera proteína para acomodar mejor los cambios realizados a partir de los ACR de trasplante.

25 Otra solución evolutiva parcial más puede expresarse también en residuos hallados en el grupo de Residuos Variantes donde se examinan residuos de al menos tres ramas filogenéticas y se pronostica que el residuo seleccionado influye en al menos un Residuo Diana y se encuentra en la secuencia consenso de la segunda proteína de una rama mientras que es diferente de los residuos presentes en las secuencias consenso de al menos otras dos ramas, y donde las secuencias consenso de las primeras proteínas son también diferentes en la misma posición. Dicho residuo se denomina Residuo Candidato Divergente (DCR). En cuanto a un Residuo Candidato, un DCR debe también satisfacer los criterios de selección relacionados con la influencia en un Residuo Diana relevante.

30 Un propósito para introducir un DCR puede ser proporcionar residuos para injerto que se sospecha que están asociados con la propiedad mejorada o deseable, pero que también se puede esperar que produzcan una mayor variación de las propiedades expresadas cuando se sustituyen en primeras proteínas (anfitrionas) *diferentes* de las al menos dos ramas filogenéticas de lo que se esperaría si la sustitución fuera con un Residuo Candidato.

35 Así, utilizando el ejemplo de especificidad de CO<sub>2</sub> en Rubisco, el Residuo Candidato Divergente puede ser un residuo que se expresa en la secuencia consenso de algas rojas, pero que es diferente del residuo expresado en las secuencias consenso de plantas con flores y en la secuencia consenso de cianobacterias, mientras que al mismo tiempo los residuos de la planta con flores y las secuencias consenso de cianobacterias también difieren en esta posición. Se identifican seis Residuos Candidato Divergentes identificados como capaces de influir en los Residuos Diana en la Región 1 (ASN123, Glu60 y TYR20) mediante sombreado de color gris en el alineamiento de la Figura 7 y se muestran gráficamente en la Figura 13. La Tabla 2 muestra la variación en el residuo entre las plantas con flores y las cianobacterias en estas seis posiciones.

45 **Tabla 2.** Tabla de composición de residuos (en porcentaje) para cianobacterias, plantas con flores y algas rojas en los 20 Residuos Candidato actuales (incluidos tres del dominio C-terminal de la LSU asociada) y 6 sitios de Residuos Candidato Divergentes. El número de secuencias utilizadas se muestra entre paréntesis. Los DCR están en la parte inferior de la tabla (36, 86, 116, 117, 138, 140).

Núm. Res.	Cianobacterias (11)		Plantas con flores (134)		Algas rojas (9)		Sustitución en <i>Synechococcus</i> PCC6301 por residuos en			Subregión
	Res	%	Res	%	Res	%	algas rojas	plantas con flores	cianobacterias	
18	K	73	K	100	I	67	K18I		K18Q	1A
	Q	27			-	33				
19	D	81	D	92	P	67	D19P	D19E	D19E	1A
	E	18	E	8	-	33				
23	T	100	T	99	G	100	T23G	T23N		1A

ES 2 640 747 T3

Núm. Res.	Cianobacterias (11)		Plantas con flores (134)		Algas rojas (9)		Sustitución en <i>Synechococcus</i> PCC6301 por residuos en			Subregión
	Res	%	Res	%	Res	%	algas rojas	plantas con flores	cianobacterias	
			N	1						
25	Y	82	Y	98	W	100	Y25W	Y25H		
	W	18	H	2						
51	E	91	E	100	V	56	D51V		D51E	
	D	9			I	44	D51 I			
54	G	55	G	99	S	78	G54A	G54R		1C
	A	45	R	1	A	22	G54S			
59	A	100	A	100	G	100	A59G			
64	G	100	G	100	A	100	G64A			
68	T	100	T	99	V	100	T68V			1A
			A	1						
81	K	100	K	100	R	100	K81R			1A
84	C	100	C	100	A	89	C84A			1C
					C	11				
87	I	82	I	98	V	100	I87V	I87L		1C
	L	9	L	2						
	V	9								
88	E	100	E	100	D	78	E88D			1C
					E	22				
104	P	100	P	100	D	89	P104D			1A
					E	11	P104E			
114	T	100	T	100	A	100	T114A			
118	T	100	T	100	A	100	T118A			
121	V	100	V	100	I	100	V121I			1B
271	T	100	T	100	V	100	T271V			
297	M	100	M	100	G	100	M297G			
300	V	100	V	100	T	100	V300T			
36	L	73	I	100	V	89	L36V			
	I	18			I	11	L36I			
	V	9								
86 <sup>†</sup>	D	63	H	57	K	89	H86K	H86G	H86D	1C
	R	18	G	20	R	11	H86R	H86D	H86R	
			D	16						
116	V	64	M	99	L	100	I116L	I116M	I116V	1B
	M	18	L	1						
	I	18								

ES 2 640 747 T3

Núm. Res.	Cianobacterias (11)		Plantas con flores (134)		Algas rojas (9)		Sustitución en <i>Synechococcus</i> PCC6301 por residuos en			Subregión
	Res	%	Res	%	Res	%	algas rojas	plantas con flores	cianobacterias	
117	L	100	F	99	T	100	L117T	L117F		1B
			L	1						
138	I	82	L	100	M	100	I138M	I138L	I138L	1B
	L	18								
140	F	82	I	97	L	56	F140L	F140V		1B
	I	18	V	2	I	44	F140I	F140S		
			S	1						

† Para el residuo 86, sólo se muestran los residuos alternos con más del 5% de ocurrencia para cianobacterias y plantas con flores.

5 Como se muestra en la Tabla 2, la mayoría de los CR tienen un residuo alternativo en uno o más de los 3 grupos (cianobacterias, plantas con flores, algas rojas) aunque en la mayoría de los casos el residuo consenso tiene una frecuencia >70%. Las excepciones son 59, 64, 81, 114, 118 y 121, y los tres CR (271, 297 y 300) en los segmentos del dominio C-terminal de la LSU adyacente (Subregión 1B). Sin embargo, la magnitud de esta variación es de poca importancia ya que los conjuntos de secuencia para cada grupo no están filogenéticamente equilibrados. Por el contrario, esta variación sólo debe tomarse como una indicación aproximada del grado de conformidad con la definición de CR. El Mutante Núm. 4 más fuerte (T23G/K81R) (véanse los Ejemplos 7 y 11) casi no mostró variación. De manera similar, el mutante más fuerte de la Subregión 1B, Mutante Núm. 8 (V121I/M297G/V300T) (véase el Ejemplo 7), tampoco mostró ninguna variación de los 3CR.

(iii) Agrupamiento de Residuos Candidato, Residuos Candidato Alternativos y Residuos Candidato Divergentes en Mutantes Candidato

15 Más de un Residuo Candidato (CR), Residuo Candidato Alternativo (ACR) o Residuo Candidato Divergente (DCR), o combinaciones de los mismos, pueden contribuir a cambios en una única región de interacción contigua entre las secuencias de la primera y segunda proteína. Por ejemplo, un solo residuo de aminoácido no variante puede interactuar con dos CR, dos ACR, dos DCR, o una combinación de dos residuos derivados de dos de estos grupos, de tal manera que ambas interacciones cambiadas pueden afectar al mismo Residuo Diana. Los dos CR, dos ACR, dos DCR o una combinación pueden agruparse en un único Mutante Candidato. De manera similar, un CR, ACR o DCR dados pueden contribuir a cambios en más de una región de interacción contigua que afecta a un Residuo Diana. Por lo tanto, puede haber otros CR y/o ACR y/o DCR con los cuales un CR, ACR o DCR dados pueden agruparse para formar otros Mutantes Candidato. Estos CR, ACR, DCR agrupados o combinaciones de los mismos pueden contribuir a una enzima Mutante Candidato propuesta en la que cada GR, ACR, DCR agrupados o combinación de los mismos se "injerta" sobre el anfitrión, reemplazando los residuos correspondientes en la secuencia de la primera proteína (anfitriona).

(iv) Combinación de Mutantes Candidato para producir un Efecto Acumulativo

30 Si las mutaciones de residuos de dos o más Mutantes Candidato (grupos de uno o más CR y/o ACR y/o DCR) afectan a la misma unidad estructural secundaria que alberga un Residuo Diana y se espera que las mutaciones de cada uno de estos Mutantes Candidato actúen de una manera coordinada en el cambio de la interacción, los Mutantes Candidato pueden combinarse adicionalmente para formar un solo Mutante Candidato combinado. Por ejemplo, si un Residuo Diana es parte de una hélice con un Mutante Candidato interactuando con el extremo N-terminal de la hélice y otro Mutante Candidato interactuando con el extremo C-terminal de la hélice, implicando ambas interacciones la adición de una interacción hidrófoba fuerte en el paso de la secuencia de la primera proteína a la secuencia de la segunda proteína, se podría predecir que un Mutante Candidato combinado con ambos Mutantes Candidato injertados en la secuencia de la primera proteína mostraría efectos concertados en el reposicionamiento de la hélice.

40 (v) Clasificación de mutantes

A pesar de la gran reducción del número de residuos a considerar para la mutación que puede efectuarse mediante la selección inicial de los Residuos Variantes frente a Residuos Diana para producir la lista de Residuos Candidato

(CR) y, opcionalmente, Residuos Candidato Alternativos (ACR) Residuos Candidato Divergentes (DCR), el número de Mutantes Candidato pronosticado que podría resultar de agrupar los CR, los ACR y los DCR aún podría ser grande. En consecuencia, es útil clasificar los Mutantes Candidato basándose en su potencial pronosticado para mostrar mejoras de la propiedad funcional deseada. Los Mutantes Candidato de mayor rango pueden ser entonces la primera opción para evaluaciones computacionales y experimentales adicionales, minimizando así el esfuerzo y el coste. Como se ilustra en el diagrama de flujo 2, en la clasificación se pueden considerar varios principios;

(a) Residuo o residuos Diana afectados por el Mutante Candidato. Diferentes Residuos Diana pueden tener diferentes niveles de participación en la etapa de reacción, y, por lo tanto, importancia relativa. Por ejemplo, un Residuo Diana puede estar directamente implicado en una transferencia de protones, mientras que otro puede ser simplemente un donador/aceptor de enlace de hidrógeno pasivo. Para un Residuo Diana con mayor participación en la etapa de reacción, se pronostica que el efecto de la mutación de un Mutante Candidato relacionado es mayor y, por tanto, el Mutante Candidato correspondiente se clasifica más alto.

(b) Nivel de interacción de los CR y/o los ACR y/o los DCR en el Mutante Candidato. Si el cambio pronosticado en las interacciones asociadas con las mutaciones para un Mutante Candidato implica la formación, o la eliminación, de una fuerte interacción electrostática o hidrófoba, entonces la clasificación será mayor.

(c) Número de cambios de interacción que acompañan a las mutaciones de un Mutante Candidato. Incluso si las mutaciones implican la adición, o la eliminación, únicamente de las interacciones débiles, los cambios totales pueden implicar muchas de tales interacciones dentro del entorno proteico de Residuo Diana, dando lugar a un gran efecto acumulativo neto sobre la estructura y la electrostática, y, por consiguiente, produce un efecto pronunciado sobre la propiedad funcional. Por lo tanto, tales mutaciones pueden ser clasificadas altamente.

#### (vi) Predicciones ampliadas de injerto filogenético

Las predicciones ampliadas de injerto filogenético pueden utilizarse para explotar la capacidad del procedimiento de injerto filogenético para "aprender de la experiencia" a través de la interpretación de los resultados (éxitos y fracasos) de los ciclos de aplicación del método básico de predicción y prueba de Mutantes Candidato para proteínas de interés antes descrito. Si bien el método básico descrito anteriormente se centra en el reconocimiento de Residuos Candidato, Residuos Candidato Alternativos y Residuos Candidato Divergentes y agrupándolos en Mutantes Candidato basados en redes de interacciones de estos residuos, los resultados iniciales para Rubisco detallados en el Ejemplo 3 mostraron que una proporción de estos Mutantes Candidatos eran ineficaces o incluso ligeramente deletéreos para la función de Rubisco, aunque no mucho. La estrategia de injerto filogenético ampliada permitió que este conocimiento acumulado, éxitos y fracasos, fuera re-interpretado y construido en un método adaptado adicionalmente para la aplicación de proteínas en particular, y que pueda continuar refinándose a medida que se obtengan resultados adicionales. La interacción entre la tecnología de injerto filogenético ampliada y el método básico descrito anteriormente se muestra en los diagramas de flujo 1 a 3.

La estrategia ampliada de injerto filogenético tiene tres componentes principales. En primer lugar, el refinamiento del concepto de redes interactivas de Residuos Candidato y/o Residuos Candidato Alternativos y/o Residuos Candidato Divergentes como base del método básico para agruparse en Mutantes Candidato. El examen de los éxitos y fracasos iniciales de los mutantes de la Región 1 de Rubisco, y en particular la comparación del Mutante Núm. 6a de *Synechococcus* y sus Mutantes componente Núm. 4 y Núm. 1a (comentados adicionalmente en los Ejemplos 3 y 4) sugirieron que un marco mejorado para la predicción implicaría reconocer que hay nucleótidos con una frecuencia de mutación muy superior al resto "hotspots" para la adaptación evolutiva que contienen las soluciones evolutivas parciales para Rubisco mejorada, y que la identificación de estas "Subregiones mutables" proporciona una base mejor o complementaria para la identificación de Residuos Candidato, Residuos Candidato Alternativos y Residuos Candidato Divergentes que se deben agrupar preferentemente para formar Mutantes Candidato con el fin de manipular las propiedades funcionales de Rubisco, en lugar de utilizando únicamente las redes de interacción dentro de las Regiones, como en el método básico. Por lo tanto, en el método ampliado, el enfoque de la investigación se centraba más en la identificación de regiones espaciales que en la identificación de redes interactivas de residuos.

En segundo lugar, la reinterpretación de los resultados iniciales de los mutantes de la Región 1 de Rubisco dentro de este marco permitió la identificación de volúmenes espacialmente contiguos de estructura proteica, denominados Subregiones, que contenían un subconjunto de CR de la Región (incluyendo ACR y DCR inicialmente identificados), que se podría predecir que influirían preferentemente en las propiedades de un determinado Residuo Diana o Residuo Diana conectado a la región. La identificación de estos subconjuntos de CR, ACR y DCR de la Región proporciona un medio por el cual se pueden agrupar preferentemente para formar Mutantes Candidato.

En tercer lugar, la identificación de las Subregiones con límites imprecisos y solapantes proporciona una base para la identificación de CR, ACR, CvR y DCR adicionales que se pronostica que interactúan con los subgrupos básicos de CR, ACR y DCR y que pueden ser reclutados a los subconjuntos principales para proporcionar residuos

adicionales para el agrupamiento en Mutantes Candidato.

Por otra parte, la identificación de las Subregiones como nucleótidos con una frecuencia de mutación muy superior al resto de variación de secuencia natural proporciona los medios para identificar y explotar otros tipos de datos de diversidad de secuencias, tales como los Residuos Variantes específicos de especie (SsVR) que varían entre especies estrechamente relacionadas y que pueden representar evoluciones parciales tales como la adaptación a entornos particulares tales como caliente/seco o frío/húmedo. Utilizando Rubisco como ejemplo, la Figura 14 ilustra la variación de secuencia natural entre dos especies de algas rojas y cinco especies de plantas con flores.

Por ejemplo, se identificaron tres Subregiones de la región 1 de la LSU de Rubisco como nucleótidos con una frecuencia de mutación muy superior al resto mutables, y las regiones marcadas 1A, 1B y 1C, como se muestra en las Figuras 15 y 17. Como se detalla en el Ejemplo 4, los autores de la presente invención han usado el método ampliado para reclutar residuos adicionales al conjunto de residuos preferidos de la Región 1 para su mutación y para formar Mutantes Candidato adicionales que se pronosticó que eran mejoras de los mejores mutantes "líder" producidos por la aplicación del método básico, detallado en el Ejemplo 3.

En resumen, el desarrollo del método de injerto filogenético extendido proporciona un medio para mejorar las propiedades funcionales de los mutantes "líder". El desarrollo del concepto de Subregiones como nucleótidos con una frecuencia de mutación muy superior al resto evolutivos y como marco para construir una base de datos de secuencia de proteínas correlacionada con los datos experimentales a partir de la prueba de predicciones, también posiciona la tecnología de injerto filogenético para aprovechar mejor los datos de diversidad de secuencias. Por ejemplo, la base de datos antes mencionada puede ser interrogada contra datos de diversidad de secuencias para identificar residuos que pueden ser reclutados como Residuos Variantes específicos de especie (SsVR) para su posible inclusión en Mutantes Candidato. Por ejemplo, para Rubisco se dispone de datos de secuencias y, en algunos casos, de datos cinéticos para organismos fotosintéticos que crecen en ambientes variados o atípicos, incluyendo especies de plantas C3 relacionadas como las encontradas en las Islas Baleares con diferentes tolerancias a la sequía ya la temperatura (Galmés *et al.*, 2005), la judía de Marama de África meridional adaptada a la sequía (Parry *et al.*, 2007), y algas rojas *Cyanidiales* extremófilas (Ciniglia *et al.*, 2004). Alternativamente para Rubisco, en los casos en que dicha diversidad de secuencias y otros datos funcionales están disponibles para plantas de cultivo tales como trigo (Evans y Austin, 1986), la base de datos antes mencionada puede usarse para identificar especies naturales con propiedades funcionales mejoradas que se pueden usar como germoplasma en la reproducción selectiva.

#### Producción de proteínas

Las proteínas con al menos un Residuo Candidato y/o Residuo Alternativo y/o Residuo Candidato Divergente, y que incluyen opcionalmente otras sustituciones de CvR y/o SsVR, de la segunda proteína pueden ser modeladas *in silico*, o pueden ser diseñadas *in vitro* y/o *in vivo*, por ejemplo mediante mutagénesis dirigida a un sitio de un polinucleótido que codifica la proteína y a continuación expresadas en un sistema de expresión.

#### Detección de proteínas mutantes

Las proteínas Mutantes Candidatas que comprenden la combinación de Residuos Diana con uno o más Residuos Candidato y/o Residuos Candidato Alternativos y/o Residuos Candidato Divergentes y que incluyen opcionalmente otras sustituciones de Residuos Covariantes (CvR) y/o Residuos de Variantes Específicos (SsVR) se escrutan posteriormente para identificar aquellas proteínas Mutantes Candidatas que tienen dicha propiedad funcional mejorada. El procedimiento de escrutinio de las proteínas Mutantes Candidatas puede utilizar una o más de varias técnicas y puede comprender la evaluación catalítica, la evaluación bioquímica y la evaluación fisiológica.

#### Evolución Dirigida

Las proteínas de la invención pueden ser modificadas por medio de evolución dirigida. Por consiguiente, la función de una proteína producida por los métodos de la invención puede mejorarse o modificarse de otro modo. En general, la evolución dirigida puede implicar la mutagenización de una o más moldes moleculares parentales e identificar cualquier molécula deseable entre las moléculas de la progenie. Las moléculas de la progenie pueden ser escrutadas a continuación para detectar la propiedad deseada, evaluando, por ejemplo, la actividad de la molécula, la estabilidad de la molécula y/o la frecuencia de mutación en la molécula. Pueden seleccionarse a continuación moléculas de progenie con propiedades deseables y realizarse más rondas de mutagénesis y escrutinio. Los métodos mediante los cuales se puede llevar a cabo la evolución dirigida son bien conocidos en la técnica. Los métodos ilustrativos incluyen, entre otros, métodos de evolución dirigida racional descritos en la Solicitud de Patente de Estados Unidos Núm. de Serie 10/022.249; y la Solicitud de Patente de Estados Unidos Publicada Núm. US-2004-0132977-A1. Para una descripción general de la metodología experimental y las técnicas implicadas en la

evolución dirigida, se puede hacer referencia a Sambrook et al., *Molecular Cloning, A Laboratory Manual* 2ª ed., Cold Spring Harbor Laboratory Press, 1989.

En una realización de la invención, las proteínas mutantes producidas por los métodos de la invención pueden usarse como un "punto de partida" para la evolución dirigida. Tomando el ejemplo de una proteína Rubisco, las proteínas Mutantes Candidatas que se ha encontrado que tienen una actividad enzimática mejorada pueden seleccionarse y optimizarse adicionalmente mediante evolución dirigida. Esta optimización puede facilitar, por ejemplo, el alivio de los conflictos estéricos por reclutamiento de SvRs y otros residuos variantes de origen natural que pueden ser identificados como complementarios a mutaciones en las proteínas Mutantes Candidatas. Las proteínas Mutantes Candidatas con actividad enzimática mejorada pueden servir así como un nuevo punto de partida para la evolución dirigida, ya que tienen un potencial diferente para explorar el espacio de secuencias en comparación con la proteína Rubisco de tipo salvaje. Los sistemas adecuados para la evolución dirigida de Rubisco incluyen, pero no se limitan a los que emplean la cepa MMI de *E. coli* como se ha informado recientemente (Mueller-Cajar et al., 2007).

#### Proteínas Rubisco

Las proteínas generadas de acuerdo con la descripción incluyen equivalentes funcionales, variantes, fragmentos activos y proteínas de fusión. Para evitar dudas, se incluyen dentro del alcance de la descripción: equivalentes funcionales de los fragmentos activos y proteínas de fusión; fragmentos activos de los equivalentes funcionales y proteínas de fusión; y proteínas de fusión que comprenden un equivalente funcional o fragmento activo.

El término "fragmento" se refiere a una secuencia de ácido nucleico o polipéptido que codifica un constituyente o es un constituyente de la proteína completa. En términos del polipéptido, el fragmento posee actividad biológica cualitativa en común con la proteína completa. Un fragmento biológicamente activo de uso de acuerdo con la presente descripción puede poseer típicamente al menos aproximadamente 50% de la actividad de la proteína completa correspondiente, más típicamente al menos aproximadamente 60% de tal actividad, más típicamente al menos aproximadamente 70% de tal actividad, más típicamente al menos aproximadamente 80% de tal actividad, más típicamente al menos aproximadamente 90% de tal actividad, y más típicamente al menos aproximadamente 95% de tal actividad.

Los métodos para medir la identidad de la secuencia de proteínas son bien conocidos en la técnica y los expertos en la técnica entenderán que en el presente contexto, la identidad de secuencia se calcula basándose en la identidad de aminoácidos (a veces denominada "homología dura"). La identidad de secuencia se calcula después de alinear las secuencias. Los autores de la presente invención han utilizado el programa ClustalW (Thompson *et al.*, 1994) proporcionado en el BioEdit Sequence Alignment Editor (Hall, 1999) para alinear las secuencias. Existen varios paquetes de soporte lógico libres y patentados que realizan alineamientos de secuencias y producen con eficacia los mismos resultados. La identificación de Residuos Variantes también puede realizarse recopilando secuencias de Rubisco, utilizando, por ejemplo, búsquedas BLAST de uno o más grupos filogenéticos que difieren en la propiedad cinética seleccionada para la mejora, y alineándolos contra dicha primera secuencia, utilizando, por ejemplo, CLUSTALW.

Los equivalentes funcionales, fragmentos activos y proteínas de fusión de la descripción conservan la capacidad de la proteína (SEQ ID NO: 23 para *Synechococcus* sp. PCC7942 y SEQ ID NO: 72 para tabaco) para actuar como una enzima Rubisco con una eficacia mejorada. Sin embargo, el experto en la técnica podrá diseñar ensayos o medios para evaluar la actividad enzimática.

Las proteínas funcionalmente equivalentes de acuerdo con la descripción están, por lo tanto, destinadas a incluir mutantes (tales como mutantes que contienen sustituciones, inserciones o deleciones de aminoácidos). Dichos mutantes pueden incluir proteínas en las que uno o más de los residuos de aminoácidos están sustituidos por un residuo de aminoácido conservativo o no conservativo y tales residuos de aminoácidos sustituidos pueden estar codificados o no por el código genético.

Son particularmente preferidas las proteínas en las que varios, es decir, 30 y 50, entre 20 y 30, entre 15 y 20, entre 10 y 15, entre 5 y 10, 1 y 5, 1 y 3, 1 y 2 o sólo 1 aminoácidos son sustituidos, suprimidos o añadidos en cualquier combinación. Las proteínas "mutantes" también incluyen proteínas en las que uno o más de los residuos de aminoácidos incluyen un grupo sustituyente.

Dichos fragmentos pueden estar "libres", es decir, no forman parte o están fusionados con otros aminoácidos o proteínas, o pueden estar comprendidos dentro de una proteína mayor de la que forman parte o región. Cuando está comprendido dentro de una proteína más grande, el fragmento de la descripción forma una única región continua. Adicionalmente, varios fragmentos pueden estar comprendidos dentro de una sola proteína más grande.

Se proporciona en la presente memoria una proteína de fusión que comprende una proteína de la descripción fusionada a un péptido u otra proteína, tal como un marcador, que puede ser, por ejemplo, bioactivo, radioactivo, enzimático o fluorescente, o un anticuerpo.

5 Por ejemplo, a menudo es ventajoso incluir una o más secuencias de aminoácidos adicionales que pueden contener secuencias secretoras o líder, pro-secuencias, secuencias que ayudan en la purificación, o secuencias que confieren una mayor estabilidad proteica, por ejemplo durante la producción recombinante. Alternativamente o adicionalmente, la proteína madura puede fusionarse con otro compuesto, tal como un compuesto para aumentar la semivida de la proteína (por ejemplo, polietilenglicol).

#### Funciones enzimáticas

15 La proteína generada de acuerdo con la descripción es una enzima. Cuando la proteína es una enzima, la función puede ser la catálisis de al menos una reacción química. En otra descripción, la función puede ser estructural (por ejemplo, servir como una proteína del citoesqueleto). La función puede implicar el transporte activo o pasivo de una sustancia dentro de la célula o entre el interior y exterior de la célula, o entre diferentes compartimientos dentro de la célula, o entre diferentes regiones del organismo, por ejemplo cuando la proteína está implicada en un canal o un poro de membrana, o la proteína está implicada en el tráfico de materiales a compartimientos celulares específicos o la proteína actúa como una chaperona o un transportador. La función puede estar implicada en interacciones ligando/receptor, por ejemplo cuando la proteína es un factor de crecimiento, una citocina, un neurotransmisor o un ligando intracelular o extracelular, o la proteína es un receptor para el factor de crecimiento, citocina, neurotransmisor o ligando intracelular o extracelular.

25 Cuando la proteína es una enzima, la enzima puede estar implicada en el catabolismo o el metabolismo. La enzima puede estar implicada en la síntesis de al menos un producto. La enzima puede estar implicada en la degradación de al menos un sustrato. La enzima puede estar implicada en la modificación química de al menos un sustrato, por ejemplo la adición o deleción de uno o más grupos fosfato de una molécula.

30 Las enzimas pueden ser adecuadas para su uso, por ejemplo, en la degradación de plaguicidas y residuos de detergentes, para la extracción de minerales, o para procedimientos químicos "masivos" o finos, tales como amilasas. Las enzimas también pueden ser adecuadas para uso en aplicaciones médicas, y en particular pueden usarse para minimizar cambios en la estabilidad biológica y fisicoquímica.

35 Las enzimas pueden tener propiedades diseñadas específicamente, por ejemplo, la capacidad de actuar óptimamente en un intervalo de temperaturas deseado, una especificidad de sustrato más restringida, más amplia o alterada, o la capacidad para prevenir la producción y/o liberación de subproductos tóxicos o potencialmente tóxicos. La enzima se puede volver a diseñar de tal manera que sea un catalizador eficiente para una reacción menor de la enzima de tipo salvaje utilizando su sustrato natural o un sustrato alternativo para producir un producto diferente.

#### 40 Purificación de las proteínas Rubisco

La descripción proporciona un método para purificar una proteína Rubisco producida de acuerdo con los métodos de la descripción. La holoenzima de la forma funcional de Rubisco a partir de organismos eucariotas (Rubisco de forma I) es un hexadecámero elaborada a partir de 8 subunidades grandes (LSU) y 8 subunidades pequeñas (SSU), y requiere chaperonas apropiadas para plegar y ensamblar correctamente la enzima. *E. coli* es el anfitrión microbiano más utilizado para expresar el ADN y las proteínas recombinantes. Cuando el operón que codifica los genes Rubisco (*rbcLS* y *rbcSS*) de *Synechococcus* sp. PCC7942 se expresa en *E. coli* ambas subunidades se sintetizan abundantemente, sin embargo solamente aproximadamente 1 al 5% de las LSU expresadas se pliegan correctamente y se ensamblan en forma funcional acumulándose la cantidad de Rubisco funcional a ~1 - 3% (peso/peso) de la proteína soluble de *E. coli*.

Para superar estas dificultades, se puede utilizar un sistema recientemente adaptado (Baker *et al.*, 2005) para purificar proteínas Rubisco nativas o mutantes. En este caso, se utilizó el sistema antes mencionado para la purificación de Rubisco *Synechococcus* sp. PCC7942 expresada en *E. coli*. La primera etapa del método de purificación implica la fusión en un primer vector de la secuencia codificante para una secuencia de ubiquitina (Ub) etiquetada con H<sub>6</sub> (H<sub>6</sub>Ub) al extremo 5' de un gen *rbcSS*. A continuación, se co-transforma un anfitrión con el primer vector y el segundo vector que codifica la subunidad grande nativa (o mutada) y la subunidad pequeña de la proteína Rubisco, y a continuación se induce la expresión de la proteína y vectores Rubisco, produciendo los tres péptidos de subunidades de Rubisco (es decir, LSU, SSU y H<sub>6</sub>UbSSU). Algunos se ensamblan en hexadecámeros de Rubisco funcionales compuestos de núcleos octaméricos 8xLSU y diferentes proporciones de SSU (como máximo 8) y H<sub>6</sub>UbSSU. La proteína Rubisco se purifica en base a la expresión de la etiqueta de H<sub>6</sub> fusionada a la

subunidad pequeña de Rubisco. Esta purificación se puede realizar, por ejemplo, utilizando técnicas de cromatografía tales como cromatografía de afinidad de metales. Los fragmentos Ub pueden retirarse a continuación de la Rubisco utilizando, por ejemplo, una proteasa específica de Ub.

- 5 La presente invención se describirá ahora con referencia a ejemplos específicos, que no se debe interpretar que limitan de ningún modo el alcance de la invención.

### Ejemplos

#### 10 Ejemplo 1

##### Identificación de Residuos Diana - Química Computacional

- 15 Un estudio computacional del mecanismo completo de carboxilación de Rubisco y el mecanismo de oxigenación complementaria utilizando los cálculos QM *ab initio* ("desde el principio") en un complejo de fragmentos de sitio activo ampliado proporcionan la base para la estrategia.

Este modelo complejo de fragmentos comprende fragmentos de la mayor parte de los residuos de aminoácidos del sitio activo que se ha establecido o discutido que tienen un papel clave en la serie de reacciones que se catalizan en el sitio activo de Rubisco. Contiene todos los residuos directamente coordinados con  $Mg^{2+}$  o que interactúan con el centro reactivo del sustrato. Este modelo complejo de fragmentos se construyó a partir de las coordenadas de la estructura cristalina con el código PDB 8ruc (estructura cristalina del complejo de Rubisco activado con  $Mg^{2+}$  y 2-carboxiarabinitol 1,5-bisfosfato (2CABP)). Inicialmente, se construyó a partir de las coordenadas cristalográficas el modelo complejo de fragmentos de Rubisco con 2-carboxi-3-cetorabinitol 1,5-bisfosfato (2C3KABP), un análogo estructural de 2CABP y el intermedio de reacción real producido durante la actividad carboxilasa de Rubisco. Como se muestra en la Figura 2, el modelo de fragmento de 77 átomos, denominado FM20, comprende especies de fragmentos moleculares para representar los residuos de aminoácidos LYS175, LYS177, ASP203, GLU204, KCX201 (LYS201 carbamiloado), HIS294 y LYS334 y el fragmento de la forma enodiolato del sustrato RuBP, más el agua y las moléculas de dióxido de carbono. Esta estructura se optimizó utilizando el paquete de química cuántica Gaussian 03 (Frisch *et al.*, 2004). Se generaron geometrías estimadas para todas las otras especies de reacción en la ruta de reacción de la carboxilasa (mostradas en las Figuras 1 y 3) modificando esta geometría optimizada, y a continuación se obtuvieron sus estructuras óptimas de energía, utilizando también el paquete Gaussian 03. En general, se consideraron varias estructuras posibles para cada especie que diferían, por ejemplo, en el patrón de unión H o la orientación de los átomos. En consecuencia, se examinaron las funciones de estos grupos en las etapas de adición de gas y posteriores de las reacciones de carboxilasa y oxigenasa conduciendo a predicciones más seguras, en particular, para la definición de los grupos implicados directamente en la adición de gas ( $CO_2$  o  $O_2$ ) y la energética relativa de las dos etapas.

El análisis se centra enteramente en la subunidad grande (LSU). Aunque la mayoría de las Rubisco, incluyendo plantas verdes, algas y cianobacterias, son proteínas multiméricas (hexadecámeras) complejas que consisten en 8 subunidades grandes (LSU, ~475 residuos) y 8 subunidades pequeñas (SSU, -140 residuos), la química del sitio activo se lleva a cabo mediante una región de proteína que consiste en un dímero de LSU solamente, con 8 de tales sitios de dímeros activos en la proteína hexadecamérica. Un radical del dominio C-terminal (barril TIM") de una LSU contiene la mayor parte de los residuos del sitio activo mientras que una región más pequeña del dominio N-terminal de la LSU adyacente completa el sitio activo. Sin embargo, las predicciones que surgen más generalmente de los estudios de bioinformática sugieren que pueden estar implicadas otras regiones en la modulación de la química, p.ej., contactos entre subunidades (LSU-LSU o LSU-SSU).

La Figura 1 muestra un esquema del mecanismo de reacción de la carboxilasa propuesto basado en los cálculos QM. Las flechas curvas indican el flujo de electrones, lo que significa la formación de enlaces y los eventos de ruptura de enlaces, que conducen a las sucesivas especies de reacción. El esquema define la participación de residuos de aminoácidos en cada etapa de reacción, y la etapa de adición de gas (carboxilación), en particular. Los residuos que se pronostica que tengan un papel en la etapa de adición de gas comprenden el conjunto de Residuos Diana definidos previamente. Las Figuras 3 y 4 proporcionan geometrías y energías relativas, respectivamente, de la especie de reacción de carboxilasa.

Las características más significativas del mecanismo de reacción propuesto se discuten a continuación. En primer lugar los autores de la presente invención han realizado el sorprendente descubrimiento de que  $H_2O[Mg]$  no es desplazado de la coordinación de Mg por  $CO_2$  durante la carboxilación. La molécula de agua ayuda de hecho a unir el  $CO_2$  al sitio activo y contribuye a la estabilidad del producto carboxilado y el TS correspondiente. La misma molécula de agua actúa como agua de hidratación en la etapa posterior. Anteriormente, este papel de hidratación se había asignado a una molécula de agua encontrada en las intermediaciones de la esfera de coordinación.

Los autores de la presente invención han realizado el descubrimiento aún más sorprendente de que el átomo O2 permanece no protonado en el intermedio enodiolato, a pesar de las expectativas de los principios químicos generales de que tendría que desprotonarse para dirigir la carboxilación exclusivamente a C2, en lugar de a C3. Las cargas atómicas derivadas de ESP también muestran que O3 es más negativo que O2. Sin embargo, este resultado inesperado se explica por la observación de puentes de hidrógeno fuertes entre LYS175 y KCX201 protonado con O2, lo que impide de hecho que O2 dirija la carboxilación a C3. Además, puesto que LYS334 está unido mediante H al grupo P1-fosfato en la enzima, su interacción con el CO<sub>2</sub> sustrato se interrumpiría si tuviera lugar la carboxilación de C3, sin dejar margen para la estabilización del TS correspondiente y el intermedio.

Otras características del mecanismo de reacción elucidado por los autores de la presente invención son las siguientes. KCX201 tiene un papel directo sólo en la reacción de enolización inicial y permanece en un estado protonado después de la enolización. KCX201 y LYS 175 tienen un papel en la obstaculización de la carboxilación de C3 neutralizando parcialmente la carga negativa en O2. HIS294 tiene un papel importante en la catálisis multietapa de Rubisco. Traslada el protón entre N<sub>ε</sub> y O3, modulando apropiadamente la longitud del enlace C3-O3. GLU204 activa la molécula de agua coordinada con Mg para la hidratación mediante la extracción de su protón. Así, tanto la carboxilación como la hidratación tienen lugar sobre la misma superficie del intermedio enodiolato. El protón H3 se transfiere finalmente a O2 sólo después de la formación del *aci*-ácido intermedio (VII). La carga en el intermedio *aci*-ácido está estabilizado por LYS 175 y LYS334. LYS175 asegura la protonación estereoespecífica del carbono C2 para producir los productos finales. LYS334 comparte su protón con LYS175.

Basándose en las conclusiones anteriores, los autores de la presente invención han identificado dos residuos de aminoácidos, uno que actúa como una base, el otro como un ácido, para cada etapa de reacción. Para la etapa de adición de gas, HIS294 actúa como una base sustrayendo el protón de O3, mientras que LYS334 es el ácido que dona un protón para estabilizar el grupo carboxilato formado por la adición de CO<sub>2</sub>. La alteración en el entorno estérico o electrónico de estos dos residuos clave, o cualquier otro residuo que sea estructural o químicamente (a través de interacciones electrostáticas) relacionadas con ellos, afectaría a la especificidad de la enzima y probablemente también afectaría a  $k_{cat}$ .

Los residuos TYR20 GLU60, ASN123 también se identificaron como cruciales para la orientación apropiada de las moléculas de gas con respecto al sustrato antes de la adición, y para la estabilidad del aducto de gas y las estructuras de estado de transición correspondientes. Estos cinco residuos comprenden un grupo inicial de Residuos Diana para un examen adicional.

## Ejemplo 2

Análisis filogenético para identificar Residuos Variantes que contengan residuos que determinan la especificidad

Las secuencias LSU de Rubisco de phyla disponibles de los organismos fotosintéticos se recogieron para el análisis filogenético de bases de datos públicamente disponibles en NCBI ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) y JGI (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>) mediante la realización de búsquedas de proteínas BLAST (Altschul *et al.*, 1997) utilizando la secuencia LSU de Rubisco de espinaca como secuencia de consulta. Puesto que las secuencias LSU son tan distintivas, y la conservación relativamente alta en comparación con la mayoría de las clases de homólogos de proteínas a lo largo de tales amplias distancias evolutivas, son fáciles de identificar y funcionaría igualmente bien otro soporte lógico de búsqueda de uso libre o patentado.

El alineamiento de las secuencias extraídas de Rubisco LSU de organismos fotosintéticos pertenecientes a trece diferentes phyla que abarcan algas rojas, cianobacterias, glaucophyta y plantas (10 phyla) se llevó a cabo utilizando el soporte lógico ClustalW (Thompson *et al.*, 1994) dentro de BioEdit Sequence Alignment Editor (Hall, 1999) para evaluar su diversidad. Un alineamiento se muestra en la Figura 5; en los casos en que hay más de una secuencia de un solo phylum, se ha utilizado una secuencia consenso de 50% para la eficacia. Las secuencias consenso se obtuvieron utilizando el servidor en <http://coot.embl.de/Alignment//consensus.html>. Se podría utilizar cualquier otra herramienta de alineamiento patentada o de uso libre para el mismo propósito debido al hecho de que la conservación global de estas Rubisco es tan alta que se obtendrían de hecho los mismos resultados de alineamiento.

La Figura 5 muestra la diversidad de las secuencias LSD de Rubisco de filogenias distantes. Así, aunque la información sobre soluciones evolutivas parciales para la adaptación de las Rubisco a ambientes específicos está integrada en esta diversidad, es imposible analizar esta información directamente para identificar los residuos responsables de adaptaciones específicas, tales como el aumento de la especificidad.

Para resolver este problema, los autores de la presente invención han desarrollado el método de injerto filogenético

basado en la hipótesis y lo han aplicado a la identificación de residuos de Rubisco que ya representan soluciones evolutivas parciales naturales para la especificidad mejorada ( $S_{C/O}$ ). Es bien sabido que diversos organismos fotosintéticos exhiben valores característicamente diferentes para la  $S_{C/O}$ . La mayoría de las plantas terrestres poseen un valor  $S_{C/O}$  de 80, mientras que se sabe que las algas rojas tienen el factor de especificidad más alto (~160). El factor de especificidad para las cianobacterias tiene un modesto valor de aproximadamente 40.

Puesto que las cianobacterias son un ancestro común tanto de las plantas terrestres como de las algas rojas, y puesto que las plantas terrestres divergieron de las cianobacterias antes en la evolución en comparación con las algas rojas ([http://www.geocities.com/we\\_evolve/Plants/chloroplast.html](http://www.geocities.com/we_evolve/Plants/chloroplast.html)) la solución evolutiva parcial para mayor especificidad integrada en los cambios de residuos de aminoácidos en algas rojas se puede revelar parcialmente mediante la comparación de las secuencias LSU de Rubisco de estos tres grupos. La Figura 6 muestra el alineamiento de secuencias consenso de 50% de las LSU de Rubisco de algas rojas, cianobacterias y plantas con flores (color verde) mostradas anteriormente en la Figura 5, pero ahora se muestra con sombreado de color gris para destacar los 134 residuos en algas rojas que difieren de los comunes en cianobacterias y plantas con flores, es decir, los Residuos Variantes. Puesto que hay un acuerdo general de que los cambios de un solo residuo no darán lugar a una mejora sustancial en la especificidad de Rubisco (de lo contrario la evolución habría sido capaz de explorar el espacio de secuencia de manera razonablemente fácil para optimizar la especificidad en determinados organismos), el número de mutaciones múltiples, 2-10 residuos, que tendrían que ser sometidas a ensayo para explorar incluso el espacio de secuencia reducido de estos 134 residuos para identificar los residuos que determinan la especificidad es increíblemente grande. Al hacer uso de esta lista de Residuos Variantes, las variaciones que no son relevantes para la propiedad enzimática de interés se filtran (es decir, especificidad).

### Ejemplo 3

Identificación, Agrupamiento y Clasificación de Residuos Candidato, y Predicción de Mutantes Candidato utilizando el Método de Injerto Filogenético Básico

Con el fin de identificar los residuos determinantes de especificidad que explican el aumento de especificidad de las Rubisco de algas rojas, los autores de la presente invención utilizaron los conocimientos enzimáticos-mecánicos de los cálculos QM descritos en el Ejemplo 1 para desarrollar un procedimiento para seleccionar un subconjunto de los 134 Residuos Variantes identificados en el Ejemplo 2. Los Residuos Variantes, identificados en el Ejemplo 2, se seleccionaron frente a los Residuos Diana, identificados en el Ejemplo 1 para tener un papel funcional en la etapa de adición de gas, con el fin de identificar los Residuos Candidato. Algunos de los Residuos Candidato seleccionados también fueron Residuos Candidato Alternativos. Además, también se seleccionaron algunos Residuos Candidato Divergentes. Se utilizó un procedimiento basado en el principio general de que los cambios evolutivos de residuos que modificaban las propiedades de estos Residuos Diana alteraban su especificidad y eficiencia cinética. Por lo tanto, cualquier Residuo Variante que tenga el potencial de afectar los Residuos Diana electrónicamente o estructuralmente, y, por tanto, provocar un cambio en la propiedad funcional de la enzima, en este caso la especificidad y la eficiencia cinética, se considera parte de la solución evolutiva parcial a la optimización de la propiedad.

El procedimiento de selección se aplicó para identificar tales Residuos Candidato, Residuos Candidato Alternativos y Residuos Candidato Divergentes. La selección se llevó a cabo mediante análisis visual y comparación de las diferencias en las interacciones entre residuos con Accelrys Discovery Studio v1.5.1 (Accelrys Software Inc., San Diego, CA, 2005), utilizando las estructuras cristalinas de las Rubisco de espinaca (código PDB: 8ruc) y *Galdieria partita* (Código PDB: 1BWV). La estructura cristalina de la Rubisco de espinaca se utilizó en lugar de la Rubisco de *Synechococcus* sp. PCC6301 (código PDB: 1RB1) debido a que la resolución de las estructuras disponibles es superior para la estructura Rubisco de espinaca. Sin embargo, los autores de la presente invención explicaron los cambios de residuos entre las LSU de espinaca y las Rubisco de *Synechococcus* sp. PCC6301 al analizar las interacciones inter-residuos, y utilizaron la estructura de *Synechococcus* también como referencia.

Tres de los cinco Residuos Diana, TYR20, GLU60 y ASN123 están en el dominio N-terminal de la LSU de Rubisco, mientras que HIS294 y LYS334 están en el dominio C-terminal que forma la estructura de barril TIM (Figura 8). Con el propósito de clasificar grupos de CR, ACR y DCR que pueden formar subconjuntos de residuos de las soluciones evolutivas parciales, los Residuos Diana fueron clasificados como pertenecientes a tres Regiones diferentes. Los residuos TYR20, GLU60 y ASN123 interactúan colectivamente con el grupo carboxilato del compuesto intermedio formado después de la adición de  $CO_2$  (Figuras 8, 9 y 15). Las unidades estructurales secundarias que contienen estos tres residuos también se entrelazan con otras hélices y hojas del dominio N-terminal (Figuras 8, 9 y 15). Como resulta evidente en la Figura 15, ASN123, GLU60 y TYR20 están unidos a tres regiones de estructura secundaria separadas del dominio N-terminal y sus cadenas laterales se extienden hacia el sitio activo en una constelación tripartita. Por lo tanto, estos tres Residuos Diana se clasifican como residuos de la región 1. HIS294 está en la hebra  $\beta_4$ , que es una parte del barril TIM; la región que rodea a HIS294 se define como la Región 2. Aunque LYS334 está

en la vecindad de HIS294 y de residuos de la Región 1, está en el 'bucle 6' de la LSU (Figura 8) y en la forma abierta de la LSU está separado de los otros Residuos Diana; por lo tanto, la región que contiene LYS334 se clasifica como Región 3.

5 El análisis inicial y el diseño del Mutante Candidato se ilustran tomando en consideración los Residuos Variantes en la Región 1. Puesto que el dominio N-terminal de la LSU de Rubisco es un dominio compacto (Figuras 9 y 15), las consideraciones proteico-estructurales sugieren que las variaciones de los residuos de aminoácidos en la mayor parte del dominio pueden influir en los Residuos Diana TYR20, GLU60 y ASN123. El procedimiento de selección se inició trazando todas las interacciones electrostáticas e hidrófobas entre cada uno de los Residuos Variantes y todos los demás residuos en la LSU en ambas estructuras cristalinas. Se ignoraron los residuos que tienen interacciones equivalentes en ambas estructuras cristalinas. Las interacciones equivalentes comprenden principalmente interacciones que implican un átomo de cadena principal de un Residuo Variante, y tales interacciones se producen en gran parte dentro de una hélice o entre dos hebras  $\beta$ . Las interacciones también se consideraron equivalentes si los Residuos Variantes en espinaca (o *Synechococcus* sp. PCC6301) y *Galdieria partita* difieren sólo en la longitud de la cadena lateral y los átomos que interactúan son del mismo tipo químico y están aproximadamente a la misma distancia en ambas estructuras cristalinas. De forma similar, las interacciones que implican valina, leucina e isoleucina se consideraron equivalentes si sólo un grupo metilo de la cadena lateral está implicado en la interacción y la distancia de interacción no difiere significativamente entre las dos estructuras cristalinas. Este procedimiento eliminó algunos de los Residuos Variantes de una consideración adicional como posibles Residuos Candidato (o Residuos Candidato Divergentes). El resto de las interacciones que implican Residuos Variantes se analizaron a continuación para determinar su potencial para afectar a los Residuos Diana.

Un Residuo Variante cuya cadena lateral interactúa con la cadena principal o las cadenas laterales de los Residuos Diana o con residuos que forman parte de las unidades estructurales secundarias que albergan los Residuos Diana se consideró un Residuo Candidato (o Residuo Candidato Divergente). Como ejemplo, el Residuo Diana GLU60 está en el extremo C-terminal de la hélice  $\alpha$ B y el Residuo Variante ILE51, en el extremo N-terminal de esta hélice, interactúa con otro Residuo Variante TRP25 en la LSU de *Galdieria partita* (véanse la Figura 14 y la Tabla 2), pero esta interacción está ausente en las LSU de espinaca y *Synechococcus* sp. PCC6301. Se pronostica que el "injerto" de estos Residuos Variantes de *Galdieria partita* en *Synechococcus* sp. PCC6301 tiene un efecto sobre la orientación de la hélice  $\alpha$ B y, por tanto, sobre GLU60. Ambos Residuos Variantes 51 y 25 se consideraron, por lo tanto, como Residuos Candidato. Las posiciones de estos residuos en la estructura de LSU se muestran en la Figura 9. En este caso, es útil agrupar estos dos Residuos Candidato que comparten una región de interacción contigua en una única mutación, ya que es probable que formen una parte coordinada de la solución evolutiva parcial. Este agrupamiento más simple es numerado Mutante Candidato Núm. 1a en la Tabla 3. Tales Residuos Candidato agrupados representan mutaciones para la sustitución de residuos (injerto) de la secuencia de *Galdieria partita* sobre la secuencia de *Synechococcus* sp. PCC6301, con el fin de transferir características de mayor especificidad a Rubisco de *Synechococcus* sp. PCC6301.

Antes del injerto, los grupos de Residuos Candidato se pueden combinar adicionalmente en grupos que pueden actuar adicionalmente en la coordinación y amplificación del efecto perturbativo sobre un determinado Residuo Diana o Residuos Diana. Tal agrupamiento ampliado es útil si dos grupos diferentes de Residuos Candidato afectan a la misma unidad estructural secundaria. Por ejemplo, ambos grupos de Residuos Candidato {25, 51} y {54, 84, 87} afectan a la hélice  $\alpha$ B, que alberga GLU60, a través de dos interacciones diferentes. Por lo tanto, uno de los mutantes pronosticados comprendía estos grupos combinados (Mostrado como Mutante Núm. 7a en la Tabla 3).

Los potenciales Mutantes Candidato injertados se evaluaron adicionalmente para comprobar nuevas interacciones estéricas desfavorables introducidas injertando residuos de *Galdieria partita* en *Synechococcus* sp. PCC6301. Tales interacciones estéricas no deseables pueden ser rectificadas añadiendo mutaciones espacialmente complementarias al Mutante Candidato, o podrían ser investigadas mediante simulaciones MD para evaluar si la relajación estructural para aliviar tales contactos negativos es energéticamente accesible.

La etapa final en el procedimiento de predicción del Mutante Candidato es clasificar los potenciales mutantes injertados para desarrollar una lista clasificada para su uso en la priorización de ensayos experimentales o pre-escrutinio *computacional in silico* detallado. La clasificación refleja el grado esperado en el que se espera que las mutaciones combinadas en los Mutantes Candidato individuales cambien la propiedad funcional, en este ejemplo en la dirección hacia la mejora de la especificidad influyendo en la química y la química relativa, de las etapas de adición de gas para CO<sub>2</sub> y O<sub>2</sub>. La clasificación depende de una serie de parámetros tales como el Residuo Diana afectado por el grupo de Residuo Candidato, la fuerza de las interacciones cambiadas del grupo de Residuo Candidato dentro de la Región de Residuo Diana y el número de dichas interacciones para cada grupo de Residuo Candidato.

Además de la secuencia Rubisco de *Galdieria partita*, se consideró en el análisis la secuencia de otras especies de

algas rojas *Griffithsia monolis*, que se sabe que tiene un mejor  $k_{cat}^c$  que *Galdieria partita*. Para el conjunto inicial de Residuos Candidato seleccionados, dos residuos (51 y 54) muestran diferencias entre *G. partita* y *G. monolis* es decir, son Residuos Candidato Alternativos. Se formaron Mutantes Candidato con las dos variantes de residuos en las posiciones 51 y 54, como se muestra en la Tabla 3 para los CM Núm. 1, Núm. 5, Núm. 6, Núm. 7 y Núm. 13.

5 Como ejemplo, se pronosticó el conjunto de dieciséis grupos de Residuos Candidato mostrados en la Tabla 3 a partir del análisis inicial de los Residuos Diana de la Región 1, formando 21 Mutantes Candidato potenciales (con variantes de *G. Partita* y *G. monolis*). Para cada Mutante Candidato, la última columna de la Tabla 3 detalla el cambio estructural pronosticado asociado con las mutaciones. Para algunos Mutantes Candidato, estos cambios se explican gráficamente en una figura; la segunda columna de la tabla proporciona el número de la figura para estos CM. La Tabla 3 también muestra las clasificaciones de prioridades para la prueba experimental.

10 La Tabla 3 muestra que en este análisis inicial se seleccionaron dos Residuos Candidato Divergentes (36, 116 y 140, véase la Tabla 2) por las razones resumidas para los Mutantes Candidato relevantes (Núm. 9 y Núm. 10). Dos de estos DCR (36 y 140) también muestran diferencias entre *G. Partita* y *G. Monolis* (véase la Tabla 2); para los Mutantes Candidato Núm. 9 y Núm. 10, la variante Gm de 36 y la variante Gp de 140 se consideraron los más prometedores para el trasplante a *Synechococcus*.

15 **Tabla 3.** Mutantes Candidato pronosticados y clasificados a partir del Análisis de la región 1 en el Dominio N-terminal de la LSU de Rubisco que rodea a los Residuos Diana TYR20, GLU60 y ASN123. Los superíndices Gp y Gm denotan el residuo de *Galdieria partita* y *Griffithsia monolis*, respectivamente.

Núm.	Fig <sup>a</sup>	Mutante	Rango	Región afectada
1a	9, 10	Y25W/D511 <sup>Gp</sup>	3	Añade una nueva interacción hidrófoba entre el extremo C-terminal de αB y βA. βA está cerca de Y20 en la secuencia.
1b	9, 10	Y25W/D51V <sup>Gm</sup>	3	Añade una nueva interacción hidrófoba entre el extremo C-terminal de αB y βA. βA está cerca de Y20 en la secuencia.
2	-	A59G/G64A	6	Mutación por intercambio "Swapping mutation". Interacción entre αB y la cadena larga que conecta αB a βC. El grupo metilado intercambiado está espacialmente cerca de Y20 y adyacente a E60.
3	-	P49D/D51I	10	Altera la interacción en el lazo corto que conecta αB y βB.
4	9, 10	T23G/K81R	4	La interacción entre βA y βC se rompe. Afecta el posicionamiento de Y20.
5a	10	G54A <sup>Gp</sup> / C84A/I87V	5	Introduce una fuerte interacción hidrófoba entre αB y βC.
5b	10	G54S <sup>Gm</sup> / C84A/I87V	5	Introduce una fuerte interacción hidrófoba entre αB y βC.
6a	9	T23G/ Y25W/ D51I <sup>Gp</sup> / K81R	1	El Mutante Núm. 1a añade una interacción hidrófoba entre βA y αB, mientras que el Mutante Núm. 4 rompe la interacción de βA con βC. Estos dos conjuntos de mutaciones en conjunto pueden tener un gran efecto sobre Y20.
6b	9	T23G/ Y25W/ D51V <sup>Gm</sup> / K81R	1	El Mutante Núm. 1b añade una interacción hidrófoba entre βA y αB, mientras que el Mutante Núm. 4 rompe la interacción de βA con βC. Estos dos conjuntos de mutaciones en conjunto pueden tener un gran efecto sobre Y20.
7a	10	Y25W/ D51I <sup>Gp</sup> / G54A <sup>Gp</sup> / C84A/ I87V	2	Efecto acumulativo de los Mutantes Núm. 1a y Núm. 5a en E60.
7b	10	Y25W/ D51I <sup>Gp</sup> / G54S <sup>Gm</sup> / C84A/ I87V	2	Efecto acumulativo de los Mutantes Núm. 1a y Núm. 5b en E60.
7c	10	Y25W/ D51V <sup>Gm</sup> / G54A <sup>Gp</sup> / C84A/I87V	2	Efecto acumulativo de los Mutantes Núm. 1b y Núm. 5a en E60.
7d	10	Y25W/ D51V <sup>Gm</sup> / G54S <sup>Gm</sup> / C84A/ I87V	2	Efecto acumulativo de los Mutantes Núm. 1b y Núm. 5b en E60.
8	11	V121I/ M297G/ V300T	7	La interacción hidrófoba entre las dos LSU se rompe (los residuos 297 y 300 son de la LSU vecina que contiene el complejo de Mg

## ES 2 640 747 T3

Núm.	Fig <sup>a</sup>	Mutante	Rango	Región afectada
				del sitio activo que se considera). Afecta a N123.
9	11	L36I <sup>Gm</sup> / I116L/ F140L <sup>Gp</sup>	8	Forma una gran región hidrófoba que implica los extremos de dos cadenas $\beta$ adyacentes ( $\beta$ B y $\beta$ E) y $\alpha$ C. Podría alterar la orientación/posicionamiento de N123.
10	11	L36I <sup>Gm</sup> / I116L/ Y121I/ F140L <sup>Gp</sup> / M297G/ V300T	3	Podrían actuar juntas sobre $\alpha$ -C y tener un efecto acumulativo sobre N123.
11		T114A/ T118A/ T271V/ V121I	2	Se rompe la interacción polar de 114 y 118 con 271 en la LSU acompañante. 271 forma una nueva interacción hidrófoba con 121. Tiene efecto sobre N123
12	17	K18I/ T23G	5	La interacción polar entre las cadenas laterales de T23 y K18 se rompe. Tiene efecto sobre Y20.
13a	9, 10	Y25W/ D51I <sup>Gp</sup> / inserto A21	6	(1a + A21) La inserción de A21 separa K21 del residuo 51 y forma una nueva interacción hidrófoba con I51.
14	17	KLTTY-(21-25)- AKMGYW	4	La forma de una bobina adyacente a Y20 está alterada, afecta a la orientación de Y20. Implica una inserción (M); véase la Figura 6.
15	17	K18I/ KLTTY-(21-25)- AKMGYW/ K81R	1	El cambio de forma de la bobina junto a Y20 se asocia con cambios en su interacción con 18 (pérdida de interacción polar) y 81 (cambio de longitud) y puede tener un efecto coordinado sobre Y20.
16	-	A15S/ K18I/ T68V/ L407I	9	Posible interacción entre los residuos 18 y 68 que podrían unir la cola N-terminal al resto del dominio; el residuo conservado 69 interactúa con 407 de la LSU asociada; S15 puede formar enlaces de H fuertes con grupos carbonilo de la cadena principal de 408 y 409 de LSU asociado en algas rojas. Elige como diana Y20.

<sup>a</sup> Las figuras 9-11 y 17 muestran los sitios de mutación pronosticados.

### Ejemplo 4

5 Identificación, agrupamiento y clasificación de residuos candidato, y predicción de mutantes candidato utilizando el método de injerto filogenético ampliado

10 Como se ha mencionado anteriormente y como se muestra en los Diagramas de flujo 1 y 3, el método de injerto filogenético ampliado proporciona un medio para utilizar eficazmente el conocimiento del "mapa" de función para las mutaciones, desarrollado a partir de la experiencia adquirida a partir de ciclos de aplicación del método básico de predicción y ensayo de Mutantes Candidato. El uso del método ampliado se ilustra en este ejemplo con referencia a los resultados iniciales para los Mutantes Candidato de la Región 1 pronosticados (Ejemplo 3) proporcionados en la Tabla 5 y comentados con más detalle en el Ejemplo 7.

15 El método ampliado se utilizó por primera vez para interpretar los resultados del método básico. Esto dio como resultado la identificación de un nuevo modelo de agrupamiento de CR, ACR y DCR ya identificados (véanse los CM en la Tabla 3), que se basó en Subregiones como nucleótidos con una frecuencia de mutación muy superior al resto mutables en lugar de centrado en las redes de interacciones como en el método básico. Las tres Subregiones (1A, 1B y 1C) exhibieron la propiedad de que se podía pronosticar que influían preferentemente en las propiedades de uno de los tres Residuos Diana vinculados a la Región 1: Subregión 1A a TYR20, 1B a ASN123 y 1C a GLU60. Este "anclaje" de los tres TR a las Subregiones de la Región 1 se muestra gráficamente en la Figura 15. Las Subregiones se muestran gráficamente en una orientación alternativa en la Figura 17. Las Subregiones están confinadas enteramente al dominio N-terminal, excepto para los fragmentos del dominio C-terminal de la LSU adyacente que forman parte de la Subregión 1B.

25 La identificación de las Subregiones proporcionó la base para la identificación de CR y DCR adicionales, utilizando la inspección visual y otros análisis descritos anteriormente para identificar CR y DCR utilizando el método básico, que se puede agrupar preferentemente con los CR y DCR ya identificados por el método básico para formar nuevos Mutantes Candidato (véase la Tabla 4). Los CR y los DCR adicionales así identificados fueron las posiciones 19, 68, 88 y 104, y 86, 117 y 138, respectivamente. Alternativamente, la identificación de las Subregiones proporciona una nueva base para el reagrupamiento preferencial de los CR y DCR ya identificados por el método básico para formar

30

nuevos Mutantes Candidato.

A continuación se ilustra el uso del método ampliado y la base para reclutar los CR y DCR adicionales particulares con referencia a ejemplos para el refinamiento de la actividad de los mutantes pronosticados por el método básico y sometidos a ensayo inicialmente (Tabla 5). Los nuevos CM pronosticados se muestran en la Tabla 4.

**Tabla 4.** Los Mutantes Candidato pronosticaron en uso del método de injerto filogenético ampliado para refinar los Mutantes Candidato inicialmente más prometedores (Tabla 3). También se enumeran las mutaciones de componente de residuo único y doble adicionales de estos Mutantes Candidato pronosticados inicialmente más prometedores.

Núm.	Fig.	Mutación	Rango	Comentarios
17-1A	17	T23G/K18I/T68V/K81R	2	Predicción basada en un injerto filogenético refinado en la Subregión 1A.
18a-1A	17	T23G/K81R/P104E <sup>Gp</sup>	5	
18b-1A	17	T23G/K81R/P104D <sup>Gm</sup>	5	
19-1A	17	T23G/D19P/K81R	4	
20-1B	16, 17	I116L/L117T/V121I/I138M/F140L	3	Predicción basada en un injerto filogenético refinado en la Subregión 1B. Elige como diana el residuo N123.
21-1A,B	17	T23G/K81R/V121I/M297G/V300T	1	Combinación de dos CM pronosticados que mostraron una mejora global máxima en el estudio de evaluación cinética de la eficacia enzimática.
22-1A	9,10	T23G		Componentes de residuo único del Mutante Núm. 4.
23-1A	9,10	K81R		
24	11	V121I/M297G		Componente de residuo doble del Mutante Núm. 8.
25	11	M297G		Componentes de residuo único del Mutante Núm. 8.
26-1 B	11	V121I		

El primer ejemplo a continuación se centra en el refinamiento del mejor mutante, Mutante Núm. 4 (T23G/K81R), de la primera ronda de pruebas, y en la Subregión 1A. Como se muestra en la Tabla 3, el Mutante Núm. 4 es un componente con el Mutante Núm. 1a (Y25W/D51I<sup>Gp</sup>) de Mutante Núm. 6a (T23G/ Y25W/ D51I<sup>Gp</sup>/ K81R). Sin embargo, aunque el Mutante Núm. 6a mostró un aumento de especificidad (8,5%, véase la Tabla 5), no mostró ninguna mejora de la eficacia, y en general es inferior al Mutante Núm. 4, que mostró tanto un aumento de especificidad (10%) como de eficacia (8%). La referencia a los resultados para el Mutante Núm. 1a indicó que es un mutante relativamente malo con un aumento de especificidad de sólo 3,4% y ningún cambio en la eficacia. Por lo tanto, se concluyó que el agrupamiento de las mutaciones del Mutante Núm. 4 con las del Mutante Núm. 1a era ineficaz, ya que el efecto neto no era una mejora acumulativa, sino que los dos grupos de mutaciones aparentemente interfirieron desventajosamente. Puesto que la Subregión 1A contiene las posiciones de residuo de las mutaciones en el Mutante Núm. 4, se dedujo que el acoplamiento de las mutaciones del Mutante Núm. 4 con mutaciones de otros CR y/o DCR en la Subregión 1A, puede proporcionar una mejor estrategia para mejorar adicionalmente las propiedades de este mutante "líder".

El examen de la secuencia y de los datos estructurales mediante procedimientos de análisis visual y de otro tipo ya descritos, especialmente un análisis de datos de covarianza específicos de especies para los CR y los DCR ya identificados en la Subregión 1A, identificó tres CR adicionales, 19, 68 y 104, que pueden influir en TR TYR20 o el residuo 81 (componente del Mutante Núm. 4). Por lo tanto, los siguientes nuevos CM, que mostraban los cambios indicados en las interacciones, fueron pronosticados como posibles mejoras en el Mutante Núm. 4.

En relación con el Mutante Núm. 17-1A (Núm. 4 + K18I/T68V), se pronostica que las mutaciones de K18I y T68V dan lugar a una interacción alterada entre la cola N-terminal y la región entre  $\beta$ B y la hélice B. Combinadas con las mutaciones en el Mutante Núm. 4 (T23G y K81R) se espera que tenga un mayor impacto en TR TYR20.

En relación con el Mutante Núm. 18a(b)-1A (Núm. 4 + P104E<sup>Gp</sup>(D<sup>Gm</sup>)), la mutación de CR 104 (que es también un ACR), es decir, P104E<sup>Gp</sup> o P104D<sup>Gm</sup>, se espera que de como resultado una nueva interacción con N de la cadena principal del residuo 81. Se espera que las tres mutaciones en el CM Núm. 18-1A combinado tenga un mayor impacto en TR TYR20.

En relación con el Mutante Núm. 19-1A (Núm. 4 + D19P), el residuo 19 forma un puente salino con el residuo 21 (K

o R) en cianobacterias y plantas, que está ausente en algas rojas donde el residuo 19 es P. Se espera que la mutación D19P influirá en el alcance de TR TYR20 en el sitio activo, y que en combinación con las mutaciones del Mutante Núm. 4, esto produciría un mayor impacto en TR TYR20.

5 Estas predicciones se muestran gráficamente en las Figuras 15 y 17.

El segundo ejemplo a continuación se refiere a la identificación de una nueva zona mutacional importante (Subregión 1B) que a su vez condujo a la predicción inicial del Mutante Núm. 20-1B (V116L/L117T/V121I/I138M/F140L). Este mutante tiene el Mutante Núm. 8 (V121I/M297G/V300T), identificado por el método básico (Tabla 3), como un componente. Como se muestra en la Tabla 5, la prueba inicial del Mutante Núm. 8  
10 mostró una mejora en la especificidad (5,6%) que proporciona una base para una caracterización adicional. Tras la identificación de la Subregión 1B, tal como se ha descrito, el examen de las variaciones de los residuos cerca del residuo 121 (CR adyacente a TR N123, véase la figura 16A) condujo al descubrimiento de una red de interacciones entre la hélice C y la hoja  $\beta$  E. Esta red puede observarse más claramente en la Figura 16B. Los cambios en esta  
15 red de interacción pueden afectar la orientación de TR N123. Las interacciones implican los residuos 116, 117, 120, 121, 124, 133, 135, 138 y 140. De éstos, el residuo 121 es un CR y los residuos 120, 124, 133 y 135 se conservan en todos los grupos principales de organismos (véase la Figura 5). Los residuos 116, 117, 138 y 140 son DCR, es decir, difieren entre los tres grupos principales de organismos, algas rojas, cianobacterias y plantas con flores (Tabla 2). Se observó que los cambios en los CR y DCR que interactúan en esta región espacial (es decir, 21, y 116, 117,  
20 138 y 140) parecen estar compensándose en los principales grupos de organismos. Por lo tanto, como una prueba inicial del efecto de tales cambios en la orientación de N123, se pronosticó el Mutante Núm. 20-1B (V116L/L117T/V121I/I138M/F140L) (Tabla 4).

#### Ejemplo 5

25 Generación de Rubiscos mutantes y su escrutinio

Basándose en los Mutantes Candidato pronosticados en las Tablas 3 y 4, se realizaron estudios de prueba de principio mutando los residuos específicos de algas rojas en la Región 1 de *Synechococcus* sp. PCC7942, en grupos  
30 de mutaciones múltiples. Inicialmente catorce de los mutantes LSU pronosticados detallados en el Ejemplo 3 y enumerados en la Tabla 3 (los Mutantes Núm. 1a, Núm. 1b, Núm. 4, Núm. 5a, Núm. 5b, Núm. 6a, Núm. 7a, Núm. 7b, Núm. 7d, Núm. 8, Núm. 9, Núm. 10, Núm. 12 y Núm. 14, donde "a" y "b" se refieren a variantes donde el CR para *G. Partita* o el ACR para *G. monolis* respectivamente, fueron trasplantados) fueron modificados mediante la mutación de *Synechococcus* sp. PCC7942 *rbclS* con el kit de multi-mutagénesis QuickChange (Stratagene)  
35 utilizando cebadores apropiados. Además, varios mutantes con mutaciones únicas (Núm. 22-1A, Núm. 23-1A, Núm. 25, Núm. 26-1B) o dobles (Núm. 24), que son componentes de los mutantes pronosticados (Núm. 4, Núm. 8, Núm. 21- 1A, B), fueron diseñados como controles.

La secuencia genómica para la secuencia *rbcl-rbcS* (operón) de *Synechococcus* sp de 'tipo salvaje'. El documento PCC7942 se muestra en el SEC ID NO 22. Esta secuencia es la misma que para *Synechococcus* sp. PCC6301. En la secuencia del genoma nativo, la secuencia codificante *rbclS* lee ATG CCC (que codifica Met después Pro) pero en todas las secuencias *rbclS* clonadas es ATG GCC (que codifica Met después Ala). Esta sustitución de nucleótidos se introdujo para codificar un sitio de restricción único (*NcoI*) utilizado para la clonación del gen. También se introdujo una mutación silenciosa en el segundo último codón (Arg) en la secuencia *rbcSS* (CGA a CGC) con  
45 fines de clonación. La secuencia traducida para la subunidad grande de tipo salvaje se muestra en el SEC ID NO 24 y la secuencia traducida para la subunidad pequeña de tipo salvaje se muestra en el SEC ID NO 23. Las secuencias de nucleótidos y proteínas para los mutantes se muestran mediante los SEQ ID NO en la Tabla 5, por ejemplo los SEQ ID NO: 25 y 26, respectivamente, para el Mutante Núm. 1a.

Después de considerar estos resultados iniciales y el desarrollo del método de injerto filogenético ampliado, se diseñaron otros cuatro de los mutantes de LSU pronosticados detallados en el Ejemplo 4 y enumerados en la Tabla 4 (Mutante Núm. 17-1A, Núm. 18-1A, Núm. 19-1A, 21-1A, B) mediante el mismo método. Todos los mutantes de control (Núm. 22-1A, Núm. 23-1A, Núm. 24, Núm. 25, Núm. 26-1B) mencionados anteriormente son componentes de estos cuatro mutantes y son relevantes, en particular, para los mutantes Núm. 4 y Núm. 21-1A, B. Las secuencias de nucleótidos y proteínas para estos mutantes se muestran mediante los SEQ ID NO en la Tabla 5, por ejemplo los  
55 SEQ ID NO: 53 y 54, respectivamente, para el Mutante Núm. 17-1A.

Los genes *rbclS* mutados se secuenciaron antes de volverlos a clonar en el segundo plásmido de expresión que codificaba la LSU mutada y la SSU nativa. Las Rubisco mutantes se expresaron y purificaron utilizando un procedimiento descrito en el Ejemplo 6. Estos experimentos iniciales en los dieciocho mutantes de *Synechococcus* sp. PCC7942 y cinco mutantes control mostraron buena expresión en *E. coli* de Rubisco mutantes activas (es decir, plegadas y ensambladas correctamente), con especificidad y constantes cinéticas comparables con, o mejores que,  
60

el tipo salvaje, como se describe en el Ejemplo 7. La forma en que las pruebas experimentales y los procedimientos de optimización, detallados en los ejemplos 6-11, se integran con las etapas de predicción y de escrutinio *in silico* se muestra en el Diagrama de flujo 1. El Diagrama de flujo 1 también muestra cómo los resultados experimentales pueden ser retroalimentados en el procedimiento de predicción para refinar las predicciones de proteínas mutantes, seguido de ciclos adicionales de escrutinio *in silico* y escrutinio experimental y evaluación de la mejora de la propiedad funcional, para optimizarla.

#### Ejemplo 6

#### 10 Expresión y Purificación de Rubisco Mutantes

A pesar de que *E. coli* es el anfitrión microbiano más utilizado para expresar ADN y proteínas recombinantes, la obtención de la forma funcional de Rubisco a partir de organismos eucariotas (definida aquí como proteína de "forma I") es más compleja. Puesto que la holoenzima de forma I Rubisco es un hexadecámero elaborado a partir de 8 subunidades grandes (LSU) y 8 subunidades pequeñas (SSU), requiere chaperonas apropiadas para plegar y ensamblar correctamente la enzima correctamente. Convenientemente, sin embargo, cuando el operón que codifica los genes de Rubisco (*rbcLS* y *rbcSS*) de *Synechococcus* sp. PCC7942 se expresa en *E. coli* las subunidades LSU y SSU son abundantemente sintetizadas. Por lo tanto, *Synechococcus* sp. PCC7942 se utilizó aquí como la enzima L<sub>8</sub>S<sub>8</sub> modelo para la prueba inicial de las predicciones de Rubisco, y se utilizó *E. coli* como la elección natural de anfitrión de expresión para la producción de las Rubisco mutantes.

Sin embargo, solamente aproximadamente 1 a 5% de las LSU expresadas se pliegan correctamente y se ensamblan en forma funcional acumulándose la cantidad de Rubisco funcional a ~ 1 a 3% (peso/peso) de la proteína soluble de *E. coli*. La purificación de la Rubisco funcional por métodos tradicionales es un procedimiento laborioso y prolongado que puede tardar hasta 3 días.

El uso de etiquetas de afinidad 6xHistidine (H<sub>6</sub>) es una alternativa atractiva que podría ahorrar un esfuerzo y tiempo sustancial hacia la purificación enzimática. Pero los experimentos han demostrado que la fusión de etiquetas H<sub>6</sub> a cualquiera de los extremos de la LSU o SSU de la forma I de Rubisco puede comprometer la actividad catalítica. Para superar estas dificultades, se utilizó un sistema recientemente adaptado (Baker *et al.*, 2005) para simplificar y acelerar la purificación de Rubisco de *Synechococcus* sp. PCC7942 expresada en *E. coli*. Este sistema implicaba la construcción de un vector plasmídico único (basado en pACYC) que incorpora la fusión en marco de la secuencia codificante para una secuencia de ubiquitina (Ub) etiquetada con H<sub>6</sub> (H<sub>6</sub> Ub) al extremo 5' de *rbcSS*. El *rbcLS* de tipo salvaje en el plásmido pTrcSynLS (que contiene el operón *rbcL-rbcS* de PCC7942; Emlyn Jones *et al.*, 2006) fue reemplazado por el *rbcLS* mutado (*rbcLS\**) y a continuación co-transformado en *E. coli* con el plásmido basado en pACYC que codifica SSU de tipo salvaje etiquetada con H<sub>6</sub>Ub (H<sub>6</sub>Ub-SSU). Cuando se indujo la expresión de la subunidad Rubisco con IPTG, se produjeron los tres péptidos de la subunidad de Rubisco (es decir, LSU, SSU y H<sub>6</sub>UbSSU). Algunos fueron ensamblados en hexadecámeros de Rubisco funcionales compuestos de núcleos octaméricos 8xLSU y diferentes proporciones de SSU (como máximo 8) y H<sub>6</sub>UbSSU. Las Rubisco con una o más etiquetas de H<sub>6</sub> se purificaron fácilmente de otras proteínas de *E. coli* utilizando cromatografía de afinidad de metal inmovilizado (IMAC) y a continuación se escindió la secuencia H<sub>6</sub>Ub con una proteasa específica de H<sub>6</sub>Ub que, junto con los péptidos H<sub>6</sub>Ub no ensamblados, puede ser eliminada mediante IMAC. Utilizando este método, las Rubisco purificadas se aislaron del *E. coli* en aproximadamente 1 hora.

Dieciocho de las Rubisco mutantes y cinco mutantes de control especificados en el Ejemplo 5 e identificados en las Tablas 3 ó 4, así como el tipo salvaje, se expresaron y purificaron utilizando el procedimiento anterior.

#### Ejemplo 7

#### 50 Ensayo cinético *in vitro* de rubisco mutantes

Las proteínas Rubisco purificadas utilizando el método descrito en el Ejemplo 6 se usaron para medir las constantes de Michaelis para CO<sub>2</sub> ( $K_c$ ) y las tasas de carboxilación saturada por sustrato ( $V_c^{max}$ ) utilizando análisis de fijación de <sup>14</sup>CO<sub>2</sub> a 25°C, pH 8 de acuerdo con el método descrito por Andrews (1988).

La enzima purificada se preincubó a 25°C durante 30 min en tampón que contenía MgCl<sub>2</sub> 20 mM y NaHCO<sub>3</sub> 25 mM, y las mediciones de  $K_c$  se realizaron en viales de centelleo cerrados con tapón con septo purgados con nitrógeno. Las reacciones se iniciaron añadiendo 10 µl de enzima purificada a 0,5 ml de tampón de ensayo equilibrado con N<sub>2</sub> (EPPS 100 mM - NaOH, 20 mM de MgCl<sub>2</sub>, ribulosa 0,8 mM-P<sub>2</sub>, 0,1 mg/ml de anhidrasa carbónica) que contenía concentraciones variables de NaH<sup>14</sup>CO<sub>3</sub>.

Las constantes de Michaelis se determinaron ajustando los datos a la ecuación de Michaelis-Menten. La

5 cuantificación del contenido de Rubisco en los análisis se midió utilizando el análisis de unión de [2-<sup>14</sup>C] carboxiarabinitol - P<sub>2</sub> (<sup>14</sup>2CABP) descrito por Ruuska *et al.* (1998) y Whitney y Andrews (2001). El número de recambio de carboxilación saturada con sustrato ( $K_{cat}^c$ ) se calculó dividiendo la actividad carboxilasa máxima extrapolada ( $V_c^{max}$ ) por la concentración de sitios activos de Rubisco en el análisis. Las preparaciones de Rubisco purificadas también se utilizaron para medir la especificidad CO<sub>2</sub>/O<sub>2</sub> ( $S_{c/o}$ ) a pH 8,3 como describen Kane *et al.* (1994).

10 Un resumen de los resultados obtenidos se presenta en la Tabla 5. Para la especificidad, los resultados promedio proporcionan como % de cambio en comparación con el tipo salvaje: un valor positivo representa una mejora. Para la cinética, se proporcionan los valores y % de cambio en comparación con el tipo salvaje; un cambio de % positivo para la tasa catalítica  $K_{cat}^c$  y la eficacia catalítica  $K_{cat}^c/K_c$  representa una mejora, mientras que un cambio de % negativo para la constante de Michaelis  $K_c$  representa una mejora.

15 **Tabla 5.** Especificidad y resultados cinéticos para las Rubisco mutantes de *Synechococcus* sp. PCC7942. Los números en la columna "SEQ ID" corresponden a las secuencias en las listas de secuencias generadas por ordenador. n.m. representa no medido en el momento de la presentación.

Núm. de mutante (residuo o residuos modificados) <sup>c</sup>	SEQ ID <sup>e</sup>	Cinética			
		$S_{c/o}$ Promedio <sup>d</sup> (% cambio tipo salvaje)	$K_{cat}^c$ (s <sup>-1</sup> ) (% cambio tipo salvaje)	$K_c$ (μM) (% cambio tipo salvaje)	$K_{cat}^c/K_c$ (s <sup>-1</sup> mM <sup>-1</sup> ) (% cambio tipo salvaje)
Tipo salvaje	22-24	<b>41,6 ± 0,7</b>	<b>13,2 ± 0,2<sup>a</sup></b>	<b>203 ± 10<sup>a</sup></b>	<b>65<sup>a</sup></b>
			<b>13,0 ± 0,2<sup>b</sup></b>	<b>197 ± 10<sup>b</sup></b>	<b>66<sup>b</sup></b>
Núm. 1a (Y25W, D51I <sup>GP</sup> )	25,26	(3,4%)	11,5 ± 0,2 <sup>a</sup> (-13%)	176 ± 6 <sup>a</sup> (-13%)	65 (0%)
Núm. 1b (Y25 W, D51C <sup>Gm</sup> )	27,28	(-1,1 %)			
Núm. 4 (T23G, K81R)	29,30	(10,0 %)	14,1 ± 0,2 <sup>b</sup> (8%)	198 ± 8 <sup>b</sup> (0%)	71 (8%)
Núm. 5a (G54A <sup>GP</sup> , C84A, I87V)	31,32	(-6%)	11,3 ± 0,3 <sup>a</sup> (-14%)	182 ± 12 <sup>a</sup> (-10%)	62 (4%)
Núm. 5b (G54S <sup>Gm</sup> , C84A, I87V)	33,34	(1,9%)	10,6 ± 0,2 <sup>a</sup> (-20%)	167 ± 8 <sup>a</sup> (-18%)	63 (3%)
Núm. 6a (T23G, Y25W, D51I <sup>GP</sup> , K81R)	35,36	(8,5%)	12,2 ± 0,2 <sup>b</sup> (-6%)	185 ± 9 <sup>b</sup> (-6%)	66 (0%)
Núm. 7a (Y25W, D51I <sup>GP</sup> , G54A <sup>GP</sup> , C84A, I87V)	37,38	(8,8%)	12,4 ± 0,2 <sup>b</sup> (-5%)	195 ± 11 <sup>b</sup> (-1%)	64 (-3%)
Núm. 7b (Y25W, D51I <sup>GP</sup> , G54S <sup>Gm</sup> , C84A, I87V)	39,40	(1,4%)	n.m.	n.m.	n.m.
Núm. 7d (Y25W, D51V <sup>Gm</sup> , G54A <sup>GP</sup> , C84A, I87V)	41,42	(0,9%)	n.m.	n.m.	n.m.
Núm. 8 (V121I, M297G, V300T)	43,44	(5,6%)	n.m.	n.m.	n.m.
Núm. 9 (L36I, I116L, F140L)	45,46	(3,6%)	n.m.	n.m.	n.m.
Núm. 10 (L36I, I116L, V121I, F140L, M297G, V300T)	47,48	(2,9%)	10,8 ± 0,4 <sup>b</sup> (-17%)	325 ± 27 <sup>b</sup> (65%)	33 (-50%)
Núm. 12 (K181, T23G)	49,50	(1%)	n.m.	n.m.	n.m.
Núm. 14 (bucle AKMGYW)	51,52	(1,7%)	7,1 ± 0,4 <sup>b</sup> (-45%)	260 ± 36 <sup>b</sup> (32%)	27 (-59%)
Núm. 17-1A (T23G, K18I, T68V, K81R)	53,54	(-0,6%)	n.m.	n.m.	n.m.
Núm. 18-1A (T23G, K81R, P104E)	55,56	(-0,5%)	n.m.	n.m.	n.m.

Núm. de mutante (residuo o residuos modificados) <sup>c</sup>	SEQ ID <sup>e</sup>	S <sub>clo</sub>	Cinética		
		Promedio <sup>d</sup> (% cambio tipo salvaje)	$k_{cat}^c$ (s <sup>-1</sup> ) (% cambio tipo salvaje)	$K_c$ (μM) (% cambio tipo salvaje)	$k_{cat}^c/K_c$ (s <sup>-1</sup> mM <sup>-1</sup> ) (% cambio tipo salvaje)
Núm. 19 - 1A (T23G, D19P, K81R)	57,58	(-2,0%)	n.m.	n.m.	n.m.
Núm. 21 - 1A, B (T23G, K81R, V121I, M297G, V300T)	59,60	(-0,9%)	n.m.	n.m.	n.m.
Núm.22-1A(T23G)	61,62	(5,0%)	n.m.	n.m.	n.m.
Núm. 23-1A (K81R)	63,64	(7,0%)	n.m.	n.m.	n.m.
Núm. 24 (V121I, M297G)	65,66	n.m.	n.m.	n.m.	n.m.
Núm. 25 (M297G)	67,68	n.m.	n.m.	n.m.	n.m.
Núm. 26 - 1B (V121I)	69,70	n.m.	n.m.	n.m.	n.m.

a, b - se refiere a las mediciones repetidas en diferentes muestras de Rubisco realizadas en días diferentes.

c - cuando se proporcionan, las mutaciones especificadas están en las Subregiones 1A, 1A, B o 1B.

d - promedio de mediciones repetidas en diferentes muestras de Rubisco realizadas en días diferentes.

e Los números de ID de secuencia corresponden a los del archivo de secuencia.

De los dieciocho mutantes y dos mutantes individuales (control) para los cuales se realizaron mediciones de especificidad, todos mostraron especificidad comparable con el tipo salvaje, es decir, ninguno se alteró significativamente, mientras que cinco mutantes y los dos controles mostraron mejoras de 5% o mejores. De interés, cuatro de ellos incluyen las mutaciones T23G y/o K81R. También, es de interés que los mutantes con variantes de ACR (Núm. 1a y Núm. 1b para la posición 51, Núm. 5a y Núm. 5b para la posición 54, y Núm. 7a, Núm. 7b y Núm. 7d para las posiciones 51 y 54) mostraron diferencias significativas (del orden de 8% para las variantes de los mutantes Núm. 5 y Núm. 7) demostrando la sensibilidad de la especificidad a estos cambios que espacialmente están relativamente alejados del sitio activo (véase la Figura 10). La mejora en el valor S<sub>CO</sub> de 5,6% en el Mutante Núm. 8 era notable ya que las mutaciones estaban en una parte diferente del dominio N-terminal (véase la Figura 15) que la mayoría del conjunto inicial de mutantes pronosticado por el método básico.

De los 9 mutantes para los cuales se analizó  $k_{cat}^c$ , la mayoría mostraron valores de ligeramente a significativamente menores (es decir, peores) en comparación con el tipo salvaje con la excepción, Mutante Núm. 4, que mostró una mejora significativa de 8%. Los correspondientes 9 mutantes analizados para determinar  $K_c$  mostraron una gama de valores, mostrando 4 unión a CO<sub>2</sub> moderadamente mejorada (Núm. 1a, Núm. 5a, Núm. 5b, Núm. 6a), mostrando 2 poco cambio (Núm. 4, Núm. 7a) y mostrando 2 un deterioro significativo (Núm. 10, Núm. 14). Los valores globales de eficiencia catalítica ( $k_{cat}^c/K_c$ ) muestran un rango de mejora de pequeña a significativa (Núm. 4, Núm. 5a, Núm. 5b), a poco cambio (Núm. 1a, Núm. 6a, Núm. 7a) a deterioro significativo (Núm. 10, Núm. 12) en comparación con el tipo salvaje. Es notable que en el tabaco (véase el Ejemplo 11), el Mutante Núm. 23-1A mostró un valor de  $k_{cat}^c$  y también eficacia catalítica global sustancialmente peores, mientras que el Mutante Núm. 4 mostró mejoras significativas en las tres medidas cinéticas.

A partir de los resultados del conjunto inicial de mutantes pronosticado por el método básico (Núm. 1-Núm. 14), los mutantes destacados en términos de rendimiento global fueron los Mutantes Núm. 4 y Núm. 6a, que mostraron mejoras en la especificidad y la eficiencia cinética de 10 % y 8%, y 8,5% y 0%, respectivamente. Las propiedades del mutante más complejo (Núm. 6a), que comprendía las mutaciones en Núm. 4 y Núm. 1a, eran en general inferiores al Mutante Núm. 4 (el mejor mutante en la lista actual). Esta observación sugiere que pueden agruparse más ventajosamente otras mutaciones con las del Mutante Núm. 4 que las del Mutante Núm. 1a. Como se comentó en el Ejemplo 4, esta observación impulsó el desarrollo del concepto de Subregiones y el método ampliado, que condujo a las primeras predicciones de los Mutantes Núm. 17-1A, Núm. 18-1A y Núm. 19-1A, así como un mutante más complejo incluyendo los CR de la Subregión 1B también (Núm. 21-1A, B). Los resultados de especificidad para estas predicciones de método ampliado mostraron poco cambio en comparación con el tipo salvaje.

### 35 Ejemplo 8

Evolución dirigida de mutantes de *Synechococcus* en *E. coli*

Los mutantes injertados filogenéticos representan una estrategia "dirigida" para explorar regiones del espacio de secuencia no muestreadas naturalmente. Sin embargo, cualquier aumento en la actividad de estos mutantes puede verse afectado debido a algunas zonas de "ajustes de secuencia pobres", ya que pueden no estar optimizados para la estructura de Rubisco del anfitrión. Aunque el método de injerto filogenético ampliado proporciona una estrategia *in silico* que puede utilizarse para optimizar los mutantes de "líder", incluyendo el alivio de los conflictos estéricos mediante el reclutamiento de SvR y otros residuos variantes naturales que pueden ser identificados como complementarios a mutaciones en los "líderes", una opción alternativa para minimizar los efectos de estos conflictos puede ser utilizar un método de evolución dirigida experimental para optimizarlos, es decir, utilizar estas Rubiscos parcialmente optimizadas como puntos de partida. Estos mutantes también proporcionan en sí mismos un nuevo punto de partida para la evolución dirigida ya que tienen un potencial diferente para explorar el espacio de secuencias en comparación con el tipo salvaje.

Como se detalla en el Ejemplo 6, a diferencia de todas las Rubisco de la Forma I (es decir, hexadecamérica) de organismos eucariotas, la Rubisco de *Synechococcus* PCC7942 puede ensamblarse correctamente en *E. coli* y se ha elegido para un procedimiento de escrutinio de mutantes. Utilizando los métodos descritos en los Ejemplos 5-7, las predicciones de Mutantes Candidato (que pueden incluir grupos de mutaciones correlacionadas, y mutaciones independientes en diferentes regiones estructurales que rodean a los Residuos Diana) pueden escrutarse inicialmente en *E. coli* para confirmar que están activos y obtener constantes cinéticas *in vitro* para su comparación entre sí y con el tipo salvaje. Los mutantes con hasta 10-12 mutaciones se pueden producir rutinariamente utilizando la tecnología actual. Según se detalla en los Ejemplos 5-7, se pueden preparar mutantes únicos seleccionados para probar la hipótesis general subyacente al método de injerto filogenético de que los mutantes individuales son probablemente poco activos/inactivos y que son necesarias al menos dos mutaciones correlacionadas para producir una enzima aceptablemente activa.

Un sistema adecuado para la evolución dirigida de Rubisco en *E. coli* ha sido referido recientemente (*Mueller-Cajar et al.*, 2007). Éste utiliza una cepa de *E. coli*, MM1, cuyo crecimiento puede depender de la expresión funcional de Rubisco, cuando se co-expresa con fosforribuloquinasa (PRK). La glicólisis en MM1 se bloqueó por delección del gen de gliceraldehído 3-fosfato deshidrogenasa (*gapA*) y se introdujo una desviación de la derivación metabólica que comprende una Rubisco de *Synechococcus* PRK y *Rhodospirillum rubrum*. Como resultado, MM1 depende de la expresión funcional de Rubisco para metabolizar el producto de la catálisis de PRK, la ribulosa-1,5-bisfosfato, que es tóxica para *E. coli*.

Este método general puede utilizarse para evaluar si las Rubisco con actividad significativamente mejorada pueden evolucionar más eficazmente a partir de formas inactivadas de los más prometedores mutantes injertados de *Synechococcus* sp. PCC7942 detallados en el Ejemplo 7 y la Tabla 5. Para este propósito, las bibliotecas mutagenizadas aleatoriamente (preparadas utilizando los métodos referidos por *Mueller-Cajar et al.*, 2007) de estos genes inactivados se pueden transformar en células MM1 cultivadas bajo diferentes condiciones selectivas (p.ej., variando las presiones de CO<sub>2</sub>/O<sub>2</sub> de crecimiento, cambiando el alcance de la producción de PRK). Se pueden aislar, secuenciar las colonias que expresan variantes de Rubisco evolucionadas con una adaptabilidad mejorada (es decir, aquellas que sobreviven al escrutinio) y caracterizar la cinética de las Rubisco mutadas purificadas como se detalla en el Ejemplo 7.

#### Ejemplo 9

#### Pruebas de la competencia bioquímica y fisiológica de mutantes de *Synechococcus in vivo*

Las pruebas funcionales *in vitro* del Ejemplo 7 identificaron varias *Synechococcus* mutantes con actividad de Rubisco mejorada, y que también, necesariamente, se plegaron y ensamblaron correctamente cuando se expresaron a partir del sistema de expresión de *E. coli* y se purificaron como se describe en el Ejemplo 6. En la estrategia de reestructuración de Rubisco, se ha utilizado *Synechococcus* como anfitrión inicial más conveniente para la prueba experimental de las predicciones de mutantes para identificar candidatos "líder". Debido a los recientes avances en el diseño de Rubisco mutante en una planta modelo, el tabaco, utilizando la transformación de plástidos, como se describe en el Ejemplo 10, en el trabajo descrito aquí un Mutante Candidato (Núm. 4) identificado como un mutante de *Synechococcus* prometedor se sometió a ensayo directamente en la planta con flores de ensayo (tabaco) sin emprender la etapa intermedia, mostrada en el Diagrama de flujo 1, de confirmar que se puede aclimatar de nuevo a su anfitrión nativo (*Synechococcus*) sin perjuicio de su fenotipo. Este método mejorado permitió la evaluación del ensamblaje y la cinética de las Rubisco mutantes en tabaco en el plazo de 7-9 semanas, lo cual es más rápido que lo requerido para obtener y cultivar transformantes de cianobacterias. Cuando el propósito último del desarrollo mutante es para diseñar Rubisco mejorada en plantas, se podría preferir esta estrategia. Sin embargo, si el propósito final es para diseñar Rubisco mejorada en cianobacterias, en ese caso se prefiere realizar esta etapa de prueba fisiológica en *Synechococcus* para evaluar si las mutaciones catalíticamente beneficiosas podrían introducir problemas de incompatibilidad que perturben el plegamiento-ensamblaje productivo por complejos de chaperonas de

cianobacterias u otras proteínas interactuantes tales como las implicadas en la formación del carboxisoma. A continuación se describe un método para realizar estas pruebas en *Synechococcus*.

5 En *Synechococcus* sp. PCC7942, los genes de Rubisco (*rbcLS* y *rbcSS*) están codificados por un solo operón en el cromosoma (de los cuales hay típicamente 5 copias de cromosomas por célula) y, análogamente a *E. coli*, esta cepa de cianobacterias es naturalmente competente y puede ser transformada genéticamente mediante modificaciones dirigidas a su cromosoma (por ejemplo delección de genes, sustitución de genes) mediante recombinación homóloga o por retención estable de vectores lanzadera plasmídicos dentro de sus células.

10 Se han desarrollado cepas mutantes de *Synechococcus* PCC7942 (véase Emlyn-Jones *et al.*, 2006 y Price *et al.*, 1993 como ejemplos). Se puede utilizar una cepa mutante de *Synechococcus* PCC7942 en la que se elimina el operón *rbcLS-rbcSS* cromosómico (es decir, una cepa 7942 $\Delta$ *rbcLS*) para facilitar la reintroducción de los genes *rbcLS-rbcSS* de Rubisco mutante. Puesto que *Synechococcus* sp. PCC7942 no puede crecer heterotróficamente (es decir, en una fuente externa de carbono) y requiere una Rubisco para el crecimiento, las cepas 7942 $\Delta$ *rbcLS* se pueden generar: (1) introduciendo un segundo gen de Rubisco (p.ej. *rbcM* que codifica el homodímero de Rubisco de Forma II (L<sub>2</sub>) estructuralmente diferente de la bacteria *Rhodospirillum rubrum* o la Rubisco PCC7942 L<sub>3</sub>S<sub>8</sub> nativa) en un vector lanzadera de plásmido en *Synechococcus* sp. PCC7942 y a continuación (2) mediante recombinación homóloga en un gen de resistencia a antibióticos km<sup>R</sup> para reemplazar el operón *rbcLS-rbcSS* en cada copia del cromosoma (es decir, para que la mutación pueda ser completamente segregada). Las células de *Synechococcus* PCC7942 transformadas con *rbcM* expresado en un vector lanzadera pueden transformarse posteriormente con otro plásmido para reemplazar homológamente la región codificante *rbcLS-rbcSS* cromosómica por un gen km<sup>R</sup>. Tras el aislamiento de los transformantes  $\Delta$ *rbcLS-rbcSS*::km<sup>R</sup> completamente segregados (es decir, todos los cromosomas tienen *rbcS-rbcSS* reemplazado por km<sup>R</sup>) las células PCC7942 $\Delta$ *rbcLS* pueden usarse para reintroducir homológamente los genes *rbcLS\** y *rbcSS* mutados y las células curadas del vector lanzadera.

25 Utilizando técnicas establecidas (p.ej., véase Emlyn-Jones *et al.*, 2006), El fenotipo de las células transformadas puede caracterizarse a continuación bioquímicamente, por ejemplo, evaluando si las subunidades LSU de Rubisco mutadas son fácilmente plegadas y ensambladas adecuadamente, y fisiológicamente, por ejemplo, evaluando si hay diferencias en la capacidad fotosintética, partición de carbono inorgánico o tasa de crecimiento.

30 Ejemplo 10

Generación de transformantes de plastoma de Rubisco en tabaco

35 Una estrategia para la reestructuración de Rubisco que es aplicable a Rubisco de plantas superiores es por transformación de plastoma utilizando genes *rbcL\** mutados. El plastoma del tabaco es fácilmente transformable (Andrews y Whitney, 2003) y se utilizó como modelo para llevar a cabo ensayos de prueba de principio para la transformación de otras plantas.

40 El Mutante Núm. 4 de Rubisco de *Synechococcus* más prometedor se utilizó como caso de prueba inicial. También se sometieron a ensayo los mutantes individuales componentes, T23G (Mutante Núm. 22-1A) y K81R (Mutante Núm. 23-1A). Los SEQ ID NO de nucleótidos y proteínas para los Mutantes Núm. 4 y Núm. 23-1A del tabaco se proporcionan en la Tabla 6. Los SEQ ID NO para el Mutante Núm. 22-1A son 77 y 78. Dado que estos residuos son Residuos Candidato (véase la Tabla 2), estas mismas mutaciones se usaron en tabaco como en *Synechococcus*.

45 Se dispone de líneas de tabaco transplastómicas que permiten un escrutinio más rápido de la cinética de las Rubisco de tabaco mutadas que los largísimos métodos tradicionales de transformación de cloroplastos en los que los genes *rbcL* nativos en el plastoma son sustituidos por versiones mutadas (Andrews y Whitney, 2003 (*supra*)) utilizando la técnica de transformación biolística descrita en Svab y Maliga (1993). En una reciente mejora del método para volver a transformar genes *rbcL* mutados o foráneos al plastoma de tabaco el gen *rbcL* nativo se reemplazó por el gen *rbcM* de *Rhodospirillum rubrum* y el gen marcador seleccionable *aadA* (que codifica la resistencia a espectinomicina) y a continuación el gen *aadA* se eliminaron para producir líneas transplastómicas tabaco-*rubrum* libres de marcadores ( $\Delta$ *aadA*). Estas líneas se generaron mediante la transformación biolística del plastoma de tabaco silvestre (*Nicotiana tabacum* L. cv Petit Havana [N,N]) con el plásmido p<sup>cm</sup>trLA (número de acceso de Genbank AY827488). El gen *aadA* en los transformantes está flanqueado por sitios *LoxP* de 34 pb que permiten su escisión mediante recombinación CRE-lox. Pare escindir *aadA*, se bombardearon biolísticamente hojas de una línea transformada biológicamente con p<sup>cm</sup>trLA con el plásmido pKO27 que expresa CRE (Corneille *et al.*, 2001). Las hojas bombardeadas se diseccionaron (~0,5 cm<sup>2</sup>) y se propagaron en medio selectivo de kanamicina (sales de Murashige-Skoog solidificadas con agar conteniendo sacarosa al 3% (p/v), 15  $\mu$ g ml<sup>-1</sup> de kanamicina y hormonas (Svab y Maliga, 1993). Las primeras plántulas que emergieron de las secciones de hojas bombardeadas blanqueadas se transfirieron a medio MS (medio selectivo sin kanamicina u hormonas) y pérdida de *aad* confirmada por análisis rutinarios de PCR.

Las líneas de tabaco-*rubrum*  $\Delta aadA$  permiten la reutilización del gen marcador *aadA* para transformar posteriormente su plastoma, por ejemplo volviéndolo a transformar en variantes *rbcl*\* de tabaco mutadas para reemplazar *rbclM*. La eficacia de transformación para reemplazar *rbclM* en tabaco-*rubrum*  $\Delta aadA$  con genes *rbcl*\* variantes es de 3 a 10 veces mayor que la transformación de tabaco de tipo salvaje, y es inmune a los eventos de recombinación no deseados que pueden ocurrir al transformar genes *rbcl*\* en plastomas de tabaco de tipo salvaje. El método permitió la producción de transformantes que contenían sólo copias de plastoma transformadas con *rbcl*\* (es decir, transformantes homoplásicos) en el plazo de 6 a 8 semanas puesto que las copias de plastoma que contienen el gen *rbclM* se eliminaron rápidamente. Puesto que Rubisco de *R. rubrum* es un pequeño homodímero de LSU ( $L_2$ , ~100 kDa), las líneas transformadas con *rbcl*\* que producen las Rubisco  $L_8S_8$  de forma I más grandes (~520 kDa) se pueden identificar separando la proteína de hoja soluble mediante electroforesis en gel de poliacrilamida no desnaturizante como describen Whitney y Sharwood (2007) y se puede medir la homoplasmicidad por la ausencia de Rubisco  $L_2$ .

Este sistema es potencialmente adaptable a otras plantas donde se ha referido la transformación de plástidos. La transformación genética de plástidos en una variedad de plantas diferentes ha sido referida, por ejemplo por Koop *et al.* (1997).

Se realizaron mutaciones para los mutantes Núm. 4, Núm. 22-1A y Núm. 23-1A en la secuencia codificante de *rbcl* de plastoma de tabaco (*Nicotiana tabacum*) de tipo salvaje (Stratagene) utilizando cebadores apropiados, e introducidos en el plásmido pLEV1 transformante de plastoma de tabaco donde la selección de transformantes es facilitada por la incorporación de un gen *aadA* sin promotor aguas abajo de *rbcl* (Whitney *et al.*, 1999). Los plásmidos transformantes derivados de pLEV1 que codifican copias de *rbcl*\* de tabaco mutagenizadas con las secuencias de nucleótidos que codifican los mutantes Núm. 4, Núm. 22-1A y Núm. 23-1A, así como el tipo salvaje, se transformaron biolísticamente en una línea de tabaco-*rubrum*  $\Delta aadA$  y en plántulas resistentes a espectinomicina seleccionadas como se ha descrito (Svab y Maliga, 1993). No se aislaron plántulas para el Mutante Núm. 22-1A. Los transformantes donde *rbclM* había sido reemplazado por los genes *rbcl*\* o *rbcl* fueron identificados mediante la producción de Rubisco  $L_8S_8$  utilizando electroforesis en gel de poliacrilamida no desnaturizante (ver Figura 20).

#### Ejemplo 11

##### Caracterización bioquímica y fisiológica de los transformantes de Rubisco de tabaco

Los siguientes métodos se utilizaron para extraer y purificar la Rubisco expresada en células de tabaco transplastómicas homoplasmáticas y de tipo salvaje para llevar a cabo análisis cinéticos.

Se utilizaron análisis de fijación de  $^{14}CO_2$  radiomarcado para medir el número de recambio con saturación de sustrato ( $k_c^{cat}$ ) y la constante de Michaelis para  $CO_2$  a 0% ( $K_c$ ) o 21% de  $O_2$  ( $K_c^{aire}$ ) utilizando extracto soluble de proteína de la hoja. Se obtuvieron discos foliares (~1 cm<sup>2</sup>) durante el fotoperiodo y se extrajeron en hielo utilizando homogeneizadores de vidrio (Wheaton, EE.UU.) en 0,8 ml de tampón de extracción sin  $CO_2$  (Bicina 50 mM-NaOH, pH 8,0, EDTA 1 mM, DTT 2 mM, cóctel inhibidor de proteasa vegetal al 1% (v/v) de (Sigma-Aldrich) y PVPP al 1% (p/v). La muestra se centrifugó (36.000 g, 5 min, 4°C) y la proteína soluble se incubó (activó) con  $NaHCO_3$  y  $MgCl_2$  (15 mM cada uno) durante 15 minutos y se utilizó para medir el contenido de Rubisco en alícuotas duplicadas incubadas con 40  $\mu M$  de  $^{14}C$ -CABP y la cantidad de Rubisco-bound- $^{14}C$ -CABP se recuperó mediante filtración en gel (Ruuska *et al.*, 1998), o se utilizó para medir  $K_c$  y  $K_c^{aire}$  a 25°C, pH 8,0 utilizando análisis de fijación de  $^{14}CO_2$  (Andrews 1988, Whitney y Sharwood, 2007). Para confirmar que las muestras utilizadas eran homoplásicas (es decir, sólo contienen copias de plastoma transformadas con los genes *rbcl*\* y ninguno con *rbclM*), la proteína también se separó en geles de poliacrilamida no desnaturizantes (Figura 20). En cuanto a los controles de tipo salvaje, sólo Rubisco  $L_8S_8$  mutada y no Rubisco  $L_2$  se expresó en los transformantes de los Mutantes Núm. 4 y Núm. 23-1A.

Los ensayos cinéticos se iniciaron añadiendo el extracto de proteína soluble activado a viales de centelleo con tapón con septo que contenían tampón de análisis (Bicina 100 mM - NaOH,  $MgCl_2$  15 mM, ribulosa- $P_2$  0,6 mM, anhidrasa carbónica de 0,1 mg · ml<sup>-1</sup>) equilibrado con  $N_2$  (para  $K_c$ ) o aire sin  $CO_2$  (para  $K_c^{aire}$ ) que contenía  $^{14}CO_2$  de 0 a 90  $\mu M$ . La ribulosa- $P_2$  se sintetizó de acuerdo con (Kane *et al.*, 1998). Los análisis se detuvieron después de 1 min con 0,5 volúmenes de ácido fórmico al 25% (v/v), las reacciones se secaron a 80°C y el residuo se disolvió en agua, se añadieron dos volúmenes de líquido de centelleo, se sometieron a vórtice y se midió el  $^{14}C$  por conteo de centelleo. Las  $K_c$  y  $K_c^{aire}$  se calcularon a partir del gráfico de Michaelis-Menten de la velocidad de carboxilación frente a  $[CO_2]$ .

Mediciones de  $k_c^{cat}$  se calcularon utilizando análisis de fijación de  $^{14}CO_2$  que contenían  $NaH^{14}CO_3$  15 mM y dividiendo la velocidad de carboxilación con saturación de sustrato,  $V_c^{m\acute{a}x}$ , mediante la concentración de sitios activos de Rubisco medida por  $^{14}C$ -CABP (véase más arriba).

Las mediciones de especificidad se realizaron utilizando Rubisco purificada. Se extrajo la proteína soluble de la hoja como se ha descrito anteriormente y 0,2 mL se sometieron a cromatografía a través de una columna Superdex 200HR 10/30 equilibrada con tampón de especificidad (Trietanolamina 30 mM de pH 8,3, acetato de Mg 30 mM) utilizando un sistema explorador ÄKTA (APBiotech). Las tres fracciones pico (0,3 ml) que contenían Rubisco L<sub>8</sub>S<sub>8</sub> se agruparon (sitios de la subunidad L de ~100-150 pmoles) y se usaron para medir la especificidad de CO<sub>2</sub>/O<sub>2</sub> a 25°C como se describe (Kane *et al.*, 1994) después de equilibrar con una atmósfera que contenía 500 ppm de CO<sub>2</sub> en O<sub>2</sub> controlado mediante tres bombas de mezcla de gas de precisión Wösthoff.

Un resumen de los resultados obtenidos se presenta en las Tablas 6 y 7. Para la especificidad, los resultados promedio se proporcionan como % de cambio en comparación con el tipo salvaje: un valor positivo representa una mejora. Para la cinética, se proporcionan los valores y % de cambio en comparación con el tipo salvaje; un cambio de % positivo para la tasa catalítica  $k_{cat}^c$  y la eficacia catalítica  $k_{cat}^c/K_c$  representa una mejora, mientras que un cambio de % negativo para la constante de Michaelis  $K_c$  representa una mejora.

**Tabla 6.** Especificidad y resultados cinéticos medidos en aire (21% de O<sub>2</sub>) para las Rubisco mutantes de tabaco.

Núm. de mutante (residuos o residuos modificados)	ID de Seq <sup>a</sup>	S <sub>c/o</sub>			
		Promedio (% cambio del tipo salvaje)	$k_{cat}^c$ (s <sup>-1</sup> ) (% de cambio del tipo salvaje)	$K_c^{aire}$ (μM) (% de cambio del tipo salvaje)	$k_{cat}^c/K_c^{aire}$ (s <sup>-1</sup> mM <sup>-1</sup> ) (% de cambio del tipo salvaje)
tipo salvaje	71,72	81,7 ± 2,1	3,59 ± 0,02	20,8 ± 1,4	173
Núm. 4 (T23G, K81R)	73,74	79,3 ± 1,4 (-3%)	3,84 ± 0,05 (7%)	18,2 ± 0,6 (-13%)	209 (21%)
Núm. 23-1A (K81R)	75,76	80,7 ± 1,9 (-1%)	2,46 ± 0,17 (-31%)	17,8 ± 1,0 (-14%)	138 (-20%)

a - Los números de ID de secuencia corresponden a los de la lista de secuencias; el primer número es el SEQ ID NO de nucleótido y el segundo número es el SEQ ID NO de proteína.

Los resultados de la Tabla 6 para las mediciones en aire (21% de O<sub>2</sub>) muestran mejoras significativas sobre el tipo salvaje en ambos valores de  $k_{cat}$  y  $K_c^{aire}$  para el Mutante Núm. 4 de 7% y -13%, respectivamente, produciendo una mejora global en la eficacia catalítica de 21%. Sin embargo, aunque el Mutante individual Núm. 23-1A muestra una mejora similar de  $K_c^{aire}$  (-14%), tiene un índice significativamente deteriorado (-31%) produciendo una disminución total de la eficacia catalítica de 20%. Los resultados cinéticos (análisis por triplicado) se obtuvieron utilizando un extracto soluble de proteína de hoja de dos transformantes distintos (tipo salvaje, Núm. 23-1A) o tres (Núm. 4) diferentes. Las Rubisco aisladas replicadas proporcionaron resultados similares, contribuyendo a los pequeños errores de las mediciones.

Los resultados de especificidad para ambos mutantes muestran poco cambio en comparación con el tipo salvaje. Estos valores se obtuvieron a partir de mediciones por cuadruplicado utilizando Rubisco purificadas a partir de dos transformantes de tipo salvaje, dos mutantes Núm. 4 y dos mutantes Núm. 23-1A.

La consideración de los resultados de especificidad y cinéticos en conjunto indica que la mejora en la eficacia de carboxilación para el Mutante Núm. 4 se ha visto acompañada por una mejora comparable en la eficacia de la oxigenación. Como prueba inicial de si los cambios en la eficiencia de la oxigenación podrían deberse a cambios en la  $k_{cat}^o$  o  $K_o$ ,  $K_c$  se midió en 0% de O<sub>2</sub>. Estos resultados, proporcionados en la Tabla 7, muestran que ambos mutantes tienen peores valores de  $K_c$  que el tipo salvaje, 18% y 9% más alto para los mutantes Núm. 4 y Núm. 23-1A, respectivamente. Esto indica que los valores de  $K_o$  para los mutantes son mejores que el tipo salvaje, es decir, son menos inhibidos por O<sub>2</sub>. Los valores aproximados para  $K_o$ , mostrados en la última columna, se calcularon de acuerdo con Whitney *et al.* (1999). Puesto que estas mediciones de  $K_o$  se basan en mediciones de  $K_c$  a sólo dos concentraciones de O<sub>2</sub> (es decir, 0 y 21% (v/v) en nitrógeno), se deben considerar sólo estimaciones. Sin embargo, suponiendo que sean fiables, se puede concluir que la  $k_{cat}^o$  para los mutantes debe mejorarse para explicar la mejora en la eficacia de oxigenación sugerida por los resultados de especificidad.

**Tabla 7.** Resultados de especificidad y cinéticos medidos en 0% de O<sub>2</sub> para las Rubisco de tabaco mutantes.

Núm. de mutante (residuo o residuos modificados)	cinética			
	$k_{cat}^c$ (s <sup>-1</sup> ) (% cambio del tipo salvaje)	$K_c$ (μM) (% cambio del tipo salvaje)	$k_{cat}^c/K_c$ (s <sup>-1</sup> mM <sup>-1</sup> ) (% de cambio del tipo salvaje)	$K_i(O_2)$ (μM)
tipo salvaje	3,59 ± 0,02	12,7 ± 0,1	283	395

Núm. de mutante (residuo o residuos modificados)	cinética			
	$k_{cat}^c$ ( $s^{-1}$ ) (% cambio del tipo salvaje)	$K_c$ ( $\mu M$ ) (% cambio del tipo salvaje)	$k_{cat}^c/K_c$ ( $s^{-1} mM^{-1}$ ) (% de cambio del tipo salvaje)	$K_i(O_2)$ ( $\mu M$ )
Núm. 4 (T23G, K81R)	3,84 ± 0,05 (7%)	15,0 ± 0,7 (18%)	256 (-10%)	1157
Núm. 23 - 1A (K81R)	2,46 ± 0,17 (-31%)	13,9 ± 0,6 (9%)	177 (-37%)	884

En resumen, los resultados muestran que el Mutante Núm. 4 conserva sus propiedades superiores en el tabaco, aunque se expresan más como mejoras en la eficacia catalítica que como mejoras aproximadamente iguales en la especificidad y la eficacia, como es el caso del Mutante Núm. 4 de *Synechococcus*.

5 Las líneas transplastómicas homoplásmicas y los controles de tabaco de tipo salvaje pueden cultivar hasta su madurez en tierra en cámaras de crecimiento de entorno controlado o invernaderos de vidrio y a continuación se pueden llevar a cabo ensayos fisiológicos convencionales. Estos ensayos se realizan sólo para las Rubisco mutantes más prometedoras que muestran mejoras de la especificidad y/o la cinética sobre el tipo salvaje e incluyen la determinación de la capacidad fotosintética, la tasa de crecimiento y la bioquímica foliar, incluyendo el nivel de expresión de Rubisco, el metabolito, los niveles de carbono y nitrógeno utilizando metodologías establecidas (Whitney *et al.*, 1999; Whitney y Andrews, 2001; Whitney *et al.*, 2001)

### Referencias

- 15 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- Andrews TJ. (1988) Catalysis by cyanobacterial ribulosebisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* 263, 12213-12220.
- Andrews TJ, Whitney SM. (2003) Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Arch. Biochem. Biophys.* 414, 159-169.
- 20 Baker RT, Catanzariti AM, Karunasekara Y, Soboleva TA, Sharwood R, Whitney S, Board PG. (2005) Using deubiquitylating enzymes as research tools. *Methods Enzymol.* 398, 540-554.
- Case DA, Darden TA, Cheatham III, TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. (2006) AMBER 9, University of California, San Francisco. Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacharya D. (2004) Hidden biodiversity of the extremophilic Cyanidiales red algae. *Mol. Ecol.* 13, 1827-1838.
- Corneille S, Lutz K, Svab Z, Maliga P. (2001) Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system. *Plant J.* 27, 171-178.
- 30 Cummins PL. (1996) Molecular Orbital Programs for Simulations (MOPS), Australian National University, Canberra.
- Cummins PL, Gready JE. (1997) A coupled semiempirical molecular orbital and molecular mechanical model (QM/MM) for organic molecules in aqueous solution. *J. Comput. Chem.* 18, 1496-1512.
- 35 Cummins PL, Gready JE. (1998) A molecular dynamics and free energy perturbation (MD/FEP) study of the hydride-ion transfer step in dihydrofolate reductase using a combined quantum and molecular mechanical (QM/MM) model. *J. Comput. Chem.* 19, 977-988.
- Cummins PL, Gready JE. (1999) Coupled semiempirical quantum mechanics and molecular mechanics model (QM/MM) calculations on the aqueous solvation energies of ionised molecules. *J. Comput. Chem.* 20, 1028-1038.
- 40

- Cummins PL, Gready JE. (2003) Computational methods for the study of enzymic reaction mechanisms II: An overlapping mechanically embedded method for hybrid semiempirical-QM/MM calculations. *THEOCHEM* 632, 245-255.
- 5 Cummins PL, Gready JE. (2005) Computational methods for the study of enzymic reaction mechanisms III: a perturbation plus QM/MM approach for calculating relative free energies of protonation. *J. Comput. Chem.* 26, 561-568.
- Cummins PL, Rostov I, Gready JE. (2007) Calculation of a complete enzymic reaction surface: reaction and activation free energies for hydride-ion transfer in dihydrofolate reductase. *J. Chem. Theor. Comput.* 3, 1203-1211.
- 10 Emlyn-Jones D, Woodger FJ, Price GP, Whitney SM. (2006) RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* 47, 1630-1640.
- Evans JR, Austin RB. (1986) The specific activity of ribulose-1,5-bisphosphate carboxylase in relation to genotype in wheat. *Planta* 167, 344-350.
- Fersht A. (1998) *Structure and mechanism in protein science: guide to enzyme catalysis and protein folding.* W. H. Freeman & Co., 1998.
- 15 Frey PA, Hegeman A. (2007) *Enzymatic reaction mechanisms.* Oxford University Press USA, 2007.
- Frisch MJ, (80 co-authors) and Pople JA. (2004) *Gaussian 03, Revision C.02,* Gaussian Inc., Wallingford, CT.
- Galmés J, Flexas J, Keys AJ, Cifre J, Mitchell RAC, Madgwick PJ, Haslam RP, Medrano H, Parry MAJ. (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ.* 28, 571-579.
- 20 Gready JE, Rostov I, Cummins PL. (2006) Simulations of enzyme reaction mechanisms in active sites: accounting for an environment which is much more than a solvent perturbation. In "Modelling Molecular Structure and Reactivity in Biological Systems", K.J. Naidoo, M. Hann, J. Gao, M. Field and J. Brady, eds, Royal Society of Chemistry, London, pp. 101-118.
- 25 Hall TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95.
- Kane HJ, Viil J, Entsch B, Paul BK, Morell MK, Andrews TJ. (1994) An improved method for measuring the CO<sub>2</sub>/O<sub>2</sub> specificity of ribulosebisphosphate carboxylase-oxygenase. *Aust. J. Plant Physiol.* 21, 449-461.
- 30 Kane HJ, Wilkin JM, Portis AR, Andrews TJ. (1998) Potent inhibition of ribulosebisphosphate carboxylase by an oxidized impurity in ribulose-1,5-bisphosphate. *Plant Physiol.* 117, 1059-1069.
- Koop H-U, Herz S, Golds TJ, Nickelsen J. (2007) The genetic transformation of plastids. *Topic in Current Genetics.* DOI 10.1007/4735\_2007\_0225 / Published online: 15 May 2007.
- Mueller-Cajar O, Morell M, Whitney SM. (2007) Directed evolution of Rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the Form II enzyme. *Biochemistry*, in press, Sept 2007.
- 35 Parry MA, Andralojc PJ, Mitchell RA, Madgwick PJ, Keys AJ. (2003) Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exptl Bot.* 54, 1321-1333.
- Parry MAJ, Madgwick PJ, Carvalho JFC, Andralojc PJ. (2007) Prospects from increasing photosynthesis by overcoming the limitations of Rubisco. *J. Ag. Sci.* 145, 31-43.
- 40 Price GD, Howitt SM, Harrison K, Badger MR. (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp. strain PCC7942 involved in carboxysome assembly and function. *J. Bacteriol.* 175, 2871-2879.

Ruuska S, Andrews TJ, Badger MR, Hudson GS, Laisk A, Price GD, von Caemmerer S. (1998). The interplay between limiting processes in C-3 photosynthesis studied by rapid-response gas exchange using transgenic tobacco impaired in photosynthesis. *Aust. J. Plant Physiol.* 25, 859-870.

Svab Z, Maliga P. (1993) High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc. Natl Acad. Sci. USA* 90, 913-917.

Spreitzer RJ, Salvucci ME. (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant. Biol.* 53, 449-475.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Whitney SM, von Caemmerer S, Hudson GS, Andrews TJ. (1999) Directed mutation of the Rubisco large subunit of tobacco influences photorespiration and growth. *Plant Physiol.* 121, 579-588.

Whitney SM, Andrews TJ. (2001) Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) supports photosynthesis and growth in tobacco. *Proc. Natl Acad. Sci. USA* 98, 14738-14743.

Whitney SM, Baldet P, Hudson GS, Andrews TJ. (2001). Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* 26, 535-547.

Whitney SM, Sharwood RE. (2007) Linked Rubisco subunits can assemble into functional oligomers without impeding catalytic performance. *J. Biol. Chem.* 282, 3809-3818.

LISTA DE SECUENCIAS

<110> Universidad Nacional Australiana  
 <120> Procedimiento para la generación de Proteína y usos del mismo  
 <130> N.107705  
 5 <140> EP 07815347.5  
 <141> 10-10-2007  
 <150> AU 2006905622  
 <151> 10-10-2006  
 <160> 78  
 10 <170> PatentIn version 3.2  
 <210> 1  
 <211> 475  
 <212> PRT  
 <213> Cyanophora paradoxa  
 15 <400> 1

```

Met Ser Ser Gln Ala Arg Thr Gln Thr Arg Ala Gly Phe Lys Ala Gly
1      5      10
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Thr Pro Lys
20      25      30
Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
35      40      45
Pro Pro Glu Glu Cys Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
50      55      60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65      70      75      80
Lys Gly Arg Ser Tyr Gly Phe Glu Pro Val Pro Gly Glu Glu Asn Gln
85      90      95
Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
100     105     110
Val Thr Asn Met Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
115     120     125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Val Gly Tyr
130     135     140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Thr Val Glu Arg Asp
145     150     155     160
Lys Leu Asn Lys Tyr Gly Arg Ala Leu Leu Gly Cys Thr Ile Lys Pro
165     170     175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
180     185     190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
195     200     205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Tyr Val Met Asp Ala
210     215     220
Ile Lys Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225     230     235     240
Asn Ala Thr Ala Pro Thr Thr Glu Glu Met Ile Lys Arg Ala Glu Phe
245     250     255
Ala Ala Glu Leu Asp Ala Pro Ile Ile Met His Asp Tyr Ile Thr Ala
260     265     270
Gly Phe Thr Ser Asn Thr Thr Leu Ala Arg Trp Cys Arg Asp Asn Gly
275     280     285
Pro Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
290     295     300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Thr Leu Arg Met
305     310     315     320
Ser Gly Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu
325     330     335
    
```

# ES 2 640 747 T3

```

Gly Asp Arg Ala Gly Thr Leu Gly Phe Val Asp Leu Met Arg Asp Asp
      340      345      350
His Ile Glu Gln Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp Trp
      355      360      365
Ala Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Ile
      370      375      380
Trp His Met Pro Ala Leu Val Asp Ile Phe Gly Asp Asp Ser Cys Leu
      385      390      395
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
      405      410      415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
      420      425      430
Glu Gly Arg Asn Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
      435      440      445
Ala Arg Phe Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
      450      455      460
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Ile
      465      470      475

```

<210> 2

<211> 488

5

<212> PRT

<213> Artificial

<220>

<223> secuencia consenso del 50% para LSU de Rubisco de rodofita (9 especies)

<220>

10

<221> RASGO\_MISC

<222> (2) .. (2)

<223> el residuo es uno de A, C, D, G, N, P, S, T o V

<220>

<221> RASGO\_MISC

15

<222> (3) .. (3)

<223> el residuo es uno de C, D, E, H, K, N, Q, R, S o T

<220>

<221> RASGO\_MISC

20

<222> (6) .. (6)

<223> el residuo es uno de C, D, E, H, K, N, Q, R, S o T

<220>

<221> RASGO\_MISC

<222> (93) .. (93)

<223> el residuo es uno de A, G o S

25

<220>

<221> RASGO\_MISC

<222> (97) .. (97)

<223> el residuo es uno de S o T

<220>

30

<221> RASGO\_MISC

<222> (159)..(159)

<223> el residuo es uno de I, L o V

<220>  
<221> RASGO\_MISC  
<222> (219)..(219)  
<223> el residuo es uno de H, K o R  
5  
<220>  
<221> RASGO\_MISC  
<222> (233)..(233)  
<223> el residuo es uno de C, D, E, H, K, N, Q, R, S o T  
10  
<220>  
<221> RASGO\_MISC  
<222> (236)..(236)  
<223> el residuo es uno de S o T  
<220>  
<221> RASGO\_MISC  
15  
<222> (239)..(239)  
<223> el residuo es uno de I, L o V  
<220>  
<221> RASGO\_MISC  
<222> (253)..(253)  
20  
<223> el residuo es uno de D o E  
<220>  
<221> RASGO\_MISC  
<222> (269)..(269)  
<223> el residuo es uno de I, L o V  
25  
<220>  
<221> RASGO\_MISC  
<222> (309)..(309)  
<223> el residuo es uno de A, C, D, G, N, P, S, T o V  
<220>  
30  
<221> RASGO\_MISC  
<222> (343)..(343)  
<223> el residuo es uno de I, L o V  
<220>  
<221> RASGO\_MISC  
35  
<222> (346)..(346)  
<223> el residuo es uno de H, K o R  
<220>  
<221> RASGO\_MISC  
<222> (355)..(355)  
40  
<223> el residuo es uno de S o T  
<220>  
<221> RASGO\_MISC

ES 2 640 747 T3

<222> (358)..(358)

<223> el residuo es uno de D o E

<220>

<221> RASGO\_MISC

5

<222> (359)..(359)

<223> el residuo es uno de I, L o V

<400> 2

```

Met Xaa Xaa Ser Val Xaa Glu Arg Thr Arg Ile Lys Asn Glu Arg Tyr
1      5      10      15
Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly Tyr Trp Asp Pro Asp
      20      25      30
Tyr Val Val Lys Asp Thr Asp Val Leu Ala Leu Phe Arg Val Thr Pro
      35      40      45
Gln Pro Gly Val Asp Pro Val Glu Ala Ser Ala Ala Val Ala Gly Glu
      50      55      60
Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr Asp Leu Leu Thr Ala
65      70      75      80
Cys Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val Asp Xaa Val Pro Asn
      85      90      95
Xaa Ser Asp Gln Tyr Phe Ala Tyr Ile Ala Tyr Asp Ile Asp Leu Phe
      100     105     110
Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser Ile Ile Gly Asn Val

```

ES 2 640 747 T3

115 120 125  
 Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu Glu Asp Met Arg Leu  
 130 135 140  
 Pro Val Ala Tyr Leu Lys Thr Phe Gln Gly Pro Ala Thr Gly Xaa Ile  
 145 150 155 160  
 Val Glu Arg Glu Arg Met Asp Lys Phe Gly Arg Pro Phe Leu Gly Ala  
 165 170 175  
 Thr Val Lys Pro Lys Leu Gly Leu Ser Gly Lys Asn Tyr Gly Arg Val  
 180 185 190  
 Val Tyr Glu Gly Leu Lys Gly Gly Leu Asp Phe Leu Lys Asp Asp Glu  
 195 200 205  
 Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Xaa Glu Arg Tyr Leu Tyr  
 210 215 220  
 Ser Met Glu Gly Val Asn Arg Ala Xaa Ala Ala Xaa Gly Glu Xaa Lys  
 225 230 235 240  
 Gly His Tyr Leu Asn Val Thr Ala Ala Thr Met Glu Xaa Met Tyr Glu  
 245 250 255  
 Arg Ala Glu Phe Ala Lys Glu Leu Gly Ser Val Ile Xaa Met Ile Asp  
 260 265 270  
 Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met Ala Ile Trp Ala Arg  
 275 280 285  
 Lys Asn Asp Met Ile Leu His Leu His Arg Ala Gly Asn Ser Thr Tyr  
 290 295 300  
 Ser Arg Gln Lys Xaa His Gly Met Asn Phe Arg Val Ile Cys Lys Trp  
 305 310 315 320  
 Met Arg Met Ala Gly Val Asp His Ile His Ala Gly Thr Val Val Gly  
 325 330 335  
 Lys Leu Glu Gly Asp Pro Xaa Met Ile Xaa Gly Phe Tyr Asn Thr Leu  
 340 345 350  
 Leu Leu Xaa His Leu Xaa Xaa Asn Leu Pro Gln Gly Ile Phe Phe Glu  
 355 360 365  
 Gln Asp Trp Ala Ser Leu Arg Lys Val Thr Pro Val Ala Ser Gly Gly  
 370 375 380  
 Ile His Cys Gly Gln Met His Gln Leu Leu Asp Tyr Leu Gly Asp Asp  
 385 390 395 400  
 Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly His Pro Asp Gly Ile  
 405 410 415  
 Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ser Met Val Ile  
 420 425 430  
 Ala Arg Asn Glu Gly Arg Asp Tyr Val Ala Glu Gly Pro Gln Ile Leu  
 435 440 445  
 Arg Asp Ala Ala Lys Thr Cys Gly Pro Leu Gln Thr Ala Leu Asp Leu  
 450 455 460  
 Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr Asp Thr Ala Asp Phe  
 465 470 475 480  
 Val Glu Thr Pro Thr Ala Asn Val  
 485

<210> 3

<211> 476

<212> PRT

5 <213> Artificial

<220>

<223> secuencia consenso del 50% para Rubisco de cianobacteria

<220>

<221> RASGO\_MISC

10 <222> (6) .. (6)

<223> El residuo es uno de A, C, F, G, H, I, K, L, M, R, T, V, W o Y

<220>

<221> RASGO\_MISC

<222> (8) .. (8)

15 <223> El residuo es uno de S o T

<220>

<221> RASGO\_MISC  
<222> (10) .. (10)  
<223> El residuo es uno de S o T  
<220>  
5 <221> RASGO\_MISC  
<222> (12) .. (12)  
<223> El residuo es uno de A, G o S  
<220>  
10 <221> RASGO\_MISC  
<222> (15) .. (15)  
<223> El residuo es uno de C, D, E, H, K, N, Q, R, S o T  
<220>  
<221> RASGO\_MISC  
<222> (43) .. (43)  
15 <223> El residuo es uno de A, C, F, G, H, I, K, L, M, R, T, V, W o Y  
<220>  
<221> RASGO\_MISC  
<222> (51) .. (51)  
<223> El residuo es uno de A, C, D, G, N, P, S, T o V  
20 <220>  
<221> RASGO\_MISC  
<222> (99) .. (99)  
<223> El residuo es uno de F, H, W o Y  
<220>  
25 <221> RASGO\_MISC  
<222> (145)..(145)  
<223> El residuo es uno de F, H, W o Y  
<220>  
<221> RASGO\_MISC  
30 <222> (227)..(227)  
<223> El residuo es uno de D, E, H, K o R  
<220>  
<221> RASGO\_MISC  
<222> (246)..(246)  
35 <223> El residuo es uno de A, C, D, G, N, P, S, T o V  
<220>  
<221> RASGO\_MISC  
<222> (252)..(252)  
<223> El residuo es uno de A, C, F, G, H, I, K, L, M, R, T, V, W o Y  
40 <220>  
<221> RASGO\_MISC  
<222> (263)..(263)

<223> El residuo es uno de C, D, E, H, K, N, Q, R, S o T  
<220>  
<221> RASGO\_MISC  
<222> (290)..(290)  
5 <223> El residuo es uno de I, L o V  
<220>  
<221> RASGO\_MISC  
<222> (341)..(341)  
<223> El residuo es uno de A, G o S  
10 <220>  
<221> RASGO\_MISC  
<222> (353)..(353)  
<223> El residuo es uno de D o E  
<220>  
15 <221> RASGO\_MISC  
<222> (354)..(354)  
<223> El residuo es uno de F, H, W o Y  
<220>  
<221> RASGO\_MISC  
20 <222> (357) .. (357)  
<223> El residuo es uno de D, E, H, K o R  
<220>  
<221> RASGO\_MISC  
<222> (437)..(437)  
25 <223> El residuo es uno de D, E, H, K, R  
<220>  
<221> RASGO\_MISC  
<222> (439)..(439)  
<223> El residuo es uno de A, C, D, E, G, H, K, N, Q, R, S, T  
30 <220>  
<221> RASGO\_MISC  
<222> (443)..(443)  
<223> El residuo es uno de A, C, D, G, N, P, S, T o V  
<220>  
35 <221> RASGO\_MISC  
<222> (458) .. (458)  
<223> El residuo es uno de I, L o V  
<400> 3

ES 2 640 747 T3

<400> 3

Met Ser Tyr Ala Gln Xaa Lys Xaa Gln Xaa Lys Xaa Gly Tyr Xaa Ala  
 1 5 10 15  
 Gly Val Lys Asp Tyr Arg Leu Thr Tyr Thr Pro Asp Tyr Thr Pro  
 20 25 30  
 Lys Asp Thr Asp Leu Leu Ala Ala Phe Arg Xaa Thr Pro Gln Pro Gly  
 35 40 45  
 Val Pro Xaa Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr  
 50 55 60  
 Gly Thr Trp Thr Thr Val Trp Thr Asp Leu Leu Thr Asp Leu Asp Arg  
 65 70 75 80  
 Tyr Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Asp Asn  
 85 90 95  
 Gln Tyr Xaa Ala Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly  
 100 105 110  
 Ser Val Thr Asn Val Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe  
 115 120 125  
 Lys Ala Leu Arg Ala Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala  
 130 135 140  
 Xaa Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg  
 145 150 155 160  
 Asp Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys  
 165 170 175  
 Pro Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu  
 180 185 190  
 Cys Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn  
 195 200 205  
 Ser Gln Pro Phe Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu  
 210 215 220  
 Ala Ile Xaa Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr  
 225 230 235 240  
 Leu Asn Val Thr Ala Xaa Thr Cys Glu Glu Met Xaa Lys Arg Ala Glu  
 245 250 255  
 Phe Ala Lys Glu Leu Gly Xaa Pro Ile Met His Asp Phe Leu Thr  
 260 265 270  
 Gly Gly Phe Thr Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn  
 275 280 285  
 Gly Xaa Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg  
 290 295 300  
 Gln Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg  
 305 310 315 320  
 Leu Ser Gly Gly Asp His Leu His Thr Gly Thr Val Val Gly Lys Leu  
 325 330 335  
 Glu Gly Glu Arg Xaa Ile Thr Met Gly Phe Val Asp Leu Leu Arg Glu  
 340 345 350  
 Xaa Xaa Val Glu Xaa Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp  
 355 360 365  
 Trp Ala Ser Met Pro Gly Val Met Ala Val Ala Ser Gly Gly Ile His  
 370 375 380  
 Val Trp His Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val  
 385 390 395 400  
 Leu Gln Phe Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro  
 405 410 415  
 Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg  
 420 425 430  
 Asn Glu Gly Arg Xaa Leu Xaa Arg Glu Gly Xaa Asp Ile Ile Arg Glu  
 435 440 445  
 Ala Ala Lys Trp Ser Pro Glu Leu Ala Xaa Ala Cys Glu Leu Trp Lys  
 450 455 460  
 Glu Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu  
 465 470 475

<210> 4

5

<211> 475

<212> PRT

<213> Anthoceros formosae

<400> 4

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110

Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala  
 210 215 220  
 Ile Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala His Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Met Pro Ile Val Met His Asp Tyr Leu Thr Gly

260 265 270  
 Gly Phe Thr Ala Asn Thr Thr Leu Ala Arg Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Arg Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Ser Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Asp Ile Ile Arg Glu Ala  
 435 440 445  
 Ser Lys Trp Ser Pro Glu Leu Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Val Phe Glu Thr Ile Asp Thr Leu  
 465 470 475

<210> 5  
 <211> 475  
 <212> PRT  
 <213> Artificial  
 <220>

# ES 2 640 747 T3

<223> secuencia consenso del 50% para bryophyta (28 especies)

<400> 5

```

Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Val Gly Phe Lys Ala Gly
1      5      10      15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Gln Thr Lys
20      25      30
Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
35      40      45
Pro Ala Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
50      55      60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65      70      75      80
Lys Gly Arg Cys Tyr Asp Ile Glu Ala Val Pro Gly Glu Glu Asn Gln
85      90      95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
100     105     110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
115     120     125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
130     135     140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145     150     155     160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
165     170     175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
180     185     190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
195     200     205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
210     215     220
Ile Tyr Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225     230     235     240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Gln Phe
245     250     255
Ala Arg Glu Leu Gly Met Pro Ile Val Met His Asp Tyr Leu Thr Gly
260     265     270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
275     280     285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
290     295     300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu
305     310     315     320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
325     330     335
Gly Glu Arg Gln Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
340     345     350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
355     360     365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
370     375     380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385     390     395     400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
405     410     415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
420     425     430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala
435     440     445
Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
450     455     460
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Val
465     470     475

```

5

<210> 6

<211> 475

<212> PRT

<213> Artificial

<220>

ES 2 640 747 T3

<223> secuencia consenso del 50% para carofita (4 especies)

<400> 6

Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Ala Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
 20 25 30  
 Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Ala Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75  
 Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Ile Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala  
 210 215 220  
 Ile Phe Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Ala Tyr  
 245 250 255  
 Ala Lys Glu Leu Gly Val Pro Ile Ile Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Asp Val Ile Arg Glu Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Asp Thr Ile Asp Thr Leu  
 465 470 475

5

<210> 7

<211> 475

<212> PRT

<213> Artificial

<220>

# ES 2 640 747 T3

<223> secuencia consenso del 50% para clorofita (4 especies)

<220>

<221> RASGO\_MISC

<222> (282) .. (282)

5

<223> El residuo es uno de A, C, D, G, N, P, S, T o V

<400> 7

<400> 7

```

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

```

ES 2 640 747 T3

Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Gly Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala  
 210 215 220  
 Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Val Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Ala Thr Cys Glu Glu Met Leu Lys Arg Ala Val Cys  
 245 250 255  
 Ala Lys Glu Leu Gly Val Pro Ile Ile Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala Xaa Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Arg Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu  
 305 310 315 320  
 Ser Gly Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Met Arg Asp Asp  
 340 345 350  
 Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365

Cys Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ala Cys Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Thr Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Gly Asp Val Ile Arg Ser Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Asp Thr Ile Asp Lys Leu  
 465 470 475

<210> 8

<211> 475

5

<212> PRT

<213> Pinus thunbergii

<400> 8

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

ES 2 640 747 T3

130 135 140  
 Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala  
 210 215 220  
 Leu Asn Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Ile Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Arg Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Asp Val Ile Asp Arg Leu  
 465 470 475

<210> 9

<211> 475

<212> PRT

5

<213> Equisetum arvense

<400> 9

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

ES 2 640 747 T3

130 135 140  
 Ser Lys Thr Phe Ile Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala  
 210 215 220  
 Leu Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Ala Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala Phe Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Thr Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Leu Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Val Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Thr Glu Gly Asn Asp Ile Ile Arg Glu Ala  
 435 440 445  
 Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Tyr Glu Ala Met Asp Thr Leu  
 465 470 475

<210> 10

<211> 475

<212> PRT

5

<213> *Welwitschia mirabilis*

<400> 10

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

# ES 2 640 747 T3

Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Thr Ser Tyr  
 130 135 140  
 Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Thr Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Asp Ile Phe Gly Asp Asp Ala Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala  
 435 440 445  
 Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Glu Ser Val Asp Thr Leu  
 465 470 475

<210> 11

<211> 475

5 <212> PRT

<213> Artificial

<220>

<223> secuencia consenso del 50% para magnoliofita (134 especies)

<220>

10 <221> RASGO\_MISC

<222> (142) .. (142)

<223> El residuo es uno de A, C, D, G, N, P, S, T o V

<220>

<221> RASGO\_MISC

15 <222> (251)..(251)

<223> El residuo es uno de I, L o V

<400> 11

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Glu Glu Asn Gln  
 85 90 95  
 Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Xaa Ala Tyr  
 130 135 140  
 Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Xaa Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala  
 435 440 445  
 Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu  
 465 470 475

<210> 12

<211> 475

5

<212> PRT

<213> Artificial

<220>

<223> secuencia consenso del 50% para pinofita (2 especies)

<400> 12

ES 2 640 747 T3

Met Ser Pro Lys Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Ala Glu Glu Ala Gly Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Glu Asn Gln  
 85 90 95  
 Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Asn Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Ile Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Asp Ala Ile Asp Arg Leu  
 465 470 475

<210> 13

<211> 475

<212> PRT

5

<213> Artificial

<220>

<223> secuencia consenso del 50% para pteridofita (2 especies)

<400> 13

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Thr Gly Ile Gly Phe Lys Ala Gly  
1 5 10 15  
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
20 25 30  
Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val  
35 40 45  
Pro Ala Glu Glu Ala Gly Ala Val Ala Ala Glu Ser Ser Thr Gly  
50 55 60  
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
65 70 75 80  
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln  
85 90 95  
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
100 105 110  
Val Thr Asn Met Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
115 120 125  
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Leu Pro Ala Tyr  
130 135 140  
Ser Lys Thr Phe Ile Gly Pro His Gly Ile Gln Val Glu Arg Asp  
145 150 155 160  
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
165 170 175  
Lys Leu Gly Leu Cys Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
180 185 190  
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
195 200 205  
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala  
210 215 220  
Leu Phe Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
225 230 235 240  
Asn Ala Thr Ala Ala Thr Cys Glu Glu Met Ile Lys Arg Ala Ile Phe  
245 250 255  
Ala Arg Glu Leu Gly Ala Pro Ile Val Met His Asp Tyr Leu Thr Gly  
260 265 270  
Gly Phe Thr Ala Asn Thr Ser Leu Ala Phe Tyr Cys Arg Asp Asn Gly  
275 280 285  
Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
290 295 300  
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
305 310 315 320  
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
325 330 335  
Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
340 345 350  
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
355 360 365  
Val Ser Met Pro Gly Val Phe Pro Val Ala Ser Gly Gly Ile His Val  
370 375 380  
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
385 390 395 400  
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
405 410 415  
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
420 425 430  
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala  
435 440 445  
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
450 455 460  
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Val  
465 470 475

<210> 14

<211> 493

<212> PRT

5

<213> Galdieria partita

<400> 14

ES 2 640 747 T3

<400> 14

Met Ser Gln Ser Ile Glu Glu Lys Ser Val Gln Glu Arg Thr Arg Ile  
1 5 10 15  
Lys Asn Ser Arg Tyr Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly  
20 25 30  
Tyr Trp Asn Pro Asp Tyr Gln Val Lys Asp Thr Asp Val Leu Ala Leu  
35 40 45  
Phe Arg Val Thr Pro Gln Pro Gly Val Asp Pro Ile Glu Ala Ala Ala  
50 55 60  
Ala Val Ala Gly Glu Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr  
65 70 75 80  
Asp Leu Leu Thr Ala Ala Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val  
85 90 95  
Asp Gln Val Pro Asn Asn Pro Glu Gln Tyr Phe Ala Tyr Ile Ala Tyr  
100 105 110  
Glu Leu Asp Leu Phe Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser  
115 120 125  
Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu  
130 135 140  
Glu Asp Met Arg Leu Pro Leu Ala Tyr Leu Lys Thr Phe Gln Gly Pro  
145 150 155 160  
Ala Thr Gly Val Ile Leu Glu Arg Glu Arg Leu Asp Lys Phe Gly Arg  
165 170 175  
Pro Leu Leu Gly Cys Thr Thr Lys Pro Lys Leu Gly Leu Ser Gly Lys  
180 185 190  
Asn Tyr Gly Arg Val Val Tyr Glu Ala Leu Lys Gly Gly Leu Asp Phe  
195 200 205  
Val Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Arg  
210 215 220  
Glu Arg Tyr Leu Phe Thr Met Glu Ala Val Asn Lys Ala Ser Ala Ala  
225 230 235 240  
Thr Gly Glu Val Lys Gly His Tyr Leu Asn Val Thr Ala Ala Thr Met  
245 250 255  
Glu Glu Met Tyr Ala Arg Ala Asn Phe Ala Lys Glu Leu Gly Ser Val  
260 265 270  
Ile Ile Met Ile Asp Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met  
275 280 285  
Ala Lys Trp Ala Arg Asp Asn Asp Met Ile Leu His Leu His Arg Ala  
290 295 300  
Gly Asn Ser Thr Tyr Ser Arg Gln Lys Asn His Gly Met Asn Phe Arg  
305 310 315 320  
Val Ile Cys Lys Trp Met Arg Met Ala Gly Val Asp His Ile His Ala  
325 330 335  
Gly Thr Val Val Gly Lys Leu Glu Gly Asp Pro Ile Ile Thr Arg Gly  
340 345 350  
Phe Tyr Lys Thr Leu Leu Leu Pro Lys Leu Glu Arg Asn Leu Gln Glu  
355 360 365  
Gly Leu Phe Phe Asp Met Glu Trp Ala Ser Leu Arg Lys Val Met Pro  
370 375 380  
Val Ala Ser Gly Gly Ile His Ala Gly Gln Met His Gln Leu Ile His  
385 390 395 400  
Tyr Leu Gly Glu Asp Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly  
405 410 415  
His Pro Asp Gly Ile Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu  
420 425 430  
Glu Ala Met Ile Leu Ala Arg Asn Glu Asn Arg Asp Tyr Leu Thr Glu  
435 440 445  
Gly Pro Glu Ile Leu Arg Glu Ala Ala Lys Thr Cys Gly Ala Leu Arg  
450 455 460  
Thr Ala Leu Asp Leu Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr  
465 470 475 480  
Asp Thr Ser Asp Phe Val Glu Thr Pro Thr Ala Asn Ile

485 490

5

<210> 15

<211> 488

<212> PRT

<213> Griffithsia monolis

ES 2 640 747 T3

<400> 15

Met Ser Asn Ser Val Glu Glu Arg Thr Arg Ile Lys Asn Glu Arg Tyr  
 1 5 10 15  
 Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly Tyr Trp Asp Pro Asn  
 20 25 30  
 Tyr Ala Val Lys Asp Thr Asp Ile Leu Ala Leu Phe Arg Val Ser Pro  
 35 40 45  
 Gln Pro Gly Val Asp Pro Val Glu Ala Ser Ala Ala Val Ala Gly Glu  
 50 55 60  
 Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr Asp Leu Leu Thr Ala  
 65 70 75 80  
 Cys Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val Glu Ser Val Pro Asn  
 85 90 95  
 Thr Ser Asp Gln Tyr Phe Ala Tyr Ile Ser Tyr Asp Ile Asp Leu Phe  
 100 105 110  
 Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser Ile Ile Gly Asn Val  
 115 120 125  
 Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu Glu Asp Met Arg Ile  
 130 135 140  
 Pro Val Ala Tyr Leu Lys Thr Phe Gln Gly Pro Ala Thr Gly Ile Val  
 145 150 155 160  
 Val Glu Arg Glu Arg Met Asp Lys Phe Gly Arg Pro Phe Leu Gly Ala  
 165 170 175  
 Thr Val Lys Pro Lys Leu Gly Leu Ser Gly Lys Asn Tyr Gly Arg Val  
 180 185 190  
 Val Tyr Glu Gly Leu Arg Gly Gly Leu Asp Phe Leu Lys Asp Asp Glu  
 195 200 205  
 Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Lys Glu Arg Phe Leu Tyr  
 210 215 220  
 Ser Ile Glu Ala Val Asn Arg Ser Ile Ala Ala Thr Gly Glu Val Lys  
 225 230 235 240  
 Gly His Tyr Met Asn Val Thr Ala Ala Thr Met Glu Glu Met Tyr Glu  
 245 250 255  
 Arg Ala Glu Phe Ala Lys Gln Leu Gly Thr Val Ile Ile Met Ile Asp  
 260 265 270  
 Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met Gly Ile Trp Ala Arg  
 275 280 285  
 Lys Asn Asp Met Ile Leu His Leu His Arg Ala Gly Asn Ser Thr Tyr  
 290 295 300  
 Ser Arg Gln Lys Ile His Gly Met Asn Phe Arg Val Ile Cys Lys Trp  
 305 310 315 320  
 Met Arg Met Ala Gly Val Asp His Ile His Ala Gly Thr Val Val Gly  
 325 330 335  
 Lys Leu Glu Gly Asp Pro Leu Met Ile Arg Gly Phe Tyr Asn Thr Leu  
 340 345 350  
 Leu Leu Pro Tyr Leu Glu Val Asn Leu Pro Gln Gly Ile Phe Phe Gln  
 355 360 365  
 Gln Asp Trp Ala Ser Leu Arg Lys Val Thr Pro Val Ala Ser Gly Gly  
 370 375 380  
 Ile His Cys Gly Gln Met His Gln Leu Leu Asp Tyr Leu Gly Asn Asp  
 385 390 395 400  
 Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly His Pro Asp Gly Ile  
 405 410 415  
 Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ser Met Val Ile  
 420 425 430  
 Ala Arg Asn Glu Gly Arg Asp Tyr Val Ala Glu Gly Pro Gln Ile Leu  
 435 440 445  
 Arg Asp Ala Ala Lys Thr Cys Gly Pro Leu Gln Thr Ala Leu Asp Leu  
 450 455 460  
 Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr Asp Thr Ala Asp Phe  
 465 470 475 480  
 Val Glu Thr Pro Thr Ala Asn Val  
 485

<210> 16

<211> 472

<212> PRT

<213> Synechococcus elongatus PCC6301

ES 2 640 747 T3

<400> 16

Met Pro Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp  
 1 5 10 15  
 Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp  
 20 25 30  
 Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp  
 35 40 45  
 Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
 50 55 60  
 Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys  
 65 70 75 80  
 Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 17

<211> 475

<212> PRT

<213> Spinacia oleracea

<400> 17

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Leu  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Ser Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Asn Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Glu Glu Asn Gln  
 85 90 95  
 Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Val Ala Tyr  
 130 135 140  
 Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Asp Met Met Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Thr Leu Ser His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu  
 305 310 315 320

Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Tyr Thr Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Ser Trp  
 355 360 365  
 Val Ser Thr Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Thr Ile Ile Arg Glu Ala  
 435 440 445  
 Thr Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Pro Ala Met Asp Thr Val  
 465 470 475

<210> 18

<211> 475

<212> PRT

<213> Nicotiana tabacum

<400> 18

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Asp Glu Asn Gln  
 85 90 95  
 Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Ala Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Ile Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala  
 435 440 445  
 Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu  
 465 470 475

<210> 19

<211> 477

5

<212> PRT

<213> Oriza sativa

<400> 19

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr His Ile Glu Pro Val Gly Glu Asp Asn Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Thr Tyr  
 130 135 140  
 Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Cys Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Tyr Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Met Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Ile Glu Lys Asp Arg Ala Arg Gly Ile Phe Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Met Pro Gly Val Ile Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Ser Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Ala  
 450 455 460  
 Ile Lys Phe Glu Phe Glu Pro Val Asp Lys Leu Asp Ser  
 465 470 475

<210> 20

<211> 475

<212> PRT

<213> Glycine max

<400> 20

5

ES 2 640 747 T3

<400> 20

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75  
 Lys Gly Arg Cys Tyr Gly Leu Glu Pro Val Ala Gly Glu Glu Asn Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Thr Ala Tyr  
 130 135 140  
 Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Ile Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu  
 305 310 315  
 Ser Gly Gly Asp His Val His Ala Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Val Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala  
 435 440 445  
 Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu  
 465 470 475

<210> 21

<211> 476

<212> PRT

<213> Saccharum officinarum

<400> 21

5

ES 2 640 747 T3

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
1 5 10 15  
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Lys  
20 25 30  
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Leu Gly Val  
35 40 45  
Pro Pro Glu Glu Ala Gly Ala Val Ala Ala Glu Ser Ser Thr Gly  
50 55 60  
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
65 70 75 80  
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Pro Gly Asp Pro Asp Gln  
85 90 95  
Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
100 105 110  
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
115 120 125  
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
130 135 140  
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
145 150 155 160  
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
165 170 175  
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Cys Tyr Glu Cys  
180 185 190  
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
195 200 205  
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala  
210 215 220  
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
225 230 235 240  
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
245 250 255  
Ala Lys Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
260 265 270  
Gly Phe Thr Ala Asn Thr Thr Leu Ser His Tyr Cys Arg Asp Asn Gly  
275 280 285  
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
290 295 300  
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
305 310 315 320  
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
325 330 335  
Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
340 345 350  
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp Trp  
355 360 365  
Val Ser Met Pro Gly Val Ile Pro Val Ala Ser Gly Gly Ile His Val  
370 375 380  
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
385 390 395 400  
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
405 410 415  
Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn  
420 425 430  
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Lys Ala Ala  
435 440 445  
Cys Lys Trp Ser Ala Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu  
450 455 460  
Ile Lys Phe Asp Thr Phe Lys Ala Met Asp Thr Leu  
465 470 475

<210> 22

<211> 1845

5

<212> ADN

<213> Synechococcus PCC6301

<400> 22

ES 2 640 747 T3

<400> 22

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc      180
ggtaacctga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag      240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac      300
ccgctcgacc tgtttgaaga agggctcggtc accaacatcc tgacctgat cgtcggtaac      360
gtggttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctcgc      420
ttggtcaaaa ccttccaagg tctctcccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtccggcгаа      540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttctggt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcaactacc gaacgtgacc      720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg      1020
ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggcttc      1080
ttcaccnaag attggcgctc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcggg      1200
ggcggcacct tgggtcacc ctagggtaat gctcctggtg caaccgcгаа cgtggtgcc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac      1320
atccttctgt aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctgгаа      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctctgtac      1560
ctgcctcccc tcagcgatgc ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc      1620
ttcaccacct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaaagctc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgctgtgag      1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtcg gcttcgacaa catcaagcag      1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

<210> 23

<211> 111

5

<212> PRT

<213> Synechococcus PCC6301

<400> 23

```

Met Ser Met Lys Thr Leu Pro Lys Glu Arg Arg Phe Glu Thr Phe Ser
1          5          10          15
Tyr Leu Pro Pro Leu Ser Asp Arg Gln Ile Ala Ala Gln Ile Glu Tyr
20          25          30
Met Ile Glu Gln Gly Phe His Pro Leu Ile Glu Phe Asn Glu His Ser
35          40          45
Asn Pro Glu Glu Phe Tyr Trp Thr Met Trp Lys Leu Pro Leu Phe Asp
50          55          60
Cys Lys Ser Pro Gln Gln Val Leu Asp Glu Val Arg Glu Cys Arg Ser
65          70          75          80
Glu Tyr Gly Asp Cys Tyr Ile Arg Val Ala Gly Phe Asp Asn Ile Lys
85          90          95
Gln Cys Gln Thr Val Ser Phe Ile Val His Arg Pro Gly Arg Tyr
100         105         110

```

<210> 24

10

<211> 472

<212> PRT

<213> Synechococcus PCC6301

<400> 24

ES 2 640 747 T3

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp  
 1 5 10 15  
 Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp  
 20 25 30  
 Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp  
 35 40 45  
 Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
 50 55 60  
 Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys  
 65 70 75 80  
 Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 25

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 25

ES 2 640 747 T3

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattggacc  ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc     120
cctcagccgg gtgtccctgc tatcgaagct ggtgcccgga tcgcccgtga atcttcgacc     180
ggtaacctga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag     240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcggt catcgcttac     300
ccgctcgacc tgtttgaaga agggctcggtc accaacatcc tgacctcgat cgtcggtaac     360
gtggttggt  tcaaagctat ccgttcgtg  cgtctggaag acatccgctt ccccgctgcc     420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac     480
aagtaaggcc gtccgatgct gggttgacg  atcaaacca aactcggctt gtcggcgaaa     540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac     600
gaaaacatca actcgcagcc gttccaacgc tggcgcgac  gcttctgtt  tgtggctgat     660
gcaatccaca aatcgcaagc agaaaaccgt gaaatcaaag gtcactacct gaacgtgacc     720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg     780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg     840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt     900
cagcgtaac  acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt     960
gaccacctcc actccgacac cgctcgtcggc aaactggaag gcgacaaaag ttcgaccttg    1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg acccgagccg tggggctctc    1080
ttaccccaag atgggcgctc gatgcccggc gtgctgccgg ttgcttccgg tggatccac    1140
gtgtggcaca tgcccgcact ggtggaatc  ttcggtgat  actccgttct ccagttcggg    1200
ggcggcacct tgggtcaacc ctggggtaat gctcctggtg caaccgcaa  ccggttgcc    1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga agggggcgac    1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa    1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg    1440
gggagtgagc gttgctcgtt aaagctttct ccccagcctt tcgacttaac ctttcaggat    1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac    1560
ctgctcccc  tcagcgatcg ccaaatcgtc gcacaaatcg agtacatgat cgagcaaggc    1620
ttccaccct  tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg    1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtagg    1740
tgccgcagcg aatacgggta ttgctacatc cgtgtcgtg  gtttcgaaa  catcaagcag    1800
tgccaaccg  tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

<210> 26

<211> 472

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 26

```

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

```

5

ES 2 640 747 T3

Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 27

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 27

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaactcacc	60
tattggacc	cgattacac	ccccaaagac	actgacctgc	tggcggcttt	cgcttcagc	120
cctcagccg	gtgtccctgc	tgtcgaagct	ggtgcggcga	tcgcggtga	atcttcgacc	180
ggtacctgga	ccaccgtgtg	gaccgacttg	ctgaccgaca	tggatcggta	caaaggcaag	240
tgctaccaca	tcgagccggt	gcaagggcga	gagaactcct	actttgcggt	catcgcttac	300
ccgctcgacc	tgtttgaaga	agggtcggtc	accaacatcc	tgacctcgat	cgctcggtaac	360
gtgtttggct	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttggtcaaaa	ccttccaag	tcctccccac	ggtatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgacg	atcaaaccaa	aactcggctt	gtcggcgaaa	540
aactacggtc	gtgcccgtcta	cgaatgtctg	cgcgccggtc	tggacttcac	caaagacgac	600
gaaaacatca	actcgcagcc	gttccaacgc	tggcgcgac	gcttcctggt	tgtggctgat	660
gcaatccaca	aatcgcaagc	agaaacgggt	gaaatcaaag	gtcactacct	gaacgtgacc	720
gcgcccacct	gcgaagaagt	gatgaaacgg	gctgagttcg	ctaaagaact	cggcatgccg	780

ES 2 640 747 T3

```

atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaag ttcgaccttg     1020
ggctttgttg acttgatgcy cgaagaccac atcgaagctg accgcagccg tggggtcttc     1080
ttcacccaag attgggcgct gatgccgggc gtgctgccgg ttgcttccgg tggtatccac     1140
gtgtggcaca tgcgccact ggtggaatc ttcggtgatg actccgttct ccagttcggt      1200
ggcggcacct tgggtcacc ctggggtaat gctcctggtg caaccgcgaa cgtggtgccc     1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac     1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa     1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gaggcctctga ctatcgctgg     1440
gggagtgagc gttgctgcyt aaagcttctt ccccagcctt tcgacttaac ctttcaggat     1500
tttgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac     1560
ctgctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc     1620
ttccaccctt tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg     1680
tggaagctcc cctggttga ctgcaagagc cctcagcaag tcctcagatga agtgctgag     1740
tgcccgagcg aatacggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag     1800
tgcaaacccg tgagcttcat cgttcatcgt cccggccgat actaa                       1845

```

<210> 28

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 28

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Val
          35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65          70          75          80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
          85          90          95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
          100          105          110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
          115          120          125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
          130          135          140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
          145          150          155          160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
          165          170          175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
          180          185          190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
          195          200          205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
          210          215          220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
          225          230          235          240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
          245          250          255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
          260          265          270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
          275          280          285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
          290          295          300

```

ES 2 640 747 T3

Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 29

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 29

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc 60  
 tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120  
 cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc 180  
 ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcccgaag 240  
 tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcggt catcgcctac 300  
 ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctgat cgtcggtaac 360  
 gtgtttggct tcaaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctgcc 420  
 ttggtaaaaa ccttccaagg tcctcccac ggtatccaag tcgagcgcga cctgctgaac 480  
 aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa 540  
 aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600  
 gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctggt ttgtgctgat 660  
 gcaatccaca aatcgaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720  
 ggcgcgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780  
 atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaatgg 840  
 tgcccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900  
 cagcgtaac acgggattca ctccgtgctc ttggccaagt gtttgcgctc gtcgggtggt 960  
 gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaag ttcgacctg 1020  
 ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080  
 ttcacccaag attggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggatccac 1140  
 gtgtggcaca tgcccgaact ggtgaaaatc ttcggtgatg actccgttct ccagtccgtt 1200  
 ggcgcgacct tgggtcacc cttgggtaat gctcctgggt caaccgcgaa ccgtgttgcc 1260  
 ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320  
 atcctctgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctgaaa 1380  
 gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440  
 gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500  
 ttctgaaatc tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560  
 ctgctcccc tcagcgcgac ccaaatcgtc gcacaaatcg agtacatgat cgagcaaggc 1620  
 ttccacccct tgatcagatt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680  
 tggaaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgctgag 1740  
 tgcccgcagc aatacggtag ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800  
 tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

10

<210> 30

<211> 472

<212> PRT

<213> Artificial

ES 2 640 747 T3

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 30

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1      5      10      15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20      25      30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
35      40      45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50      55      60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65      70      75      80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
85      90      95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
100     105
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
115     120     125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
130     135     140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145     150     155     160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
165     170     175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
180     185     190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
195     200     205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210     215     220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225     230     235     240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
245     250     255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
260     265     270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
275     280     285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
290     295     300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305     310     315     320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
325     330     335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
340     345     350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
355     360     365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
370     375     380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385     390     395     400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
405     410     415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
420     425     430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
435     440     445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
450     455     460
Glu Phe Glu Thr Met Asp Lys Leu
465     470

```

5

<210> 31

<211> 1845

<212> ADN

<213> Artificial

ES 2 640 747 T3

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 31

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt cgccttcagc      120
cctcagccgg gtgtccctgc tgacgaagct cggcggcgga tcgcggtga atcttcgacc      180
ggtaacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag      240
gcctaccacg tggagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac      300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac      360
gtgtttggct tcaaagctat cgttccgctg cgtctggaag acatccgctt ccccgctgcc      420
ttggtcaaaa ctttccaagg tcttccccac ggtatccaag tcgagcgaga cctgctgaac      480
aagtacggcc gtccgatgct gggttgacg atcaaaccaa aactcggctt gtcggcgaaa      540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaacatca actcgagacc gttccaacgc tggcgcgata gttcctctgt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc      720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaac ttcgaccttg      1020
ggctttgttg acttgatgcy cgaagaccac atcgaagctg accgcagccg tggggtcttc      1080
ttcaccacaag attggggcgc gatgcccggc gtgctgccgg ttgcttccgg tggatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcggg      1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac      1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctgcyt aaagctttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac      1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacctgat cgagcaaggc      1620
ttcaccacct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag      1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtcg gcttcgacaa catcaagcag      1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845
    
```

5

<210> 32

<211> 472

<212> PRT

<213> Artificial

<220>

10

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 32

<400> 32

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
 1                    5                    10                    15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
                20                    25                    30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
                35                    40                    45
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
 50                    55                    60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
 65                    70                    75                    80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                    90                    95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
 100                    105                    110
    
```

ES 2 640 747 T3

Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 33

<211> 1845

5

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 33

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaactcacc	60
tattacaccc	ccgattacac	ccccaaagac	actgacctgc	tggcggcttt	ccgcttcagc	120
cctcagccgg	gtgtccctgc	tgacgaagct	agcgcggcga	tcgctggctga	atcttcgacc	180
ggtacctgga	ccaccgtgtg	gaccgacttg	ctgaccgaca	tgatcggtga	caaaggcaag	240
gcctaccacg	tgagccgggt	gcaaggcgaa	gagaactcct	actttgcgtt	catcgcttac	300
ccgctcgacc	tgtttgaaga	agggtcggtc	accaacatcc	tgacctgat	cgctcgtaac	360
gtgtttggct	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttggtcaaaa	ccttccaagg	tcctccccac	ggtatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgacag	atcaaaccac	aactcggctc	gtcggcgaaa	540
aactacggtc	gtgccgtcta	cgaatgtctg	cgccggcggtc	tgacttcac	caaagacgac	600
gaaaaatca	actcgcagcc	gttccaacgc	tggcgcgatc	gcttcctggt	tgtggctgat	660

10

# ES 2 640 747 T3

```

gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaagt gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgctcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt 960
gaccacctcc actccggcac cgctcgtcggc aaactggaag gcgacaaaagc ttcgaccttg 1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg acccgagccg tggggctctc 1080
ttcacccaag attggggcgtc gatgcccggc gtgctgcccg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcgccact ggtggaatc ttcggtgatg actccgttct ccagttcggg 1200
ggcggcacct tgggtcacc cttgggtaat gctcctggtg caaccgcaa cctggttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga agggggcgac 1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctcgtg aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgctctccc tcagcgatcg ccaaactcgt gcacaaatcg agtacatgat cgagcaaggc 1620
ttccaccctt tgatcgagtt caacgagcac tcgaaatccg aagagttcta ctggacgatg 1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgctgag 1740
tgccgcagcg aatacggtaga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800
tgcaaacccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

```

<210> 34

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 34

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20        25        30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
35        40        45
Glu Ala Ser Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50        55        60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65        70        75        80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala
85        90        95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
100       105       110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
115       120       125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
130       135       140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145       150       155       160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
165       170       175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
180       185       190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
195       200       205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210       215       220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225       230       235       240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
245       250       255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
260       265       270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
275       280       285

```

# ES 2 640 747 T3

His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 35

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 35

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaactcggc	60
tattggacc	ccgattacac	ccccaaagac	actgacctgc	tggcggcttt	cgcttcagc	120
cctcagccg	gtgtccctgc	tatcgaagct	ggtgcccgga	tcgcccgtga	atcttcgacc	180
ggtacctgga	ccaccgtgtg	gaccgacttg	ctgaccgaca	tggatcggta	ccgcccgaag	240
tgctaccaca	tcgagccggt	gcaaggcgaa	gagaactcct	actttgcggt	catcgcttac	300
ccgctcgacc	tgtttgaaga	agggtcggtc	accaacatcc	tgacctcgat	cgctcgtaac	360
gtggttggt	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttggtcaaaa	ccttccaagg	tcctccccac	ggtatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgcaag	atcaaaccaa	aactcggctt	gtcggcgaaa	540
aactacggct	gtgccgtcta	cgaatgtctg	cgccgcccgt	tggacttcac	caaagacgac	600
gaaaacatca	actcgcagcc	gttccaacgc	tgccgcgcat	gcttctctgt	tgtggctgat	660
gcaatccaca	aatcgcaagc	agaaaccggt	gaaatcaaag	gtcactacct	gaacgtgacc	720
gcccgcacct	gcaagaaat	gatgaaacgg	gctgagttcg	ctaaagaact	cggcatgccc	780
atcatcatgc	atgacttctt	gacggctggt	ttcacccgca	acaccacctt	ggcaaaatgg	840
tgcccgcgca	acggcgtcct	gctgcacatc	caccgtgcaa	tgacccgctt	gatcgaccgt	900
cagcgtaaac	acgggattca	cttccgtgct	ttggccaagt	gtttgcgtct	gtccggtggt	960
gaccacctcc	actccggcac	cgctcgtcgg	aaactggaag	gcgacaaagc	ttcgaccttg	1020
ggctttgttg	acttgatgcg	cgaagaccac	atcgaagctg	accgcagccc	tggggctctc	1080
ttcacccaag	attgggcgct	gatgccgggc	gtgctgcccg	ttgcttccgg	tggatccac	1140
gtgtggcaca	tgcccgcact	ggtggaatc	ttcggatgat	actccgttct	ccagttcggg	1200
ggcggcacct	tgggtcacc	ctggggtaat	gctcctggtg	caaccgcgaa	ccgtgttgcc	1260
ttggaagctt	gcgtccaagc	tcggaacgaa	ggtcgcgacc	tctaccgtga	aggcggcgac	1320
atccttcggt	aagctggcaa	gtggtcgcct	gaactggctg	ctgccctcga	cctctggaaa	1380
gagatcaagt	tcgaattcga	aacgatggac	aagctctaag	gagcctctga	ctatcgctgg	1440
gggagtgagc	gttgctgctg	aaagctttct	ccccagcctt	tcgacttaac	ctttcaggat	1500
ttctgaaatc	tgagcatgaa	aactctgccc	aaagagcgtc	gtttcgagac	tttctcgtac	1560
ctgcctcccc	tcagcgatcg	ccaatcgcct	gcacaaatcg	agtacatgat	cgagcaaggc	1620
ttccaccctt	tgatcgaagt	caacgagcac	tcgaatccgg	aagagttcta	ctggacgatg	1680
tggaaagctc	ccctgtttga	ctgcaagagc	cctcagcaag	tcctcgatga	agtgcgtgag	1740
tgcccgcagc	aatacggatg	ttgctacatc	cgtgtcgtg	gcttcgacaa	catcaagcag	1800
tgccaaaccg	tgagcttcat	cgttcatcgt	cccggccgat	actaa		1845

10

<210> 36

<211> 472

<212> PRT

ES 2 640 747 T3

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 36

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1      5      10      15
Tyr Lys Leu Gly Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
      20      25      30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
      35      40      45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50      55      60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65      70      75      80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
      85      90      95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
      100      105      110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
115      120      125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
130      135      140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145      150      155      160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
      165      170      175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
180      185      190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
195      200      205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210      215      220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225      230      235      240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
      245      250      255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
260      265      270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
275      280      285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
290      295      300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305      310      315      320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
      325      330      335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
340      345      350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
355      360      365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
370      375      380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385      390      395      400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
      405      410      415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
420      425      430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
435      440      445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
450      455      460
Glu Phe Glu Thr Met Asp Lys Leu
465      470

```

5

<210> 37

<211> 1845

ES 2 640 747 T3

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

5

<400> 37

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattggacct ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tatcgaagct gcggcgcgga tcgctggctga atcttcgacc      180
ggtagcctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggtga caaaggcaag      240
gcctaccacg tggagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac      300
ccgctcgacc tgtttgaaga agggtcggtc accaaccatcc tgacctcgat cgtcggtaac      360
gtggttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctgcc      420
ttggtcaaaa ccttccaagg tctccccac ggtatccaag tcgagcgcgga cctgctgaaac      480
aagtaocggc gtccgatgct gggttgacg atcaaaccac aactcggctt gtcggcgaaa      540
aactacggtc gtccgctcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaacatca actcgacgac gttccaacgc tggcgcgatc gcttctgtt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaccggg gaaatcaaag gtactactcct gaacgtgacc      720
gcgcccgcct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaac acgggatca cttccgtgtc ttggccaagt gttgcgtct gtccgggtgt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaag ttcgaccttg      1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tgggtcttc      1080
ttcacccaag attgggcgct gatgccgggc gtgctgcccg ttgcttccgg tggatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actcgttct ccagttcgtg      1200
ggcggcacct tgggtcacc cttgggtaat gctcctgggt caaccgcgaa ccgtgttgcc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac      1320
atccttcgtg aagctggcaa gtggctgcct gaactggctg ctgccctcga cctctgaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gacacctga ctatcgctgg      1440
gggagtgagc gttgctgctg aaagcttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctctgtac      1560
ctgctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaagg      1620
ttcaccctc tgatcgagtt caacgagcac tcgaaatccg aagagttcta ctggacgatg      1680
tggaaactcc ccctggttga ctgcaagagc cctcagcaag tctcgtatga agtgctgag      1740
tgccgcagcg aatacgtgta ttgctacatc cgtgtcgtg gttcgcgaaa catcaagcag      1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845
    
```

<210> 38

<211> 472

10

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 38

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
          35          40          45
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65          70          75          80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
          85          90          95
    
```

15

ES 2 640 747 T3

Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 39

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 39

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaactcacc	60
tattggacc	ccgattacac	ccccaaagac	actgacctgc	tggcggcttt	ccgcttcagc	120
cctcagccgg	gtgtccctgc	tatcgaagct	agcgcggcga	tcgaggctga	atcttcgacc	180
ggtacctgga	ccacogtgtg	gaccgacttg	ctgaccgaca	tggatcggtg	caaaggcaag	240
gcctaccacg	tggagccggg	gcaaggcga	gagaactcct	actttgcgtt	catcgcttac	300
ccgctcgacc	tgtttgaaga	agggtcggtc	accaacatcc	tgacctcgat	cgtcggtaac	360
gtgtttggct	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttggtcaaaa	ccttccaagg	tcttccccac	ggtatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgacag	atcaaacc	aactcggctc	gtcggcgaaa	540

ES 2 640 747 T3

```

aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgacg gcttctctgt tgtggctgat    660
gcaatccaca aatcgcaagc agaaacgggt gaaatcaaag gtactactct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgttaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaag ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attggggcgtc gatgccgggc gtgctgcccg ttgcttccgg tggatccac   1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcgggt   1200
ggcagctcc ttggtcaccc ctggggtaat gctcctggtg caaccgcgaa cgtggttgc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggctgcct gaactggctg ctgccctcga cctctgaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgcct gttgctcgt aaagctttct ccccagcctt cgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctctgtac   1560
ctgctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttcaccctct tgatcgagtt caacgagcac tcgaaatccg aagagttcta ctggacgatg   1680
tggaaagctc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgctgag   1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag   1800
tgcaaacccg tgagcttcat cgttcatcgt cccggccgat actaa                    1845

```

<210> 40

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 40

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
          35          40          45
Glu Ala Ser Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
          65          70          75          80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
          85          90          95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
          100          105          110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
          115          120          125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
          130          135          140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
          145          150          155          160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
          165          170          175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
          180          185          190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
          195          200          205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
          210          215          220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
          225          230          235          240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
          245          250          255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
          260          265          270

```

ES 2 640 747 T3

Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 41

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 41

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaaactcacc	60
tattggacc	ccgattacac	ccccaagac	actgacctgc	tggcggcttt	ccgcttcagc	120
ccctcagccg	gtgtccctgc	tgtcgaagct	gcggcgccga	tcgcggctga	atcttcgacc	180
ggtaacctga	ccaccgtgtg	gaccgacttg	ctgaccgaca	tggatcggtg	caaaggcaag	240
gcctaccacg	tgagaccggt	gcaaggcga	gagaactcct	actttgcgtt	catcgcttac	300
ccgctcggac	gtttgaaaga	agggctcgtc	accaaactcc	tgacctcgat	cgctcgtaac	360
gtgtttggt	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttgtcaaaa	ccttccaagg	tcctccccac	ggatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgacg	atcaaacca	aactcggctc	gtcggcgaaa	540
aactacggtc	gtgccgtcta	cgaatgtctg	cgccggcgtc	tggacttcac	caaagacgac	600
gaaaacatca	actcgcagcc	gttccaacgc	tggcgcgac	gcttctctgt	tgtggctgat	660
gcaatccaca	aatcgcaagc	agaaaccggt	gaaatcaaag	gtcactacct	gaacgtgacc	720
gcggccagct	gcgaagaaat	gatgaaacgg	gctgagttcg	ctaaagaact	cgccatgccc	780
atcatcatgc	atgacttctt	gacggctggt	ttcaccccca	acaccacctt	ggcaaaaatg	840
tgcccgacg	acggcgtcct	gctgcacatc	caccgtgcaa	tgcacgcggt	gatcgaccgt	900
cagcgtaacc	acgggattca	cttccgtgtc	ttggccaagt	gtttgcgtct	gtccggtggt	960
gaccacctcc	actccggcac	cgctcgtcgg	aaactggaag	gcgacaaaag	ttcgaccttg	1020
ggctttgttg	acttgatgcg	cgaagaccac	atcgaagctg	accgcagccg	tggggctctc	1080
ttcacccaag	attgggcgct	gatgccgggc	gtgctgccc	ttgcttccgg	tggtatccac	1140
gtgtggcaca	tgcccgcact	ggtggaatc	ttcggtgatg	actccgttct	ccagttcggt	1200
ggcggcacct	tggttcaccc	ctggggtaat	gctcctggtg	caaccgcgaa	ccgtgttgcc	1260
ttggaagctt	gcgtccaagc	tcggaacgaa	ggtcgcgacc	tctaccgtga	aggcggcgac	1320
atccttcgtg	aagctggcaa	gtggctgcct	gaactggctg	ctgccctcga	cctctggaaa	1380
gagatcaagt	tcgaattcga	aacgatggac	aagctctaag	gagcctcga	ctatcgctgg	1440
gggagtgagc	gttctgcgtg	aaagctttct	ccccagcctt	tcgacttaac	ctttcaggat	1500
ttctgaatca	tgagcatgaa	aactctgccc	aaagagcgtc	gtttcgagac	tttctcgtac	1560
ctgcctcccc	tcagcgatcg	ccaatcgctc	gcacaaatcg	agtacatgat	cgagcaaggc	1620
ttccaccctt	tgatcgagtt	caacgagcac	tcgaatccgg	aagagttcta	ctggacgatg	1680
tggaagctcc	ccctgtttga	ctgcaagagc	cctcagcaag	tcctcgatga	agtgcgtgag	1740
tgccgcagcg	aatacggtga	ttgctacatc	cgtgtcgtcg	gcttcgacaa	catcaagcag	1800
tgccaaaccg	tgagcttcat	cgttcatcgt	cccggccgat	actaa		1845

10

<210> 42

ES 2 640 747 T3

<211> 472

<212> PRT

<213> Artificial

<220>

5

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 42

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1      5      10      15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
      20      25      30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Val
      35      40
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
      50      55      60
Thr Val Trp Thr Asp Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65      70      75      80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
      85      90      95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
      100      105      110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
      115      120      125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
      130      135      140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145      150      155      160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
      165      170      175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
      180      185      190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
      195      200      205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
      210      215      220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
      225      230      235      240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
      245      250      255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
      260      265      270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
      275      280      285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
      290      295      300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
      305      310      315      320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
      325      330      335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
      340      345      350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
      355      360      365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
      370      375      380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
      385      390      395      400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
      405      410      415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
      420      425      430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
      435      440      445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
      450      455      460
Glu Phe Glu Thr Met Asp Lys Leu
      465      470

```

<210> 43

ES 2 640 747 T3

<211> 1845

<212> ADN

<213> Artificial

<220>

5 <223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 43

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcgggctt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc      180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag      240
tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac      300
ccgctcgacc tgtttgaaga agggctcggc accaacatcc tgacctcgat catcggtaat      360
gtgtttggct tcaaagctat ccgctcgtcg cgtctggaag acatccgctt ccccgctgcc      420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtaccggc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtccggcga    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctggt tgtggctgat    660
gcaatccaca aatcgcaagc agaaacgggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggg    960
gaccacctcc actccggcac cgctcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggctctc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg ttgatatccac   1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccggtct ccagttcggg   1200
ggcggcacct tgggtcacc cttggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtcatgat cgagcaaggc   1620
ttcacccctc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc cctgttttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgtcg gcttcgacaa catcaagcag   1800
tgccaaaccc tgagcttcat cgttcatcgt cccggccgat actaa                1845

```

<210> 44

10 <211> 472

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

15 <400> 44

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
          35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65          70          75          80

```

ES 2 640 747 T3

Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 45

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 45

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc 60
tattacacc ccgattacac ccccaaagac actgacatac tggcggcttt ccgcttcagc 120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag 240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcggt catcgcttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaacctcc tgacctcgat cgtcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt gcccgctgcc 420
    
```

ES 2 640 747 T3

```

ttggtcaaaa ctttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtccggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgata gctcctctgt tgtggctgat 660
gcaatccaca aatcgcgaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcccggacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020
ggctttgttg acttgatgcy cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcaccacaag attggggcgtc gatgccgggc gtgctgcccg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcgccact ggtggaatc ttcggtgatg actccgttct ccagttcggg 1200
ggcggcacct tgggtcacc cttgggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320
atccttcgtg aagctggcaa gtggctgcct gaactggctg ctgcctcga cctctgaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgcgtg 1440
gggagtgagc gttgctcgtt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctctgtac 1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacctgat cgagcaaggc 1620
ttcaccacct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtagg 1740
tgccgcagcg aatcaggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

```

<210> 46

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 46

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20          25          30
Ile Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65          70          75          80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala
85          90          95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
100         105         110
Leu Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
115         120         125
Ser Leu Arg Leu Glu Asp Ile Arg Leu Pro Val Ala Leu Val Lys Thr
130         135         140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145         150         155         160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
165         170         175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
180         185         190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
195         200         205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210         215         220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225         230         235         240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
245         250         255

```

ES 2 640 747 T3

Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 47

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 47

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc 60  
 tattacaccc ccgattacac ccccaaagac actgacatac tggcggcttt ccgcttcagc 120  
 cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc 180  
 ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggtg caaaggcaag 240  
 tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300  
 ccgctcgacc tgtttgaaga agggctcggtc accaactcct tgacctcgat catcggtaac 360  
 gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt gcccgctgcc 420  
 ttggtcaaaa ccttccaagg tcttccccac ggtatccaag tcgagcgcga cctgctgaac 480  
 aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtccggcga 540  
 aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600  
 gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttctgtt tgtggctgat 660  
 gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720  
 gcgcgcacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780  
 atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccactt ggcaaaatgg 840  
 tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt 900  
 cagcgtaac acgggattca cttccgtgct ttggccaagt gtttgcgtc gtccgggtgg 960  
 gaccacctcc actccggcac cgctcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020  
 ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080  
 ttcacccaag attggcgctc gatgccgggc gtgctgcccg ttgcttccgg tggatccac 1140  
 gtgtggcaca tgcccgact ggtggaatc ttcggtgatg actccgttct ccagttcgg 1200  
 ggcggcacct tgggtcacc cttgggtaat gctcctgggt caaccgcgaa ccgtgttgcc 1260  
 ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320  
 atccttctgt aagctggcaa gtggctgcct gaactggctg ctgccctcga cctctgaaa 1380  
 gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctg 1440  
 gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500  
 ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac ttctctgtac 1560  
 ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacctgat cgagcaaggc 1620  
 ttccaccctt tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680  
 tggaaactcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag 1740

tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800  
 tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

10

ES 2 640 747 T3

<210> 48

<211> 472

<212> PRT

<213> Artificial

5

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 48

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Ile Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
          35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65          70          75          80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala
          85          90          95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
          100          105          110
Leu Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
          115          120          125
Ser Leu Arg Leu Glu Asp Ile Arg Leu Pro Val Ala Leu Val Lys Thr
          130          135          140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145          150          155          160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
          165          170          175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
          180          185          190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
          195          200          205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210          215          220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225          230          235          240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
          245          250          255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
          260          265          270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
          275          280          285
His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His
290          295          300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305          310          315          320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
          325          330          335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
          340          345          350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
          355          360          365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
          370          375          380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385          390          395          400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
          405          410          415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
          420          425          430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
          435          440          445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
          450          455          460
Glu Phe Glu Thr Met Asp Lys Leu
465          470

```

# ES 2 640 747 T3

<210> 49

<211> 1845

<212> ADN

<213> Artificial

5

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 49

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgatagacta caaactcggc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc      180
ggtaacctga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggtg caaaggcaag      240
tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcggt catcgcttac      300
ccgctcagcc tgtttgaaga agggctcggc accaaccatcc tgacctcgat cgtcggtaac      360
gtgtttggct tcaaagctat ccgttcgtcg cgtctggaag acatccgctt ccccgtcgcc      420
ttgttcaaaa ccttccaagg tctccccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa      540
aactacggct gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaacatca actcgcagcc gttccaacgc tggcgcgcat gcttctggtt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaaccggt gaaatcaaag gtaactacct gaacgtgacc      720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaagc ttcgaccttg      1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc      1080
ttcaccceaag attgggctgc gatgccgggc gtgctgcccg ttgcttccgg tggtatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcggg      1200
ggcggcacct tgggtcaccc ctggggtaat gctcctgggt caaccgcgaa ccgtgtgccc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga agggggcgac      1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctcgtt aaagctttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac      1560
ctgcctccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc      1620
ttccaccctc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag      1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtcg gcttcgacaa catcaagcag      1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845
    
```

<210> 50

10

<211> 472

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

15

<400> 50

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Ile Asp
1          5          10          15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
          20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
          35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
          50          55          60
    
```

ES 2 640 747 T3

Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys  
65 70 75 80  
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala  
85 90 95  
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
100 105 110  
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
115 120 125  
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
130 135 140  
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
145 150 155 160  
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
165 170 175  
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
180 185 190  
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
195 200 205  
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
210 215 220  
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
225 230 235 240  
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
245 250 255  
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
260 265 270  
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
275 280 285  
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
290 295 300  
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
305 310 315 320  
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
325 330 335  
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
340 345 350  
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
355 360 365  
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
370 375 380  
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
385 390 395 400  
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
405 410 415  
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
420 425 430  
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
435 440 445  
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
450 455 460  
Glu Phe Glu Thr Met Asp Lys Leu  
465 470

<210> 51

<211> 1848

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 51

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta cgctaagatg 60  
ggatactgga cccccgatta ccccccaaaa gacactgacc tgctggcggc tttccgcttc 120  
agccctcagc cgggtgtccc tgctgacgaa gctggtgctg cgatcgcggc tgaatcttcg 180  
accggtacct ggaccaccgt gtggaccgac ttgctgaccg acatggatcg gtacaaaggc 240  
aagtgctacc acatcgagcc ggtgcaaggc gaagagaact cctactttgc gttcatcgct 300

ES 2 640 747 T3

```

taccgctcg acctgtttga agaagggtcg gtcaccaaca tcctgacctc gatcgtcgg 360
aacgtgtttg gcttcaaagc tatccggttc ctgctgctgg aagacatccg cttccccctc 420
gccttggtca aaaccttcca aggtcctccc cacggatcc aagtcgagcg cgacctgctg 480
aaccaagtac gccgtccgat gctgggttgc acgatcaaac caaaactcgg tctgtcggcg 540
aaaaactacg gtcgtgccgt ctacgaatgt ctgcgcggcg gtctggactt caccaaagac 600
gacgaaaaca tcaactcgca gccgttccaa cgctggcgcg atcgcttcct gttgtggct 660
gatgcaatcc acaaatcgca agcagaaacc ggtgaaatca aaggctacta cctgaactg 720
accgcccga cctgcgaaga aatgatgaaa cgggtgagt tcgctaaaga actcggcatg 780
ccgatcatca tgcatgactt cttgacggct ggtttcaccg ccaacaccac cttggcaaaa 840
tgggtccgcg acaacggcgt cctgctgcac atccaccgtg caatgcacgc ggtgatcgac 900
cgtcagcgta accacgggat tcacttccgt gtcttgcca agtgtttgcg tctgtccggt 960
ggtgaccacc tccactccgg caccgtcgtc ggcaaaactgg aaggcgacaa agcttcgacc 1020
ttgggctttg ttgacttgat ggcggaagac cacatcgaag ctgaccgcag ccgtggggtc 1080
ttcttaccac aagattgggc gtcgatgccc ggcgtgctgc cggttgcttc cggtggatc 1140
cacgtgtggc acatgccgc actgggtgaa atcttcggtg atgactccgt tctccagttc 1200
ggtggcggca ccttgggtca cccctggggt aatgctcctg gtgcaaccgc gaaccgtgtt 1260
gccttgaag cttgcgtcca agctcggaac gaaggtcgcg acctctaccg tgaaggcggc 1320
gacatccttc gtgaagctgg caagtggctg cctgaaactgg ctgctgccct cgacctctgg 1380
aaagagatca agttcgaatt cgaaacgatg gacaagctct aaggagcctc tgactatcg 1440
tgggggagtg agcgttgctg cgtaaagctt tctccccagc ctttcgactt aacctttcag 1500
gattttctga tcatgagcat gaaaactctg ccaaaagagc gtcgtttcga gactttctc 1560
tacctgcctc ccctcagcga tcgccaaatc gctgcacaaa tcgagtacat gatcagagca 1620
ggcttccacc ccttgatcga gttcaacgag cactcgaatc cggaagagtt ctactggac 1680
atgtggaagc tccccctgtt tgactgcaag agccctcagc aagtcctcga tgaagtgcgt 1740
gagtgccgca gcgaatacgg tgattgctac atccgtgctc ctggcttcga caacatcaag 1800
cagtgccaaa ccgtgagctt catcgttcat cgtcccggcc gatactaa 1848

```

<210> 52

<211> 473

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 52

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Ala Lys Met Gly Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr
          20          25          30
Asp Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala
          35          40          45
Asp Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp
          50          55          60
Thr Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly
65          70          75          80
Lys Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe
          85          90          95
Ala Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr
          100         105         110
Asn Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile
          115         120         125
Arg Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys
130         135         140
Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu
145         150         155         160
Asn Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu
          165         170         175
Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg
          180         185         190
Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro
          195         200         205
Phe Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His
210         215         220
Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val
225         230         235         240

```

ES 2 640 747 T3

Thr Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys  
 245 250 255  
 Glu Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe  
 260 265 270  
 Thr Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu  
 275 280 285  
 Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn  
 290 295 300  
 His Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly  
 305 310 315 320  
 Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp  
 325 330 335  
 Lys Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile  
 340 345 350  
 Glu Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser  
 355 360 365  
 Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His  
 370 375 380  
 Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe  
 385 390 395 400  
 Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr  
 405 410 415  
 Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly  
 420 425 430  
 Arg Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys  
 435 440 445  
 Trp Ser Pro Glu Leu Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys  
 450 455 460  
 Phe Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 53

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 53

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgatagacta caaactcggc 60  
 tattcacacc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120  
 cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggtga atcttcgacc 180  
 ggtacctgga ccgctcgtgt gaccgacttg ctgaccgaca tggatcggta ccgcggaag 240  
 tgcctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300  
 ccgctcagacc tgtttgaaga agggctcggtc accaacatcc tgacctgat cgtcggtaac 360  
 gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc 420  
 ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgaga cctgctgaac 480  
 aagtacggcc gtcgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa 540  
 aactacggtc gtgcccgtta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600  
 gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttctgtt tgtggctgat 660  
 gcaatccaca aatcgcgaagc agaaacgggt gaaatcaaag gtcactacct gaacgtgacc 720  
 gcgcccaccc gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780  
 atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaatgg 840  
 tgcccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgcaccg 900  
 cagcgtaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt 960  
 gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaagc ttcgaccttg 1020  
 ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080  
 ttcacccaag attgggcgct gatgccgggc gtgctgcccg ttgcttccgg tggatccac 1140  
 gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcgg 1200  
 ggcggcacct tgggtcacc cttgggtaat gctcctgggt caaccgcgaa ccggtgtg 1260  
 ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320  
 atccttcgtg aagctggcaa gtggctgcct gaactggctg ctgcccctga cctctgaaa 1380  
 gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440  
 gggagtgagc gttgctcgtt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500  
 ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560  
 ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacctgat cgagcaaggc 1620

ES 2 640 747 T3

ttccaccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680  
 tggaagctcc cctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag 1740  
 tgccgcagcg aatacggatg ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag 1800  
 tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

<210> 54

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 54

<400> 54

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Ile Asp  
 1 5 10 15  
 Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp  
 20 25 30  
 Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp  
 35 40 45  
 Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
 50 55 60  
 Val Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys  
 65 70 75 80  
 Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415

10

# ES 2 640 747 T3

```

Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
      420                               425                               430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
      435                               440                               445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
      450                               455                               460
Glu Phe Glu Thr Met Asp Lys Leu
      465                               470

```

<210> 55

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 55

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggtgta atcttcgacc      180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggtg ccgcggcaag      240
tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac      300
gagctcgacc tgtttgaaga agggctcggc accaacatcc tgacctgat  cgtcggtaac      360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc      420
ttggtcaaaa ctttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtacggcc gtcgatgct gggttgcacg atcaaaccaa aactcgttct gtcggcgaaa      540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gttcctctgt tgtggctgat      660
gcaatccaca aatcgcgaag agaaaccggg gaaatcaaag gtcactacct gaacgtgacc      720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgctct gctgcacatc caccgtgcaa tgacgcgggt gatcgaccgt      900
cagcgttaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaag ttcgaccttg      1020
ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc      1080
ttcacccaag attgggcgct gatgccgggc gtgctgcccg ttgcttccgg tggatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttccgtgatg actccgttct ccagttcggg      1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcgcgac      1320
atccttctgt aagctggcaa gtggctgcct gaactggctg ctgccctcga cctctggaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac      1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc      1620
ttcaccctct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag      1740
tgccgcagcg aatacgtgta ttgctacatc cgtgtcgtg gtttcgacaa catcaagcag      1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

10

<210> 56

<211> 472

<212> PRT

<213> Artificial

<220>

15

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 56

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1      5      10      15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
      20      25      30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
      35      40      45

```

ES 2 640 747 T3

Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
50 55 60  
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys  
65 70 75 80  
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala  
85 90 95  
Phe Ile Ala Tyr Glu Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
100 105 110  
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
115 120 125  
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
130 135 140  
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
145 150 155 160  
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
165 170 175  
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
180 185 190  
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
195 200 205  
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
210 215 220  
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
225 230 235 240  
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
245 250 255  
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
260 265 270  
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
275 280 285  
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
290 295 300  
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
305 310 315 320  
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
325 330 335  
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
340 345 350  
Ala Asp Arg Ser Arg Gly Val Phe Thr Gln Asp Trp Ala Ser Met  
355 360 365  
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
370 375 380  
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
385 390 395 400  
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
405 410 415  
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
420 425 430  
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
435 440 445  
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
450 455 460  
Glu Phe Glu Thr Met Asp Lys Leu  
465 470

<210> 57

5 <211> 1845

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

10 <400> 57

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaagcccta caaactcggc 60  
tattacacc cggattacac ccccaagac actgacctgc tggcggcttt ccgcttcagc 120  
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgctggctga atcttcgacc 180

ES 2 640 747 T3

```

ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgaggcaag 240
tgctaccaca tcgagccggg gcaagcgcaa gagaactcct actttgcgtt catcgcttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctgcc 420
ttggtcaaaa ctttccaagg tcttccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctt gtcggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caagacgac 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gttcctgtt tgtggctgat 660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaacc acgggattca cttccgtgct ttggccaagt gttgctgtt gtcgggtggt 960
gaccacctcc actccggcac cgctcgtcggc aaactggaag gcgacaaaagc ttcgaccttg 1020
ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcacccaag attgggcgct gatgccgggc gtgctgcccg ttgcttccgg tggatccac 1140
gtgtggcaca tccccgact ggtggaatc ttcggtgatg actccgttct ccagttcggg 1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcgcgac 1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctcga ctatcgctgg 1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttcttaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtcatgat cgagcaaggc 1620
ttcacccctt tgatcgagtt caacgagcac tcgaaatccg aagagttcta ctggacgatg 1680
tggaaagctc cctgtttga ctgcaagagc cctcagcaag tcctcgtatga agtgctgag 1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

```

<210> 58

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 58

```

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

```

ES 2 640 747 T3

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

<210> 59

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 59

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc 60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggaag 240
tgctaccaca tcgagccggg gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300
ccgctcgacc tgtttgaaga agggctcggc accaacatcc tgacctcgat catcggtaac 360
gtggttggtt tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctgcc 420
ttggtcaaaa cttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctt gtcggcgaaa 540
aactacggtc gtccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600
gaaaacatca actcgagccc gttccaacgc tggcgcgatc gcttcctggt tgtggctgat 660
gcaatccaca aatcgcaagc agaaccggg gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcaagaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccactt ggcaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt 900
cagcgtaac acgggattca cttccgtgct ttggccaagt gttgctgctt gtcgggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020
ggctttggtt acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttacccaag attgggcgtc gatgccgggc gtgctgcccg ttgcttccgg tggatccac 1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcggg 1200
ggcggcacct tgggtcacc cttgggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320
atccttctgt aagctggcaa gtggctgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
    
```

# ES 2 640 747 T3

```

ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc 1620
ttccaccctt tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggaagctcc cctgttttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag 1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

```

<210> 60

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 60

<400> 60

```

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1          5          10          15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20          25          30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
35          40          45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50          55          60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65          70          75          80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
85          90          95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
100         105         110
Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
115         120         125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
130         135         140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145         150         155         160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
165         170         175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
180         185         190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
195         200         205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
210         215         220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225         230         235         240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
245         250         255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
260         265         270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
275         280         285
His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His
290         295         300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305         310         315         320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
325         330         335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
340         345         350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
355         360         365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
370         375         380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385         390         395         400

```

10

ES 2 640 747 T3

Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 61

<211> 1845

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 61

atggccaaga	cgcaatctgc	cgcaggctat	aaggccgggg	tgaaggacta	caaaactcggc	60
tattacaccc	ccgattacac	ccccaaagac	actgacctgc	tggcggcttt	ccgcttcagc	120
cctcagccgg	gtgtccctgc	tgacgaagct	ggtgcccgca	tcgcccgtga	atcttcgacc	180
ggtacctgga	ccaccgtgtg	gaccgacttg	ctgaccgaca	tggatcggta	caaaggcaag	240
tgctaccaca	tcgagccggg	gcaagggcaa	gagaactcct	actttgcggt	catcgcttac	300
ccgctcgacc	tgtttgaaga	agggtcggtc	accaacatcc	tgacctcgat	cgtcggtaac	360
gtggttggct	tcaaagctat	ccgttcgctg	cgtctggaag	acatccgctt	ccccgtcgcc	420
ttgggtcaaaa	ccttccaagg	tcctccccac	ggtatccaag	tcgagcgcga	cctgctgaac	480
aagtacggcc	gtccgatgct	gggttgacag	atcaaaacaa	aactcggctc	gtcggcgaaa	540
aactacggtc	gtgcccgtca	cgaaatgtctg	cgccggcggtc	tggacttcac	caaagacgac	600
gaaaacatca	actcgagcc	gttccaacgc	tggcgcgatac	gcttctgtt	tgtggctgat	660
gcaatccaca	aatcgcaagc	agaaaccggt	gaaatcaaag	gtcactacct	gaacgtgacc	720
gcgcccagcct	gcgaagaaat	gatgaaacgg	gctgagttcg	ctaaagaact	cggcatgccc	780
atcatcatgtc	atgacttctt	gacggctggt	ttcaccgcga	acaccacctt	ggcaaaatgg	840
tgcccgagaca	acggcgtcct	gctgcacatc	caccgtgcaa	tgcacgcggt	gatcgaccgt	900
cagcgtaaac	acgggattca	cttccgtgtc	ttggccaagt	gtttgcgctc	gtccgggtgt	960
gaccacctcc	actccggcac	cgctcgtcggc	aaactggaag	gcgacaaaagc	ttcgaccttg	1020
ggctttggtg	acttgatgcg	cgaagaccac	atcgaagctg	accgcagccg	tgggtcttc	1080
ttcacccaag	attgggcgctc	gatgcccggc	gtgctgcccg	ttgcttccgg	tggatccac	1140
gtgtggcaca	tgcccgcact	ggtggaatc	ttcggtgatg	actccgttct	ccagttcggg	1200
ggcggcacct	tgggtcacc	ctggggtaat	gctcctgggtg	caaccgcgaa	ccgtgttgcc	1260
ttggaagctt	gcgtccaagc	tcggaacgaa	ggtcgcgacc	tctaccgtga	aggcggcgac	1320
atccttcgtg	aagctggcaa	gtggtcgcct	gaactggctg	ctgccctcga	cctctggaaa	1380
gagatcaagt	tcgaattcga	aacgatggac	aagctctaag	gagcctctga	ctatcgctgg	1440
gggagtgagc	ggtgctgctg	aaagctttct	ccccagcctt	tcgacttaac	ctttcaggat	1500
ttctgaatca	tgagcatgaa	aactctgccc	aaagagcgtc	gtttcgagac	tttctcgtac	1560
ctgcctcccc	tcagcgatcg	ccaaatcgct	gcacaaatcg	agtacatgat	cgagcaaggc	1620
ttccaccctc	tgatcgagtt	caacgagcac	tcgaatccgg	aagagttcta	ctggacgatg	1680
tggaagctcc	ccctgtttga	ctgcaagagc	cctcagcaag	tcctcgatga	agtgctgag	1740
tgccgcagcg	aatacgggtga	ttgctacatc	cgtgtcgtg	gcttcgacaa	catcaagcag	1800
tgccaaaccg	tgagcttcat	cgttcatcgt	cccggccgat	actaa		1845

10

<210> 62

<211> 472

<212> PRT

<213> Artificial

15

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 62

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp  
 1 5 10 15  
 Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp

# ES 2 640 747 T3

	20		25		30										
Leu	Leu	Ala	Phe	Arg	Phe	Ser	Pro	Gln	Pro	Gly	Val	Pro	Ala	Asp	
	35					40					45				
Glu	Ala	Gly	Ala	Ala	Ile	Ala	Ala	Glu	Ser	Ser	Thr	Gly	Thr	Trp	Thr
	50				55						60				
Thr	Val	Trp	Thr	Asp	Leu	Leu	Thr	Asp	Met	Asp	Arg	Tyr	Lys	Gly	Lys
65				70					75						80
Cys	Tyr	His	Ile	Glu	Pro	Val	Gln	Gly	Glu	Glu	Asn	Ser	Tyr	Phe	Ala
			85					90						95	
Phe	Ile	Ala	Tyr	Pro	Leu	Asp	Leu	Phe	Glu	Glu	Gly	Ser	Val	Thr	Asn
			100				105						110		
Ile	Leu	Thr	Ser	Ile	Val	Gly	Asn	Val	Phe	Gly	Phe	Lys	Ala	Ile	Arg
	115					120						125			
Ser	Leu	Arg	Leu	Glu	Asp	Ile	Arg	Phe	Pro	Val	Ala	Leu	Val	Lys	Thr
	130				135						140				
Phe	Gln	Gly	Pro	Pro	His	Gly	Ile	Gln	Val	Glu	Arg	Asp	Leu	Leu	Asn
145					150					155					160
Lys	Tyr	Gly	Arg	Pro	Met	Leu	Gly	Cys	Thr	Ile	Lys	Pro	Lys	Leu	Gly
			165					170						175	
Leu	Ser	Ala	Lys	Asn	Tyr	Gly	Arg	Ala	Val	Tyr	Glu	Cys	Leu	Arg	Gly
		180					185						190		
Gly	Leu	Asp	Phe	Thr	Lys	Asp	Asp	Glu	Asn	Ile	Asn	Ser	Gln	Pro	Phe
	195					200						205			
Gln	Arg	Trp	Arg	Asp	Arg	Phe	Leu	Phe	Val	Ala	Asp	Ala	Ile	His	Lys
	210				215						220				
Ser	Gln	Ala	Glu	Thr	Gly	Glu	Ile	Lys	Gly	His	Tyr	Leu	Asn	Val	Thr
225					230					235					240
Ala	Pro	Thr	Cys	Glu	Glu	Met	Met	Lys	Arg	Ala	Glu	Phe	Ala	Lys	Glu
			245					250						255	
Leu	Gly	Met	Pro	Ile	Ile	Met	His	Asp	Phe	Leu	Thr	Ala	Gly	Phe	Thr
			260					265					270		
Ala	Asn	Thr	Thr	Leu	Ala	Lys	Trp	Cys	Arg	Asp	Asn	Gly	Val	Leu	Leu
	275						280					285			
His	Ile	His	Arg	Ala	Met	His	Ala	Val	Ile	Asp	Arg	Gln	Arg	Asn	His
	290				295						300				
Gly	Ile	His	Phe	Arg	Val	Leu	Ala	Lys	Cys	Leu	Arg	Leu	Ser	Gly	Gly
305					310					315					320
Asp	His	Leu	His	Ser	Gly	Thr	Val	Val	Gly	Lys	Leu	Glu	Gly	Asp	Lys
			325						330					335	
Ala	Ser	Thr	Leu	Gly	Phe	Val	Asp	Leu	Met	Arg	Glu	Asp	His	Ile	Glu
			340					345					350		
Ala	Asp	Arg	Ser	Arg	Gly	Val	Phe	Phe	Thr	Gln	Asp	Trp	Ala	Ser	Met
	355						360						365		
Pro	Gly	Val	Leu	Pro	Val	Ala	Ser	Gly	Gly	Ile	His	Val	Trp	His	Met
	370					375					380				
Pro	Ala	Leu	Val	Glu	Ile	Phe	Gly	Asp	Asp	Ser	Val	Leu	Gln	Phe	Gly
385					390					395					400
Gly	Gly	Thr	Leu	Gly	His	Pro	Trp	Gly	Asn	Ala	Pro	Gly	Ala	Thr	Ala
			405					410						415	
Asn	Arg	Val	Ala	Leu	Glu	Ala	Cys	Val	Gln	Ala	Arg	Asn	Glu	Gly	Arg
			420					425					430		
Asp	Leu	Tyr	Arg	Glu	Gly	Gly	Asp	Ile	Leu	Arg	Glu	Ala	Gly	Lys	Trp
	435						440					445			
Ser	Pro	Glu	Leu	Ala	Ala	Ala	Leu	Asp	Leu	Trp	Lys	Glu	Ile	Lys	Phe
	450					455					460				
Glu	Phe	Glu	Thr	Met	Asp	Lys	Leu								
465					470										

<210> 63

<211> 1845

5

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 63

# ES 2 640 747 T3

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcggctga atcttcgacc      180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggtg ccgcggcaag      240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcggt catcgcttac      300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac      360
gtgtttggt tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc      420
ttgtcaaaa ccttccaagg tctccccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa      540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaaatca actcgcagcc gttccaacgc tggcgcgac gcttctgtt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc      720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgttaac acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt      960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg      1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc      1080
ttcaccceaag attgggcgct gatgccgggc gtgctgcggt ttgcttcggt tggtatccac      1140
gtgtggcaca tgcccgaact ggtggaatc ttcggtgatg actccgttct ccagttcggg      1200
ggcggcacct tgggtcacc cttgggtaat gctcctggtg caaccgcgaa ccgtgttgcc      1260
ttggaagctt ggcgccaaag tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac      1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgcctcga cctctgaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctgctt aaagcttctt ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgta      1560
ctgctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacctgat cgagcaaggc      1620
ttccaccctt tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgctgag      1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag      1800
tgccaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

<210> 64

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 64

ES 2 640 747 T3

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp  
 1 5 10 15  
 Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp  
 20 25 30  
 Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp  
 35 40 45  
 Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
 50 55 60  
 Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys  
 65 70 75 80  
 Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 65

<211> 1845

5

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 65

ES 2 640 747 T3

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc 60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt cgccttcagc 120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag 240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgctgt catcgccttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaaccatcc tgacctcgat catcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc 420
ttggtcaaaa ccttccaagg tcctcccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgat 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gtttcctggt tgtggtgat 660
gcaatccaca aatcgcgaagc agaaacgggt gaaatcaaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgcccgacga acggcgtcct gctgcacatc caccgtgcag ggcacgcggt gatcgaccgt 900
cagcgttaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaaagc ttcgaccttg 1020
ggctttggtg acttgatcgc cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcccgcact ggtgaaatc ttcggtgatg actccgttct ccagttcggc 1200
ggcggcacct tgggtcacc ctagggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga agcggcgacc 1320

atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgctcccc tcagcgcgct ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc 1620
ttcaccctct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggagctcc cctggttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag 1740
tgccgcagcg aatacgggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

```

<210> 66

<211> 472

5

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 66

ES 2 640 747 T3

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp  
 1 5 10 15  
 Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp  
 20 25 30  
 Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp  
 35 40 45  
 Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr  
 50 55 60  
 Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys  
 65 70 75 80  
 Cys Tyr His Ile Glu Pro Val Gln Gly Glu Asn Ser Tyr Phe Ala  
 85 90 95  
 Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn  
 100 105 110  
 Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg  
 115 120 125  
 Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr  
 130 135 140  
 Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn  
 145 150 155 160  
 Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly  
 165 170 175  
 Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly  
 180 185 190  
 Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe  
 195 200 205  
 Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys  
 210 215 220  
 Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr  
 225 230 235 240  
 Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu  
 245 250 255  
 Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr  
 260 265 270  
 Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu  
 275 280 285  
 His Ile His Arg Ala Gly His Ala Val Ile Asp Arg Gln Arg Asn His  
 290 295 300  
 Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly  
 305 310 315 320  
 Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys  
 325 330 335  
 Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu  
 340 345 350  
 Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met  
 355 360 365  
 Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met  
 370 375 380  
 Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly  
 385 390 395 400  
 Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala  
 405 410 415  
 Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg  
 420 425 430  
 Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp  
 435 440 445  
 Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe  
 450 455 460  
 Glu Phe Glu Thr Met Asp Lys Leu  
 465 470

<210> 67

<211> 1845

5

<212> ADN

<213> Artificial

<220>

<223> Basándose en la secuencia de nucleótidos de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 67

ES 2 640 747 T3

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc     120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc     180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag     240
tgctaccaca tcgagccggg gcaaggcga gagaactcct actttgcgtt catcgcttac     300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac     360
gtggttggtc tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgctgcc     420
ttggtcaaaa ccttccaagg tctccccac ggtatccaag tcgagcgcga cctgctgaac     480
aagtaocggc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa     540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac     600
gaaaacatca actcgcagcc gtccaacgc tggcgcgatc gcttcctggt tgtggctgat     660
gcaatccaca aatcgcaagc agaaaaccgt gaaatcaaag gtactactc gaacgtgacc     720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg     780
atcatcatgc atgacttctt gacggctggt ttcaccgcc aacaccctt ggcaaaatgg     840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcaacgggt gatcgaccgt     900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt     960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg    1020
ggctttgttg acttgatgcy cgaagaccac atcgaagctg accgcagccg tggggcttc     1080
ttcacccaag attggggcgtc gatgcccggc gtgctgcggg ttgcttcggg tggatccac     1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccgttct ccagttcggg     1200
ggcggcacct tgggtcaacc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc     1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac     1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa     1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg     1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat     1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac     1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc     1620
ttcaccacct tgatcgagtt caacgagcac tcgaaatccg aagagttcta ctggacgatg     1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag     1740
tgccgcagcg aatacgggtg ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag     1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

<210> 68

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 68

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp



ES 2 640 747 T3

```

atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc      60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc      120
cctcagccgg gtgtccctgc tgacgaagct ggtgcccgga tcgcccgtga atcttcgacc      180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag      240
tgctaccaca tcgagccggt gcaaggcga gagaactcct actttgcggt catcgcttac      300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat catcggtaac      360
gtggttggt tcaaagctat ccgctcgtg cgtctggaag acatccgctt ccccgctgcc      420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac      480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggctc gtcggcgaaa      540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac      600
gaaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttctgttt tgtggctgat      660
gcaatccaca aatcgcaagc agaaaaccgt gaaatcaaag gtcactacct gaacgtgacc      720
gcgccgacct gcgaagaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg      780
atcatcatgc atgacttctt gacggctggt ttacacgcca acaccacctt ggcaaaatgg      840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt      900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtcgggtggt      960
gaccacctcc actccggcac cgctcgtcggc aaactggaag gcgacaaaagc ttcgaccttg      1020
ggctttggtg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggctctc      1080
ttcaccacaag attgggcgctc gatgccgggc gtgctgcccg ttgcttccgg tggtatccac      1140
gtgtggcaca tgcccgcact ggtggaatc ttcggtgatg actccggttct ccagttcggg      1200
ggcggcacct tgggtcacc cttgggtaat gctcctgggt caaccgcgaa ccgtgttgcc      1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga agcgggcgac      1320
atccttcgtg aagctgcaaa gtggtcgcct gaactggctg ctgcctcga cctctggaaa      1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg      1440
gggagtgagc gttgctgctg aaagctttct ccccagcctt tcgacttaac ctttcaggat      1500
ttctgaatca tagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac      1560
ctgcctcccc tcagcgatcg ccaaatcgtc gcacaaatcg agtacatgat cgagcaaggc      1620
ttcaccacct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg      1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcggtgag      1740
tgccgcagcg aatcaggtga ttgctacatc cgtgtcgtg gcttcgacaa catcaagcag      1800
tgccaaaacc tgagcttcat cgttcatcgt cccggccgat actaa      1845

```

<210> 70

<211> 472

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la secuencia traducida de rbcLS de *Synechococcus* sp. PCC6301 de tipo salvaje

<400> 70

```

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

```

ES 2 640 747 T3

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

<210> 71

<211> 1434

<212> ADN

5

<213> Nicotiana tabacum

<400> 71

atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taagagtac 60  
 aaattgactt attatactcc tgagtaccaa accaaggata ctgatatatt ggcagcattc 120  
 cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa 180  
 tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac 240  
 aaagggcgat gctaccgcat cgagcgtggt gttggagaaa aagatcaata tattgcttat 300  
 gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatggt tacttccatt 360  
 gtaggtaacg tatttgggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc 420  
 cctcctgctt atgttaaaac tttccaaggt cgcgctcatg ggatccaagt tgaaagagat 480  
 aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa atgggggtta 540  
 tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact 600  
 aaagatgatg agaacgtgaa ctcaacaacca tttatgcggt ggagagatcg ttctttattt 660  
 tgtgccgaag cactttataa agcacaggct gaaacagggt aaatcaaagg gcattacttg 720  
 aatgctactg caggtacatg cgaagaatg atcaaaagag ctgtatattgc tagagaattg 780  
 ggcgttccga tcgtaatgca tgactactta acggggggat tcaccgcaaa tactagcttg 840  
 gctcattatt gccgagataa tgggttactt cttcacatcc accgtgcaat gcatgcggtt 900  
 attgatagac agaagaatca tggtatccac ttccgggtat tagcaaaaagc gttacgtatg 960  
 tctggtggag atcatattca ctctgttacc gtagtaggta aacttgaagg tgaagagac 1020  
 ataacttttg gctttgttga tttactgcgt gatgattttg ttgaacaaga tcgaagtcgc 1080  
 ggtatttatt tcactcaaga ttgggtctct ttaccagggt ttctaccctg ggcttcagga 1140  
 ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta 1200  
 cagttcgtg gaggaacttt aggacatcct tggggtaatg cgccaggtgc cgtagctaat 1260  
 cgagtagctc tagaagcatg tgtaaaagct cgtaaatgaag gacgtgatct tgctcagga 1320  
 ggtaaatgaaa ttattcgcga ggctttgcaaa tggagcccgg aactagctgc tgcttggaa 1380  
 gtatggaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa 1434

<210> 72

ES 2 640 747 T3

<211> 477

<212> PRT

<213> Nicotiana tabacum

<400> 72

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Glu Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Lys Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn  
 420 425 430  
 Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys  
 465 470 475

5

<210> 73

<211> 1434

<212> ADN

# ES 2 640 747 T3

<213> Artificial

<220>

<223> Basándose en la secuencia codificante de rbcL de Nicotiana tabacum wt (tipo salvaje)

<400> 73

```

atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac      60
aaattgactt attatactcc tgagtaccaa accaaggata ctgatataatt ggcagcattc      120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa      180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac      240
agagggcgat gctaccgcat cgagcgtggt gttggagaaa aagatcaata tattgcttat      300
gtagcttacc ctttagacct ttttgaagaa gtttctgtta ccaacatggt tacttccatt      360
gtaggtaacg tatttggggt caaagccctg cgcgctctac gtctggaaga tctgcgaatc      420
cctcctgctt atgttaaaac tttccaaggt cgcctcatg ggatccaagt tgaagagat      480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta      540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact      600
aaagatgatg agaacgtgaa ctcaacaacca tttatgcggt ggagagatcg tttcttattt      660
tgtgccgaag cactttataa agcacaggct gaaacagggt aaatcaaagg gcattacttg      720
aatgctactg caggtagatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg      780
ggcgttccga tcgtaatgca tgactactta acggggggat tcaccgcaa tactagcttg      840
gctcattatt gccgagataa tgggtctact cttcacatcc accgtgcaat gcatgcggtt      900
attgatagac agaagaatca tggatccac ttccgggtat tagcaaaaagc gttacgtatg      960
tctggtggag atcatattca ctctggtagc gtagtaggta aacttgaagg tgaagagac     1020
ataacttttg gctttgttga tttactgctg gatgattttg ttgaacaaga tcgaagtgcg     1080
ggatatttatt tcaactcaaga ttgggtctct ttaccagggt ttctaccggt ggcttcagga     1140
ggatttcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta     1200
cagttcgggtg gaggaacttt aggacatcct tggggtaatg cgccagggtc cgtagctaatt     1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa     1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa     1380
gtatggaaaag agatcgtatt taattttgca gcagtggacg ttttgataa gtaa         1434
    
```

5

<210> 74

<211> 477

<212> PRT

<213> Artificial

<220>

<223> Basándose en la secuencia de la subunidad L de tipo salvaje traducida de Nicotiana tabacum wt (tipo salvaje)

<400> 74

```

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

10

ES 2 640 747 T3

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

<210> 75

<211> 1434

<212> ADN

5

<213> Artificial

<220>

<223> Basándose en la secuencia codificante de rbcL de Nicotiana tabacum wt (tipo salvaje)

<400> 75

atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac 60  
 aaattgactt attatactcc tgagtaccaa accaaggata ctgatataatt ggcagcattc 120  
 cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa 180  
 tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac 240  
 agaggggcgt gctaccgcat cgagcgtggt gttggagaaa aagatcaata tattgcttat 300  
 gtacgttacc ctttagacct ttttgaagaa ggttctggtta ccaacatggt tacttccatt 360  
 gtaggtaacg tatttggggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc 420  
 cctcctgctt atgttaaaac tttccaaggt cgcctcatg ggatccaagt tgaagagat 480  
 aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta 540  
 tctgctaata actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact 600  
 aaagatgatg agaacgtgaa ctcaacaacca tttatgctgt ggagagatcg tttcttattt 660  
 tgtgccgaag cactttataa agcacaggct gaaacagggt aaatcaaaagg gcattacttg 720  
 aatgctactg caggtacatg cgaagaatg atcaaaagag ctgtatttgc tagagaattg 780  
 ggcgtttccg tcgtaatgca tgactactta acggggggat tcaccgcaaa tactagcttg 840  
 gctcattatg gccgagataa tgggtctact cttcacatcc accgtgcaat gcattgcggt 900  
 attgatagac agaagaatca tgggtatccc ttccgggtat tagcaaaaagc gttacgtatg 960  
 tctggtggag atcatattca ctctgttacc gtatgtaggta aacttgaagg tgaagagac 1020  
 ataacttttg gctttgttga tttactgcgt gatgatattt ttgaacaaga tcgaagtcgc 1080  
 ggtatttatt tcaactcaaga ttgggtctct ttaccaggtg ttctaccctg ggcttcagga 1140  
 ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta 1200

ES 2 640 747 T3

cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccagggtgc cgtagccta 1260  
 cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa 1320  
 ggtaatgaaa ttattcgcga ggcttgcaaa tggagccccg aactagctgc tgcttggtaa 1380  
 gtatggaaa agatcgtatt taattttgca gcagtgagacg ttttgataa gtaa 1434

<210> 76

<211> 477

<212> PRT

5

<213> Artificial

<220>

<223> Basándose en la subunidad L traducida de Nicotiana tabacum tipo salvaje

<400> 76

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly  
 1 5 10 15  
 Val Lys Glu Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys  
 20 25 30  
 Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val  
 35 40 45  
 Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly  
 50 55 60  
 Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr  
 65 70 75 80  
 Arg Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln  
 85 90 95  
 Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser  
 100 105 110  
 Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys  
 115 120 125  
 Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
 130 135 140  
 Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
 145 150 155 160  
 Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
 165 170 175  
 Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
 180 185 190  
 Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
 195 200 205  
 Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
 210 215 220  
 Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
 225 230 235 240  
 Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
 245 250 255  
 Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
 260 265 270  
 Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
 275 280 285  
 Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
 290 295 300  
 Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
 305 310 315 320  
 Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
 325 330 335  
 Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
 340 345 350  
 Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
 355 360 365  
 Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
 370 375 380  
 Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
 385 390 395 400  
 Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
 405 410 415  
 Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn

ES 2 640 747 T3

420 425 430  
 Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala  
 435 440 445  
 Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
 450 455 460  
 Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys  
 465 470 475

<210> 77

<211> 1434

<212> ADN

5 <213> Artificial

<220>

<223> Basándose en la secuencia codificante de rbcL de *Nicotiana tabacum* wt (tipo salvaje)

<400> 77

```

atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac 60
aaattggggtt attatactcc tgagtaccaa accaaggata ctgatataatt ggcagcattc 120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa 180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac 240
aaaggcgcgt gctaccgcat cgagcgtggt gttggagaaa aagatcaata tattgcttat 300
gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatggt tacttccatt 360
gtaggtaacg tatttggggt caaagccctg cgcgctctac gtctggaaga tctgcgaatc 420
cctcctgctt atgttaaac tttccaaggt cgcctcatg ggatccaagt tgaagagat 480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa atgggggtta 540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact 600
aaagatgatg agaacgtgaa ctcaacaacca tttatgctgt ggagagatcg tttcttattt 660
tgtgccgaag cactttataa agcacaggct gaaacagggt aaatcaaagg gcattacttg 720
aatgctactg caggtacatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg 780
ggcgttccga tcgtaatgca tgactactta acggggggat tcaccgcaa tactagcttg 840
gctcattatt gccgagataa tggctactt cttcacatcc accgtgcaat gcatgcggtt 900
attgatagac agaagaatca tggatccac ttccgggtat tagcaaaaagc gtacgtatg 960
tctggtggag atcatattca ctctggtacc gatgtaggta aacttgaagg tgaagagagc 1020
ataactttgg gctttggtga tttactcgt gatgattttg ttgaacaaga tcgaagtgcg 1080
ggtattttat tcaactcaaga ttgggtctct ttaccaggtg ttctaccctg ggcttcagga 1140
ggtattcacg tttggcatat gcctgctctg accgatctct ttggggatga ttccgtacta 1200
cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccagggtg cgtagctaat 1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa 1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa 1380
gtagtgaaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa 1434
    
```

10 <210> 78

<211> 477

<212> PRT

<213> Artificial

<220>

15 <223> Basándose en la subunidad L traducida de *Nicotiana tabacum* tipo salvaje

<400> 78

```

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

ES 2 640 747 T3

Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr  
130 135 140  
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp  
145 150 155 160  
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro  
165 170 175  
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys  
180 185 190  
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser  
195 200 205  
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala  
210 215 220  
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu  
225 230 235 240  
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe  
245 250 255  
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly  
260 265 270  
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly  
275 280 285  
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln  
290 295 300  
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met  
305 310 315 320  
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu  
325 330 335  
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp  
340 345 350  
Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp  
355 360 365  
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val  
370 375 380  
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu  
385 390 395 400  
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly  
405 410 415  
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn  
420 425 430  
Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala  
435 440 445  
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu  
450 455 460  
Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys  
465 470 475

## REIVINDICACIONES

- 5 1. Un método para generar una proteína Rubisco que tiene una propiedad funcional mejorada seleccionada del grupo que consiste en: mejora de la eficiencia cinética de Rubisco, aumento de la especificidad de Rubisco para el dióxido de carbono con respecto a oxígeno, especificidad alterada de la Rubisco para uno o más sustratos, especificidad alterada de Rubisco para uno o más productos y un intervalo de temperatura eficaz alterado para la catálisis de Rubisco, comprendiendo dicho método:
- 10 (a) identificar al menos un Residuo de aminoácido Diana en una primera proteína Rubisco, en donde dicho Residuo de aminoácido Diana está asociado a dicha propiedad funcional, en donde los Residuos de aminoácido Diana de la proteína de Rubisco son residuos de la primera envoltura que se coordinan directamente con el centro de reacción o residuos de la segunda envoltura que se coordinan directamente con uno o más residuos de la primera envoltura, y dichos residuos de la primera envoltura de Rubisco se seleccionan del grupo que consiste en: Glu60, ASN123, LYS175, LYS177, KCX201, Asp203, GLU204 e HIS294, en donde la numeración de los residuos de Rubisco de espinaca, y KCX201 denota LYS201 carbamilada;
- 15 (b) comparar al menos una segunda proteína Rubisco homóloga de la misma o diferente ramas filogenéticas que la primera proteína Rubisco con la primera proteína Rubisco e identificar al menos un Residuo de aminoácido Variante entre la primera proteína Rubisco y la segunda proteína Rubisco, en donde los residuos de la segunda proteína Rubisco utilizados para la comparación con la primera proteína Rubisco comprenden todos los residuos directamente coordinados con el sitio activo de una proteína Rubisco, interactuando todos los residuos con el centro reactivo del sustrato o especies de reacción intermedias de una proteína Rubisco, y otros residuos a una distancia aproximada de entre 3 y 28 Å de cualquier átomo de sustrato o especies de reacción intermedias del sitio activo de la proteína Rubisco o un subconjunto de tales residuos;
- 20 (c) seleccionar al menos un Residuo de aminoácido Candidato entre los Residuos de aminoácidos Variantes identificados en (b) basándose en la estimación de la proximidad espacial de los Residuos Variantes a los Residuos Diana y estimar y clasificar su capacidad para influir en la electrostática y la orientación de los Residuos Diana y modular la propiedad funcional de la proteína formada;
- 25 (d) generar al menos una proteína Mutante Rubisco Candidata *in vitro* en la que dicho al menos un Residuo de aminoácido Candidato de la segunda proteína Rubisco sustituye un residuo correspondiente en la primera proteína de Rubisco; y
- 30 (e) detectar dicha al menos una proteína Rubisco Mutante Candidata producida en (d) para identificar una proteína Rubisco que tiene una propiedad funcional mejorada.
- 35 2. El método de la reivindicación 1, en donde la etapa (a) y la etapa (b) se llevan a cabo simultáneamente.
3. El método de la reivindicación 1, en donde la etapa (b) se realiza antes de la etapa (a).
4. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde la etapa (d) comprende producir al menos una proteína Rubisco Mutante Candidata utilizando el al menos un Residuo de aminoácido Candidato de la segunda proteína Rubisco para sustituir un residuo correspondiente en una proteína Rubisco homóloga o en proteínas Rubisco homólogas distintas de o además de la primera proteína Rubisco.
- 40 5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en donde la etapa (d) comprende la generación de al menos una proteína Rubisco Mutante Candidata en la que al menos dos Residuos de aminoácidos Candidato de la segunda proteína Rubisco sustituyen los residuos correspondientes en la primera proteína Rubisco y/o en una proteína Rubisco homóloga distinta de la primera proteína Rubisco.
- 45 6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en donde el al menos un Residuo de aminoácido Diana está contenido en un conjunto de proteínas que contiene la primera proteína Rubisco.
- 50 7. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en donde el al menos un Residuo de aminoácido Diana consiste en al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 12, al menos 15, al menos 20, al menos 30 o al menos 50 Residuos de aminoácido Diana.
- 55 8. El método de la reivindicación 1, en donde los Residuos de aminoácido Diana de la proteína Rubisco están en el dominio N-terminal de la Subunidad Grande de Rubisco.
- 60 9. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en donde la etapa (c) comprende seleccionar al menos un Residuo de aminoácido Candidato Divergente y/o al menos un Residuo de aminoácido Candidato Alternativo, en lugar de al menos un Residuo de aminoácido Candidato, a partir de los Residuos de aminoácidos Variantes identificados en (b), en donde dicho Residuo de ácido amino Candidato Divergente o

Residuo Candidato Divergente (DCR) se refiere a un residuo de aminoácido que se selecciona de una pluralidad de Residuos Variantes y se cree que es capaz de influir estéricamente y/o electrostáticamente en uno o más Residuos Diana, y por lo tanto influir en la función de la proteína Rubisco mediada por uno o más Residuos Diana; y en donde dicho Residuo de aminoácido Candidato Alternativo o Residuo Candidato Alternativo (ACR) se refiere a un residuo de aminoácido sustituido en la posición de un Residuo Candidato que es expresado en una segunda proteína Rubisco, pero que es no es el aminoácido que se expresa en la secuencia consenso de la segunda proteína Rubisco y que se cree que es capaz de influir en uno o más Residuos Diana estéricamente y/o electrostáticamente, y de ese modo influir en la función de la proteína Rubisco mediada por los uno o más Residuos Diana.

- 5
  - 10
10. El método de una cualquiera de las reivindicaciones 1 a 9, que comprende la etapa adicional de realizar una evolución dirigida de dicha proteína Rubisco que tiene dicha propiedad funcional mejorada y escrutar sus productos.

Estrategia de Reestructuración de Rubisco Refinada

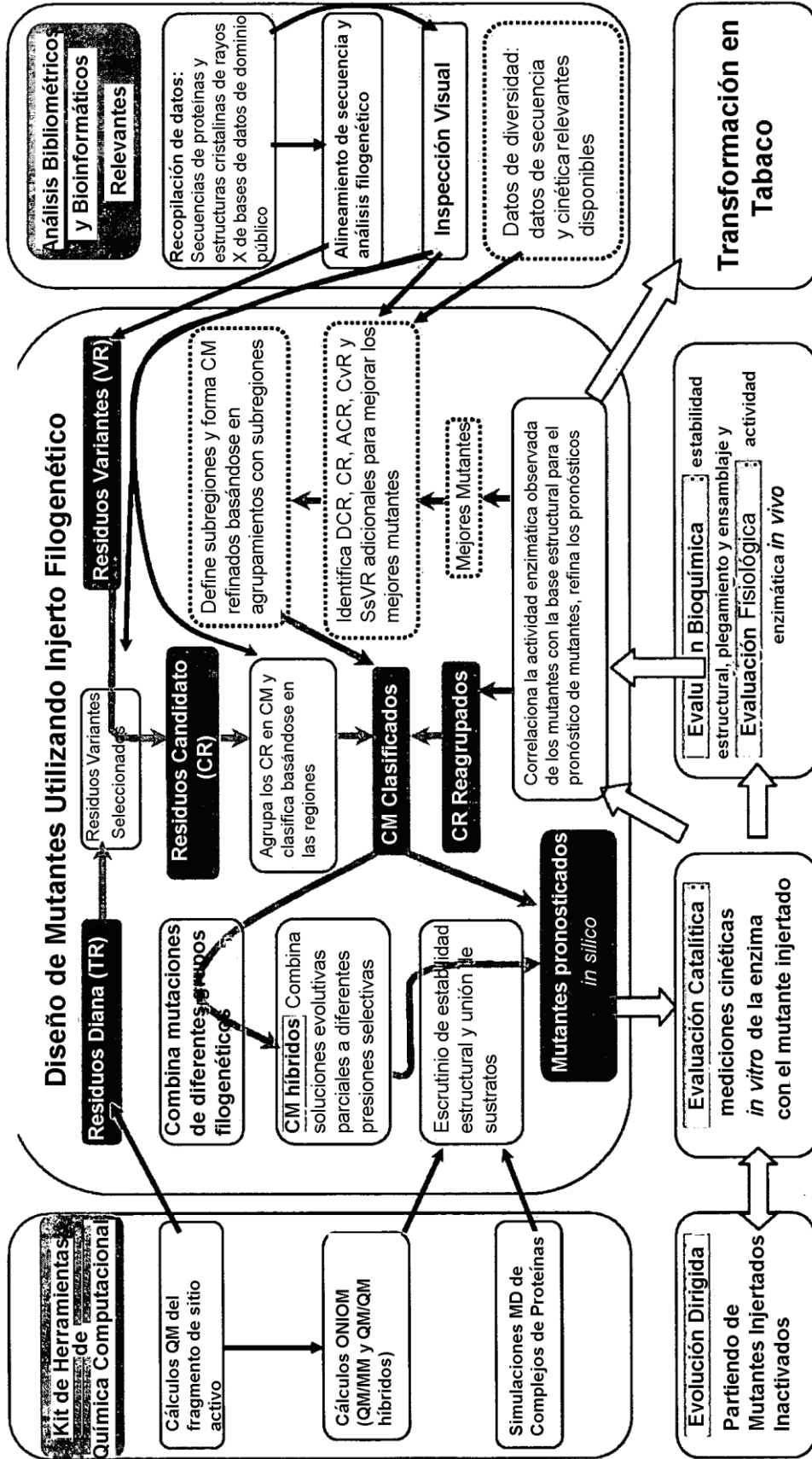


Diagrama de Flujo 1

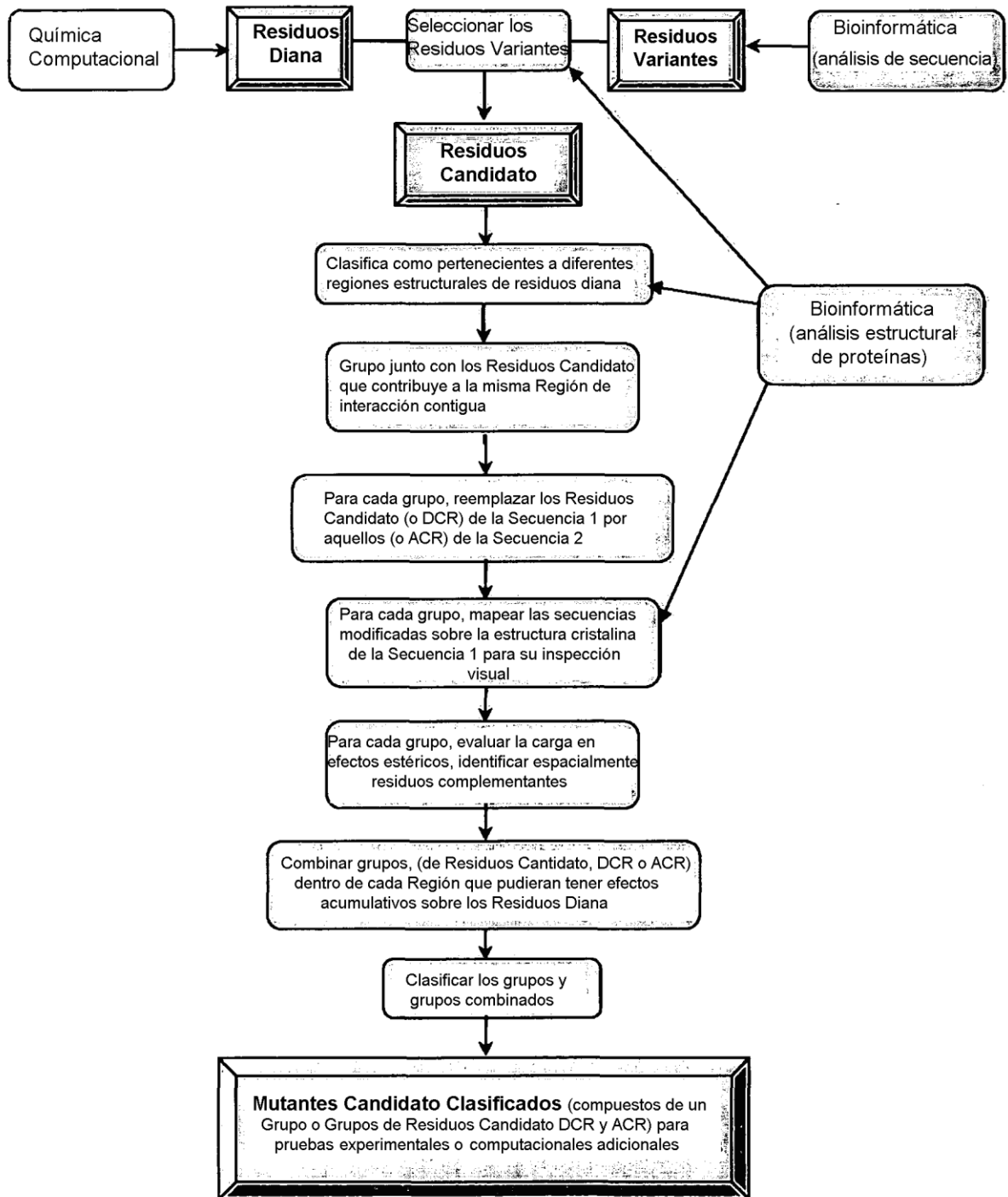


Diagrama de flujo 2

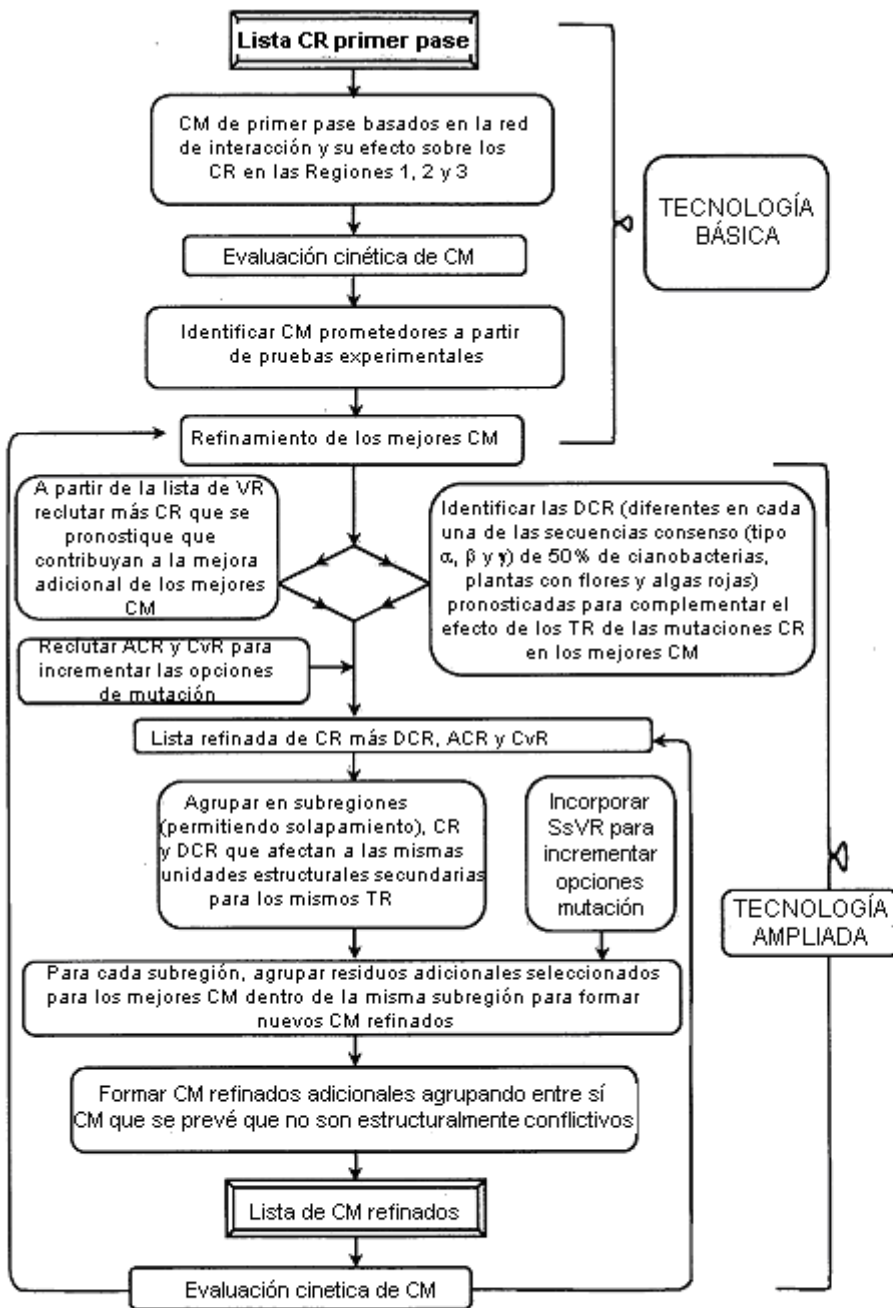


Diagrama de flujo 3

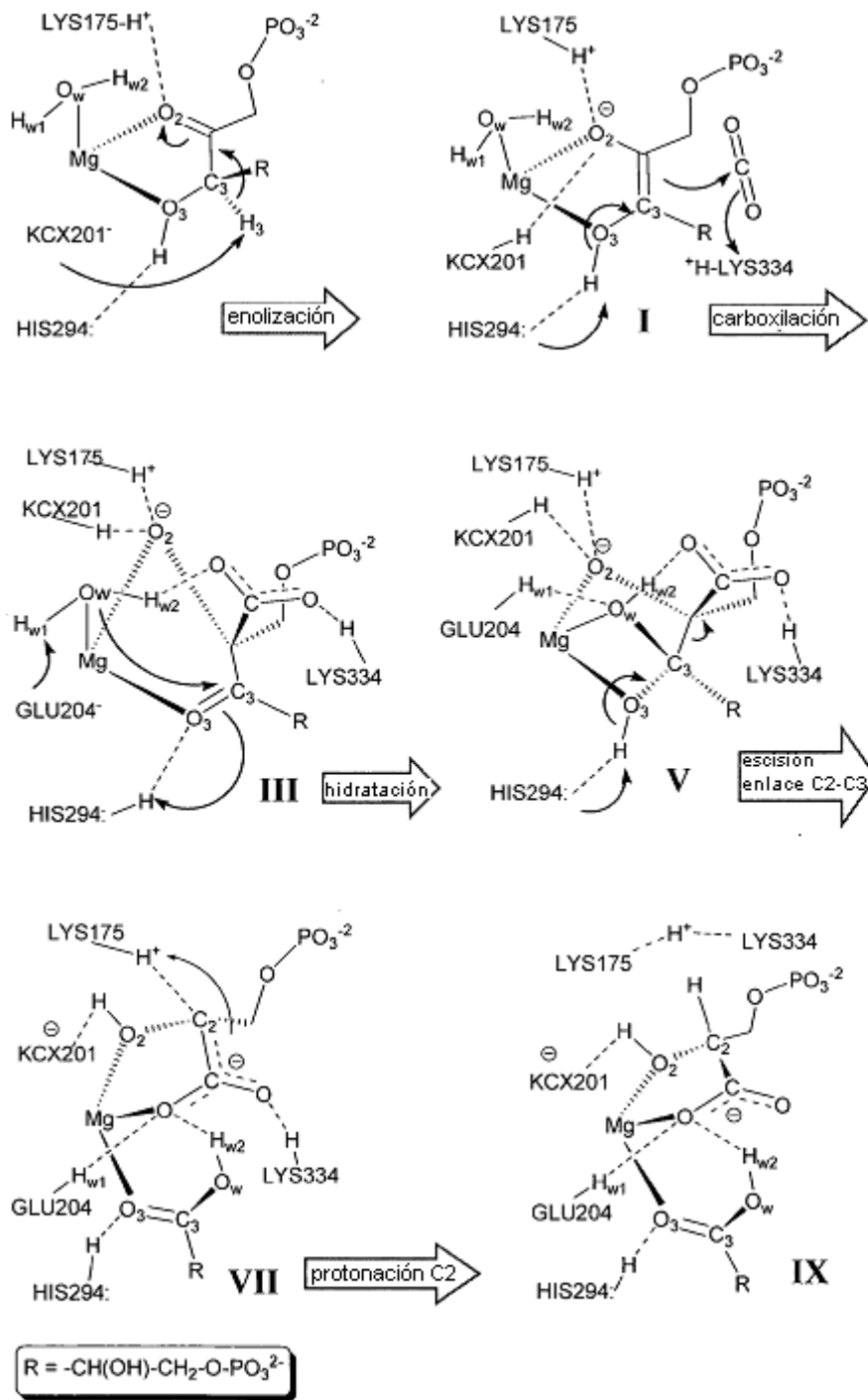


Figura 1



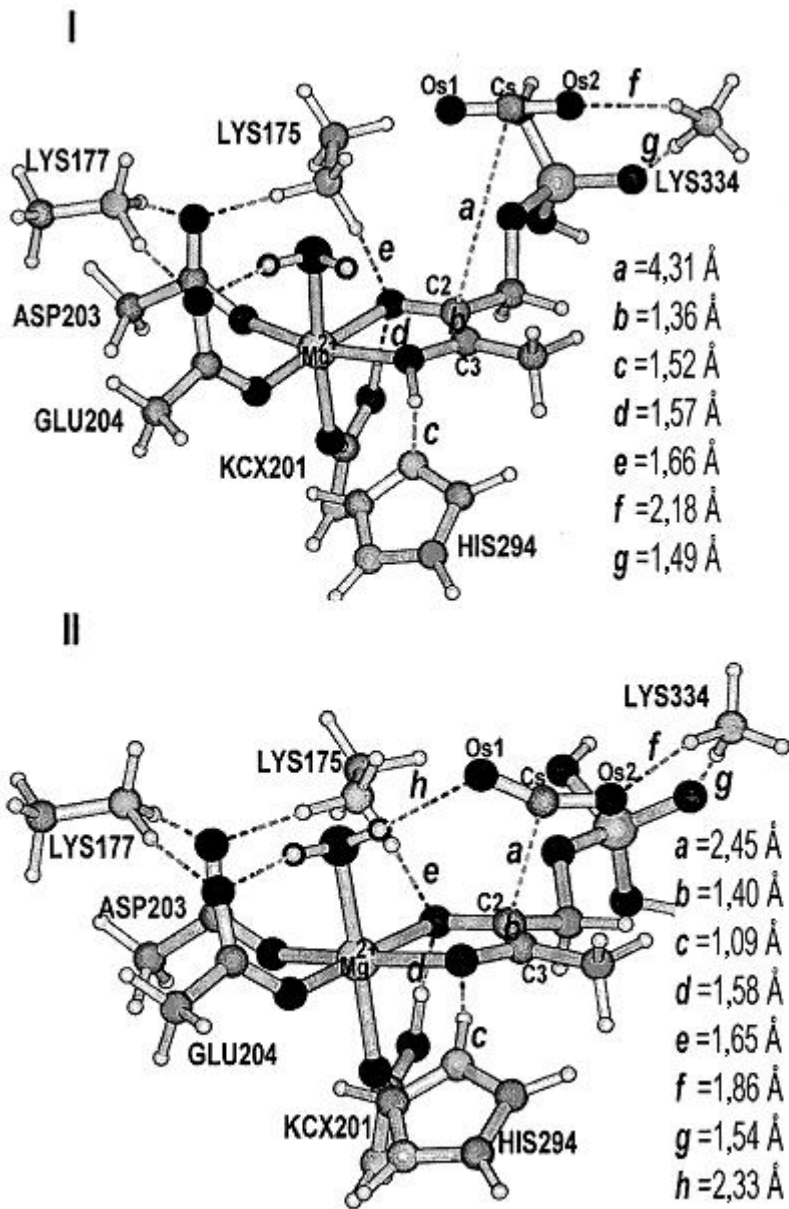


Figura 3A

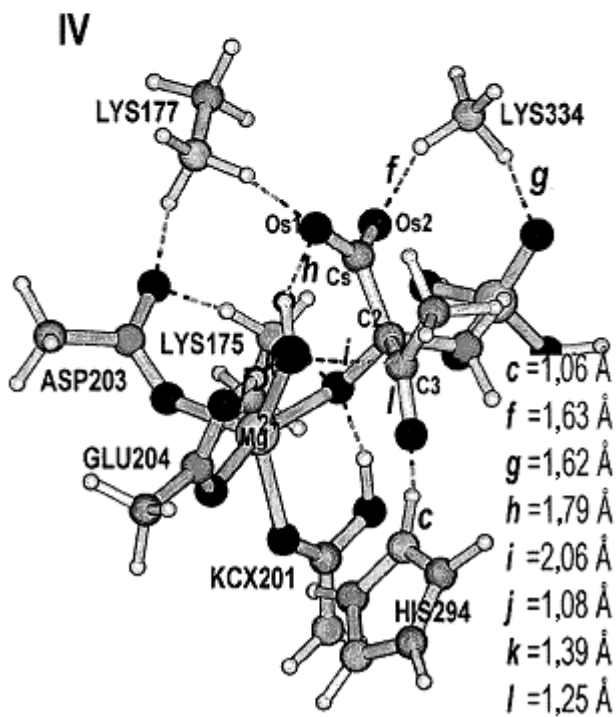
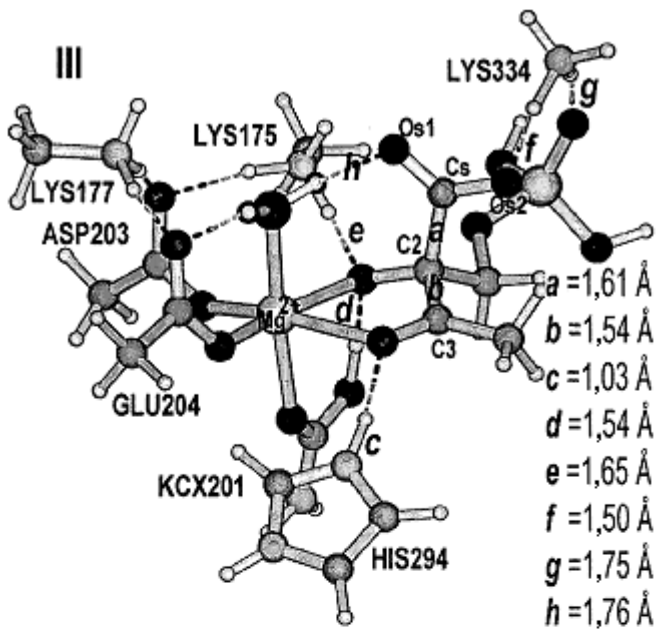
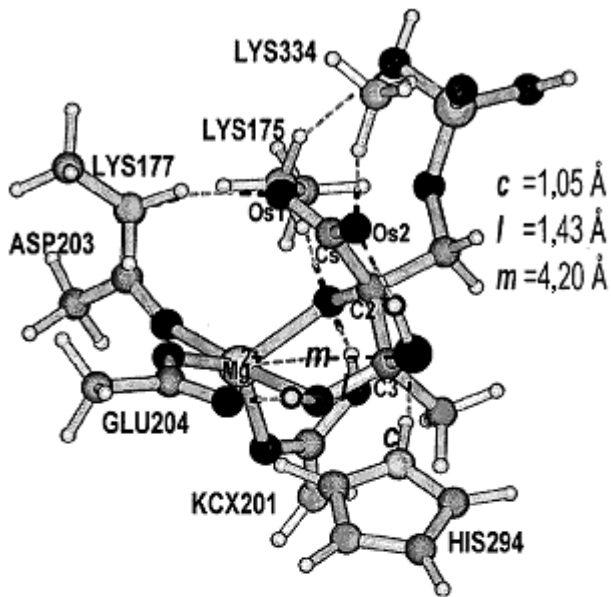


Figura 3B

V



VI

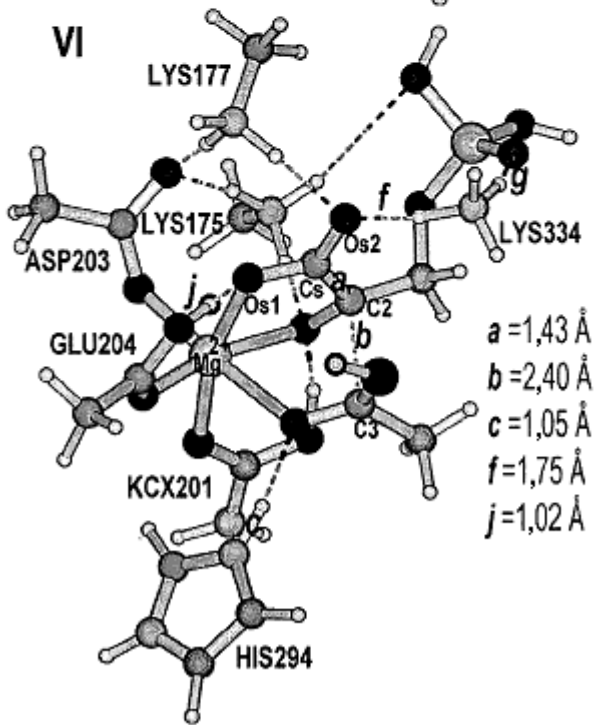


Figura 3C

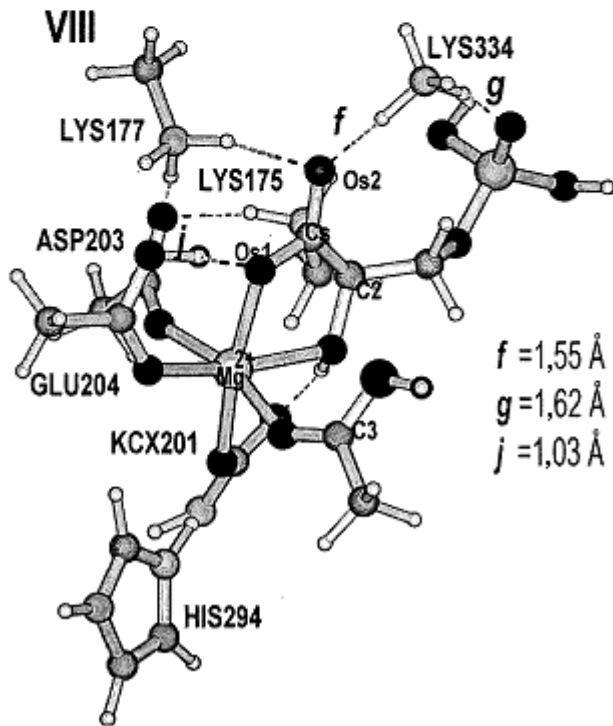
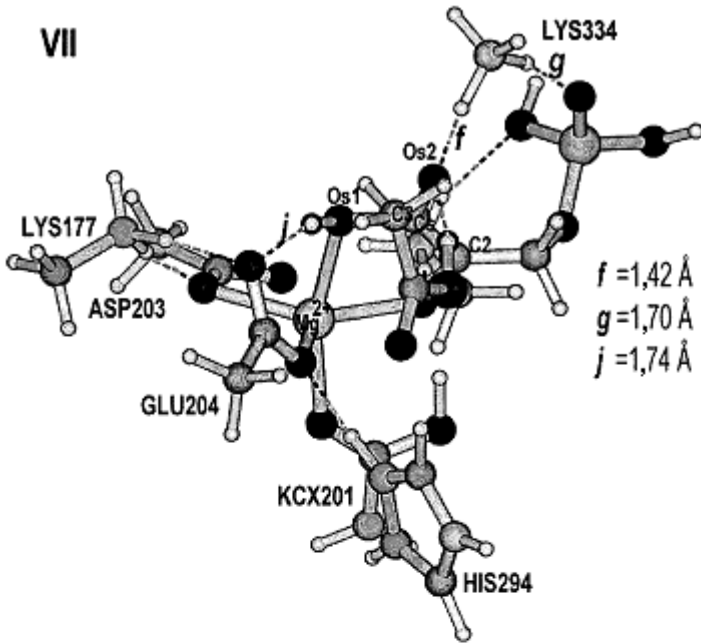


Figura 3D

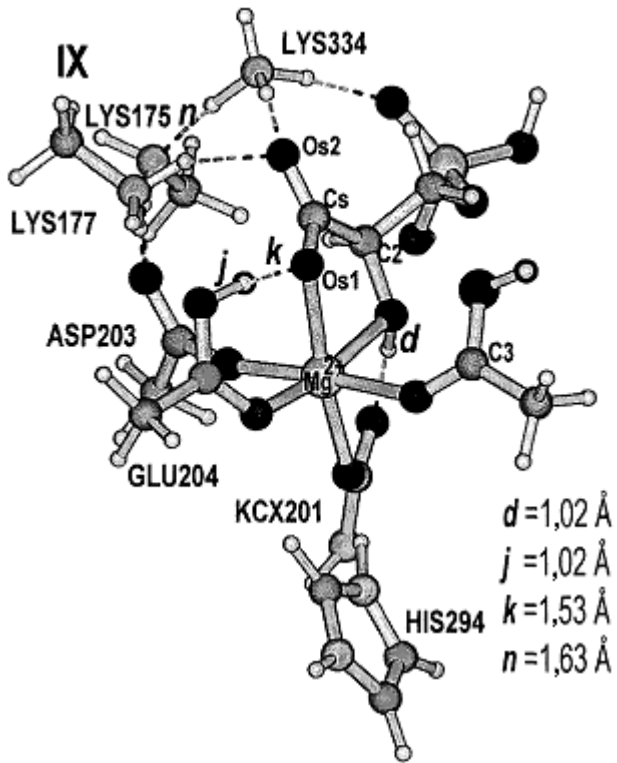


Figura 3E

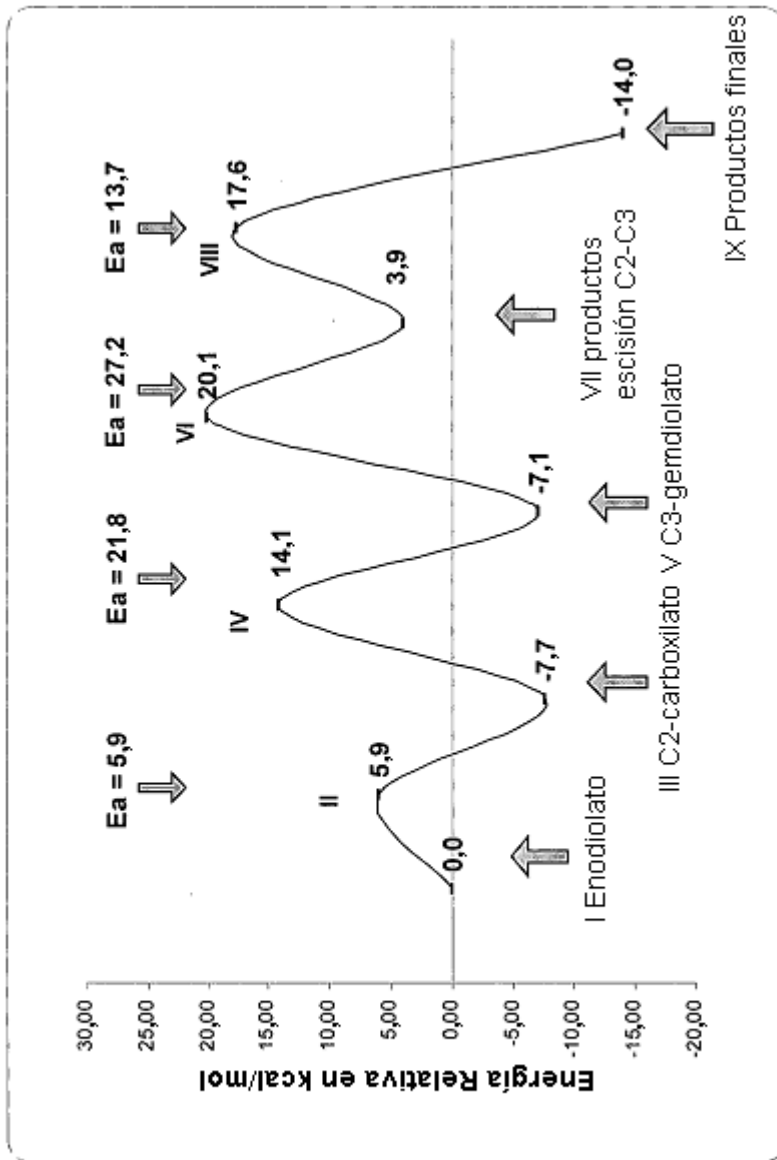


Figura 4

		10	17	26	40	50	60
		..... ..... ..... ..... ..... ..... ..... .....					
1	protista glaucophyta	MSSQARTQR	A~GFK~AGV	KDYRL~TYTT	PEYTPKRTDI	LAAFRMTFQP	GVPPECANA
1	cns50*(9) protista rhodophyta	MSPVPERTR	IKNERYESCV	IFYAKMGYWD	FDYVVKDIDV	LALFRVTFQP	GVDPVEASAA
1	cns50*(11) eubacteria cyanobacteria	MSYAQKCoO	KUGYp~AGV	KDYRL~TYTT	FDYTPKDTDL	LAAFRMTFQP	GVPSEBAGAA
1	plantae anthocerotophyta	MSPQITKAG	V~GFK~AGV	KDYRL~TYTT	FDYETKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	cns50*(28) plantae bryophyta	MSPQITKAG	V~GFK~AGV	KDYRL~TYTT	FDYQTKRTDI	LAAFRMTFQA	GVPPEEAGAA
1	cns50*(4) plantae charophyta	MSPQITKAG	A~GFK~AGV	KDYRL~TYTT	FDYETKRTDI	LAAFRMTFQA	GVPPEEAGAA
1	cns50*(4) plantae chlorophyta	MSPQITKAG	A~GFK~AGV	KDYRL~TYTT	FDYVVKDIDV	LAAFRMTFQP	GVPPECGAA
1	plantae coniferophyta	MSPKTTKAS	V~GFK~AGV	KDYRL~TYTT	PEYQTKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	plantae equisetophyta	MSPKTTKAS	V~GFK~AGV	KDYRL~TYTT	PEYQTKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	plantae gnetophyta	MSPKTTKAS	V~GFK~AGV	KDYRL~TYTT	PEYQTKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	cns50*(134) plantae magnoliophyta	MSPQITKAS	V~GFK~AGV	KDYKL~TYTT	FDYETKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	cns50*(2) plantae pinophyta	MSPKTTKAS	V~GFK~AGV	KDYRL~TYTT	PEYQTKDTDI	LAAFRMTFQP	GVPPEEAGAA
1	cns50*(2) plantae pteridophyta	MSPQITKTG	I~GFK~AGV	KDYRL~TYTT	FDYETKDTDI	LAAFRMTFQP	GVPPEEAGAA
		70	80	90	100	110	120
		..... ..... ..... ..... ..... ..... ..... .....					
57	protista glaucophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRSYG	FEFVPGBNQ	YICYVAYPLD	LFEEGSVTNN
116							
61	cns50*(9) protista rhodophyta	VAGESSTATW	TVVWTDLLTA	CDLYRAKAYK	VD~VPGNSDQ	YFAYIAYDID	LFEEGSIANL
58	cns50*(11) eubacteria cyanobacteria	VAAESSTGTW	TTVWTDLLTD	LDRYKGRCYD	IEFPVGBDNQ	YaAFIAYPLD	LFEEGSVTNV
57	plantae anthocerotophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVAGBNQ	YIAYVAYPLD	LFEEGSVTNN
57	cns50*(28) plantae bryophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEAVPGBENQ	YIAYVAYPLD	LFEEGSVTNL
57	cns50*(4) plantae charophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVAGBNQ	YIAYVAYPLD	LFEEGSVTNL
57	cns50*(4) plantae chlorophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVGBDNQ	YIAYVAYPLD	LFEEGSVTNL
57	plantae coniferophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVGBENQ	YIAYVAYPLD	LFEEGSVTNL
57	plantae equisetophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVGBDNQ	YIAYVAYPLD	LFEEGSVTNL
57	plantae gnetophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVGBDNQ	YIAYVAYPLD	LFEEGSVTNL
57	cns50*(134) plantae magnoliophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYH	IEFPVAGBNQ	YIAYVAYPLD	LFEEGSVTNN
57	cns50*(2) plantae pinophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVGBENQ	YIAYVAYPLD	LFEEGSVTNL
57	cns50*(2) plantae pteridophyta	VAAESSTGTW	TTVWTDGLTS	LDRYKGRCYD	IEFPVAGBNQ	YIAYVAYPLD	LFEEGSVTNN

Figure 5A



	250	256	270	276	280	290	296
<i>protista glaucophyta</i>	..... ..... ..... ..... ..... ..... ..... .....						
<i>cns50%(9) protista rhodophyta</i>	237 GHYLNATAPT	TEEMKRAEF	AEELDAPPIIM	HDYITAGFTS	NTTLARWCRD	NGPDLHIHRA	296
<i>cns50%(11) eubacteria cyanobacteria</i>	241 GHYLNATAAT	ME-NYERAEF	AKELGSVIIM	ID-LVIGYTA	IQTMATWARK	NDMILHLHRA	298
<i>plantae anthocerotophyta</i>	238 GHYLNVTAST	CEEMKRAEF	AKELGPPPIIM	HDFLTGGFTA	NTTLAKWCRD	NGLLLHIHRA	297
<i>cns50%(28) plantae bryophyta</i>	237 GHYLNATAGT	CEEMKRAEF	ARELGMPPIIM	HDYLTGGFTA	NTTLARYCRD	NGLLLHIHRA	296
<i>cns50%(4) plantae charophyta</i>	237 GHYLNATAGT	CEEMKRAQF	ARELGMPPIIM	HDYLTGGFTA	NTSLAHYCRD	NGLLLHIHRA	296
<i>cns50%(4) plantae chlorophyta</i>	237 GHYLNATAAT	CEEMKRAVC	AKELGVPIIM	HDYLTGGFTA	NTSLASYCRD	NGLLLHIHRA	296
<i>plantae coniferophyta</i>	237 GHYLNATACT	CEEMKRAIF	ARELGVPPIIM	HDYLTGGFTA	NTSLAHYCRD	NGLLLHIHRA	296
<i>plantae equisetophyta</i>	237 GHYLNATAGT	CEEMKRAVF	ARELGAPPIIM	HDYLTGGFTA	NTSLAFYCRD	NGLLLHIHRA	296
<i>plantae gnetophyta</i>	237 GHYLNATAGT	CEEMKRAVE	ARELGVPPIIM	HDYLTGGFTA	NTTLAHYCRD	NGLLLHIHRA	296
<i>cns50%(134) plantae magnoliophyta</i>	237 GHYLNATAGT	CEEMKRAVF	ARELGVPPIIM	HDYLTGGFTA	NTSLAHYCRD	NGLLLHIHRA	296
<i>cns50%(2) plantae pinophyta</i>	237 GHYLNATAGT	CEEMKRAIF	ARELGVPPIIM	HDYLTGGFTA	NTSLAHYCRD	NGLLLHIHRA	296
<i>cns50%(2) plantae pteridophyta</i>	237 GHYLNATAAT	CEEMKRAIF	ARELGAPPIIM	HDYLTGGFTA	NTSLAFYCRD	NGLLLHIHRA	296
	310	316	330	340	350	356	
<i>protista glaucophyta</i>	..... ..... ..... ..... ..... ..... ..... .....						
<i>cns50%(9) protista rhodophyta</i>	297 MHAVIDRQKN	HGIHFRVLAK	TLRMSGGDHL	HSCTVWGKLE	GDRAGTLGFV	DLMRDDHIEQ	356
<i>cns50%(11) eubacteria cyanobacteria</i>	299 GNSTYSRQKS	HGMNFRVICK	WMRMAGVDHI	HAGTVWGKLE	GDPIMI+GFY	NTLILLoHL-1	357
<i>plantae anthocerotophyta</i>	298 MHAVIDRQKN	HGIHFRVLAK	CLRLSGGDHL	HTGTVWGKLE	GERUITMGFV	DILRE-avec	356
<i>cns50%(28) plantae bryophyta</i>	297 MHAVIDRQKN	HGIHFRVLAK	ALRMSGGDHI	HSCTVWGKLE	GEREVTILGFV	DILRDDYIEK	356
<i>cns50%(4) plantae charophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GERQVTILGFV	DILRDDYIEK	356
<i>cns50%(4) plantae chlorophyta</i>	297 MHAVIDRQKN	HGIHFRVLAK	ALRMSGGDHI	HSCTVWGKLE	GEREVTILGFV	DILRDDYIEK	356
<i>plantae coniferophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GERDVTILGFV	DILRDDYIEK	356
<i>plantae equisetophyta</i>	297 MHAVIDRQKN	HGIHFRVLAK	ALRMSGGDHI	HTGTVWGKLE	GERDLTLGFV	DILRDDYIEK	356
<i>plantae gnetophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GEREITILGFV	DILRDDYIEK	356
<i>cns50%(134) plantae magnoliophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GEREITILGFV	DILRDDYIEK	356
<i>cns50%(2) plantae pinophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GERDVTILGFV	DILRDDYIEK	356
<i>cns50%(2) plantae pteridophyta</i>	297 MHAVIDRQKN	HGMHFRVLAK	ALRMSGGDHI	HAGTVWGKLE	GEREVTILGFV	DILRDDYIEK	356

Figura 5C



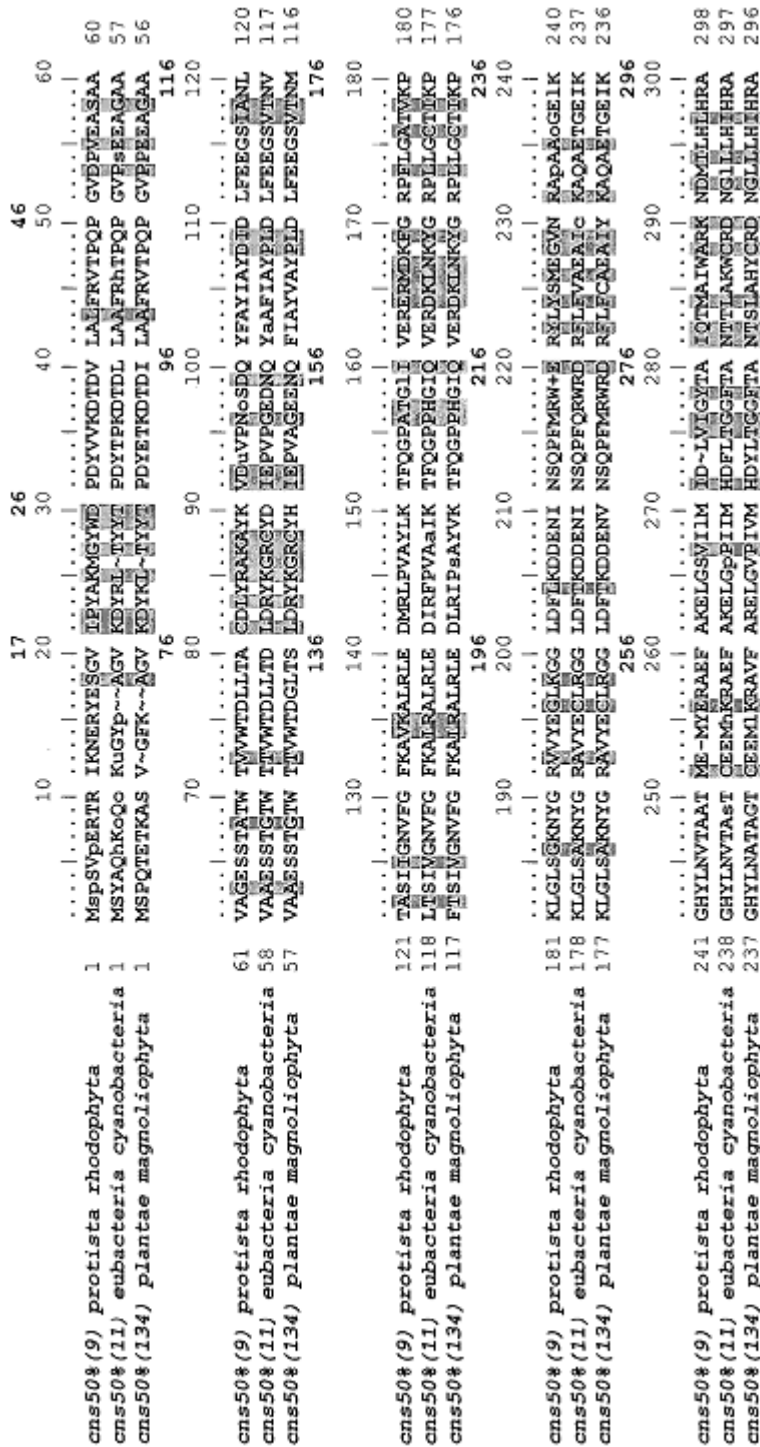


Figura 6A

	316	320	330	340	350	356	360
<i>cns50*(9) protista rhodophyta</i>	..... ..... ..... ..... ..... ..... ..... .....						
299	GNSFYSRQKs	HGMNFRVICK	WNRVACVQDHI	HAGTVVGGKLE	GDIMM+GFV	NTLLLoHL-1	357
<i>cns50*(11) eubacteria cyanobacteria</i>	298	MHAVYDRQKN	HGIFKRVLAK	CLRLSGGDHL	HGTGTVVGGKLE	GERKUTMGFV	356
<i>cns50*(134) plantae magnoliophyta</i>	297	MHAVYDRQKN	HGMNFRVICK	ALRMVSGGDHI	HAGTVVGGKLE	GERKUTMGFV	356
		376		396		416	
<i>cns50*(9) protista rhodophyta</i>	370	380	390	400	410	420	
358	NLRGQIFFEQ	DWASLRKRVTP	VASGGIHCSQ	MHQLLDYLDG	DWVLFQGGGT	LGHFPGIQAG	417
<i>cns50*(11) eubacteria cyanobacteria</i>	357	DRSRGIFFIQ	DWASMGVMA	VASGGIHVWH	MPALVELEFGD	DVYLQFGGCT	416
<i>cns50*(134) plantae magnoliophyta</i>	357	DRSRGIYFFIQ	DWVSLGVLVLP	VASGGIHVWH	MPALTELEFGD	DVYLQFGGCT	416
		436		456		480	
<i>cns50*(9) protista rhodophyta</i>	418	ATANKRVALES	MVQARNEGRD	VVAVGPGQIIR	DAAKTCGGPLQ	TALDLMKQIT	477
<i>cns50*(11) eubacteria cyanobacteria</i>	417	ATANKRVALEA	CVQARNEGRC	LVREKSDIIR	DAAKWSPFLA	IACELWKEIK	475
<i>cns50*(134) plantae magnoliophyta</i>	417	AVANKRVALEA	CVQARNEGRD	LAREGNEIIR	EASKWSPFLA	AACEVWKEIK	475

Figura 6B

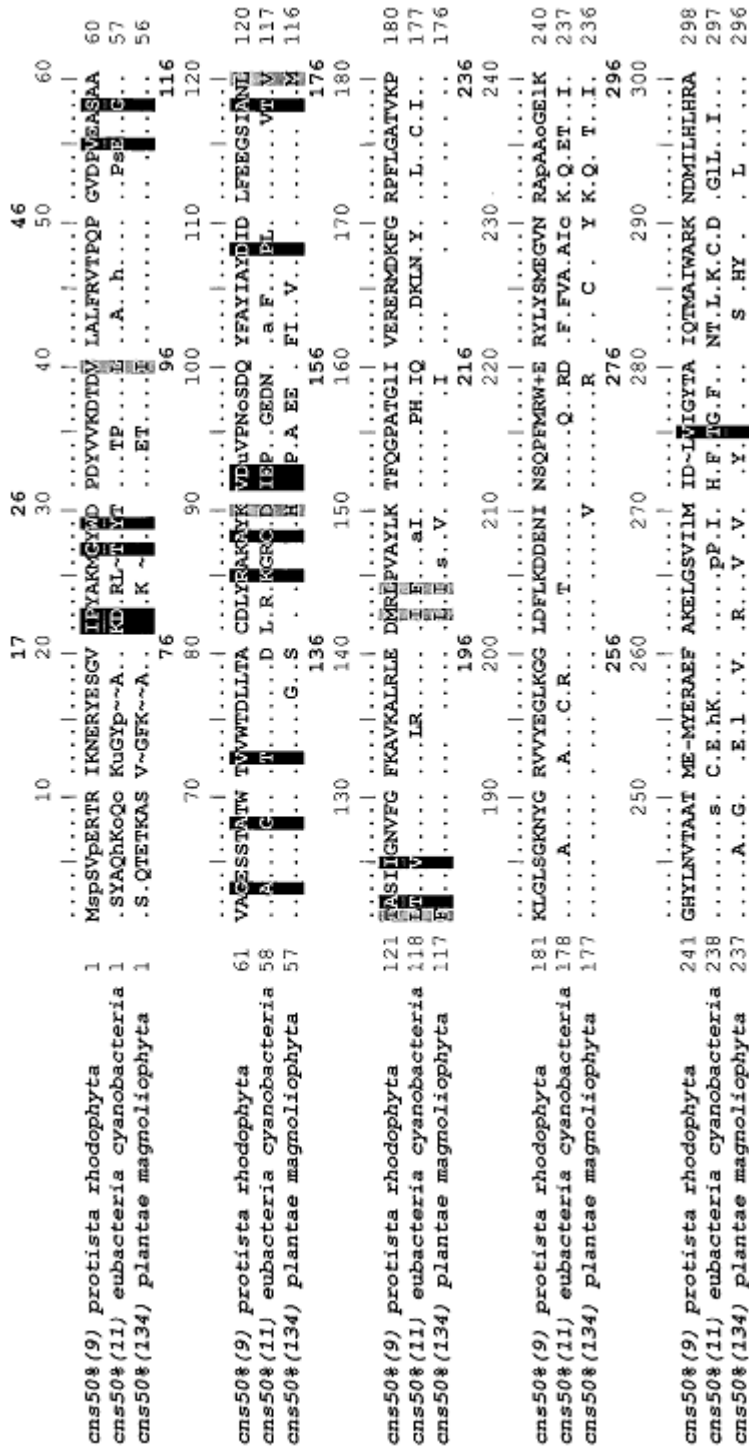


Figura 7A

	310	316	320	330	336	340	350	356	360
<i>cns50*(9) protista rhodophyta</i>	299	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(11) eubacteria cyanobacteria</i>	298	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(134) plantae magnoliophyta</i>	297	.....	.....	.....	.....	.....	.....	.....	.....
		370	376	380	390	400	410	416	420
<i>cns50*(9) protista rhodophyta</i>	358	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(11) eubacteria cyanobacteria</i>	357	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(134) plantae magnoliophyta</i>	357	.....	.....	.....	.....	.....	.....	.....	.....
		430	436	440	450	460	470	480	480
<i>cns50*(9) protista rhodophyta</i>	418	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(11) eubacteria cyanobacteria</i>	417	.....	.....	.....	.....	.....	.....	.....	.....
<i>cns50*(134) plantae magnoliophyta</i>	417	.....	.....	.....	.....	.....	.....	.....	.....

Figura 7B

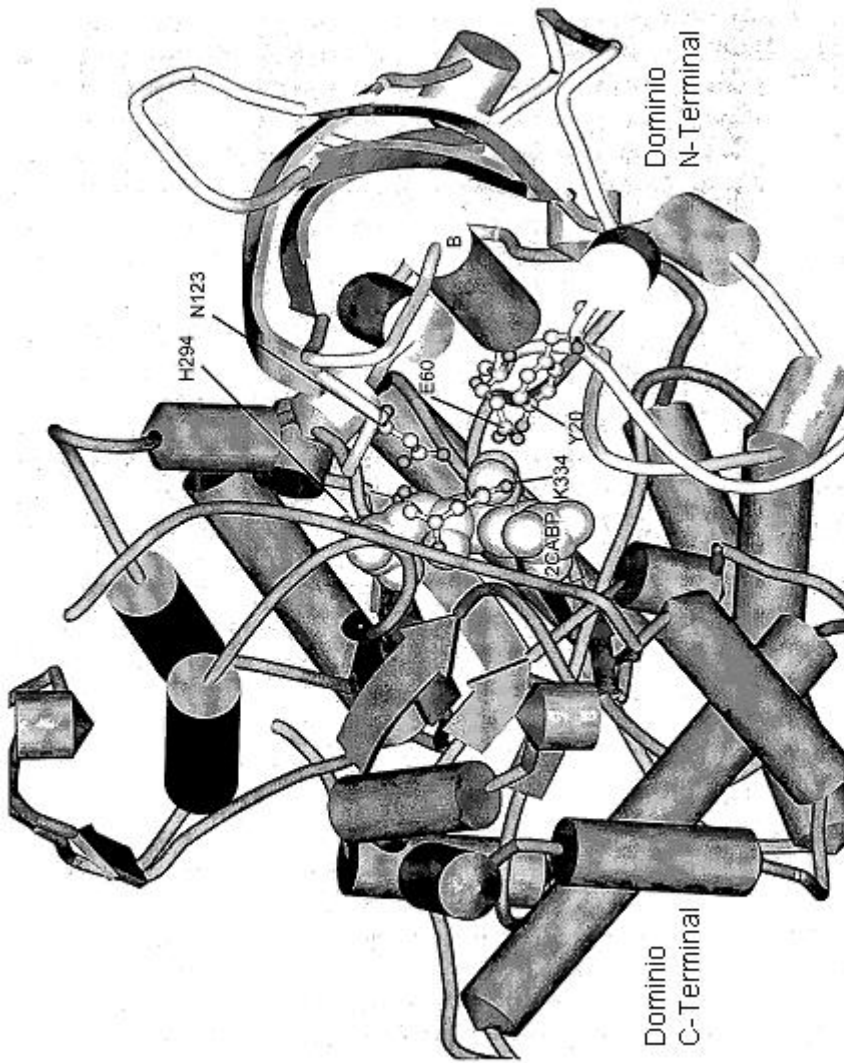


Figura 8

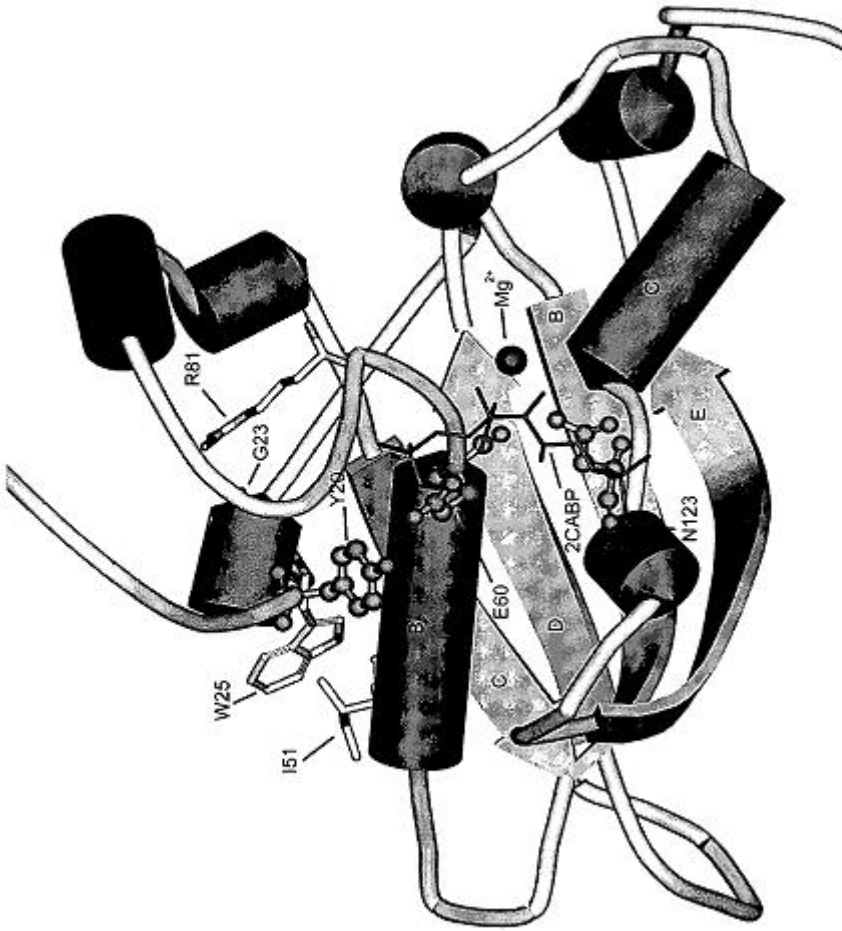


Figura 9

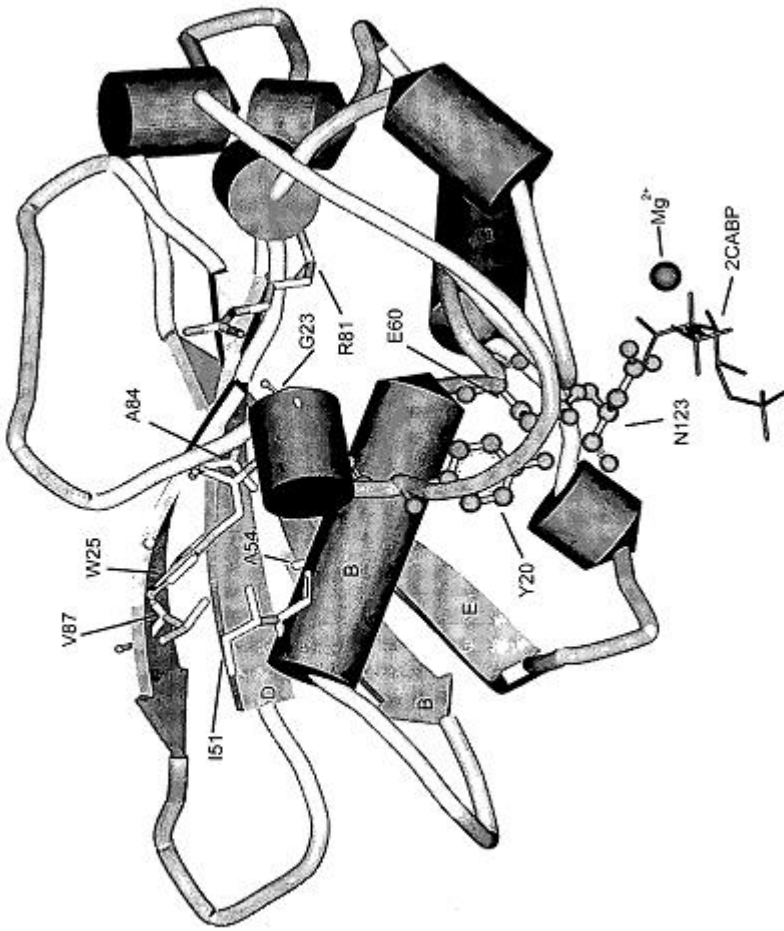


Figura 10

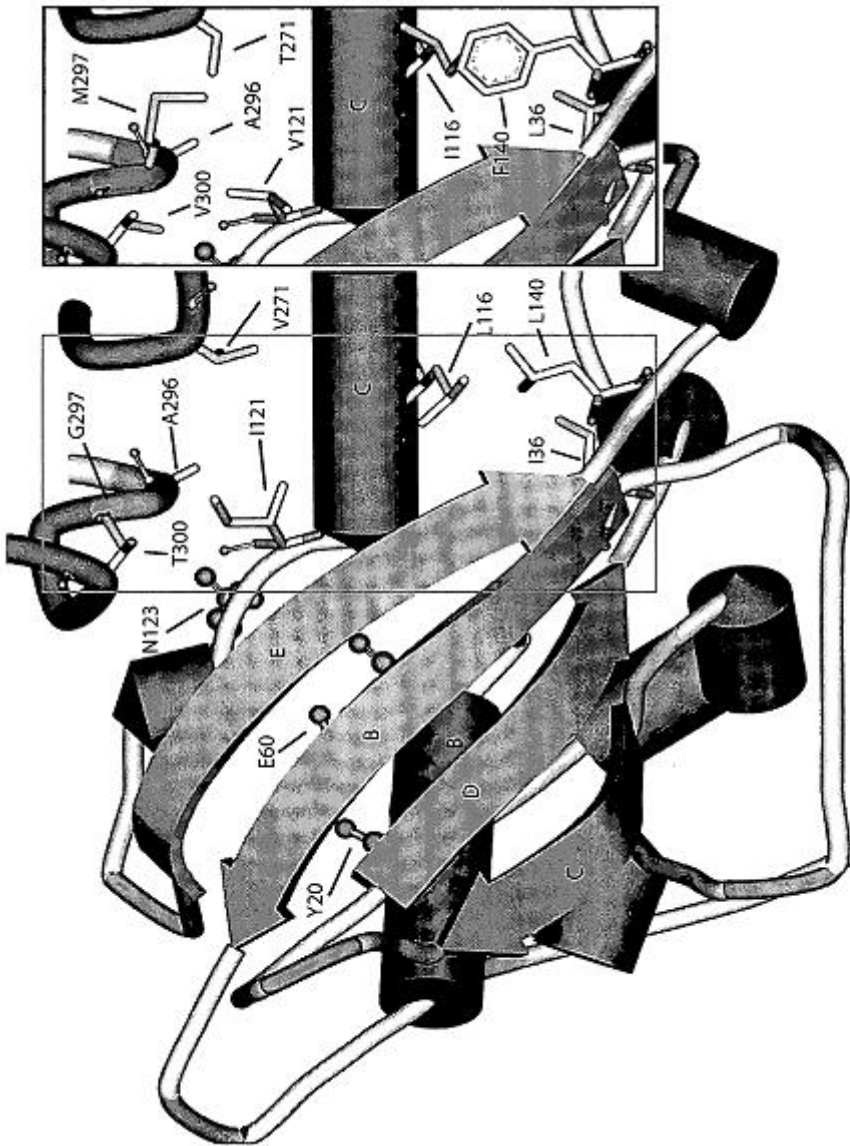


Figura 11

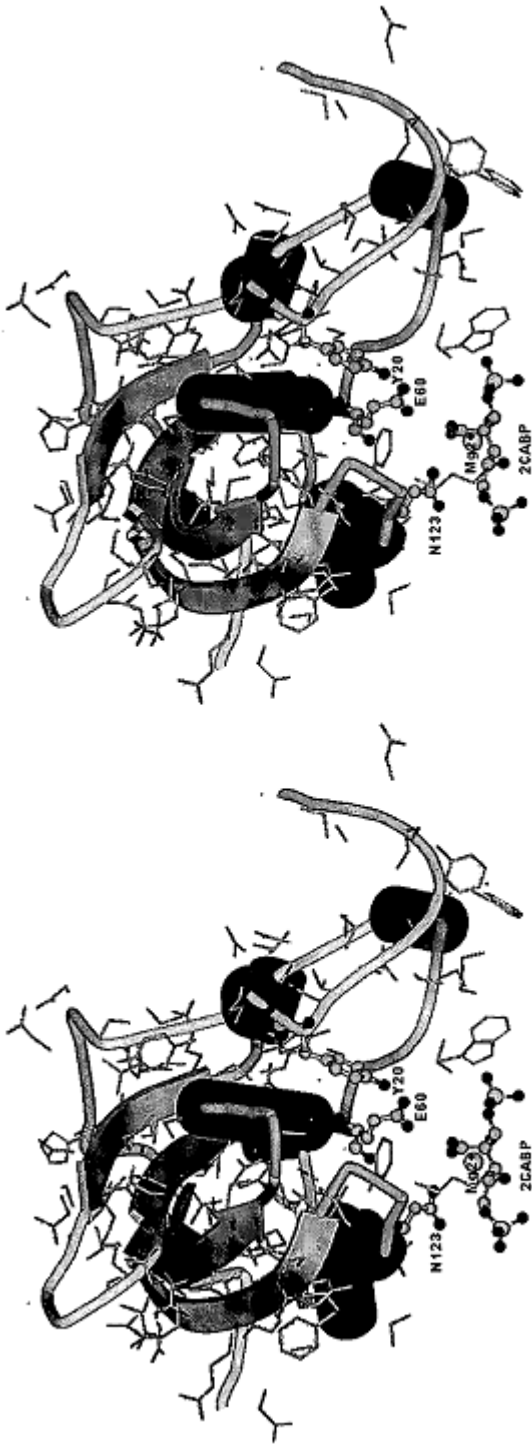


Figura 12

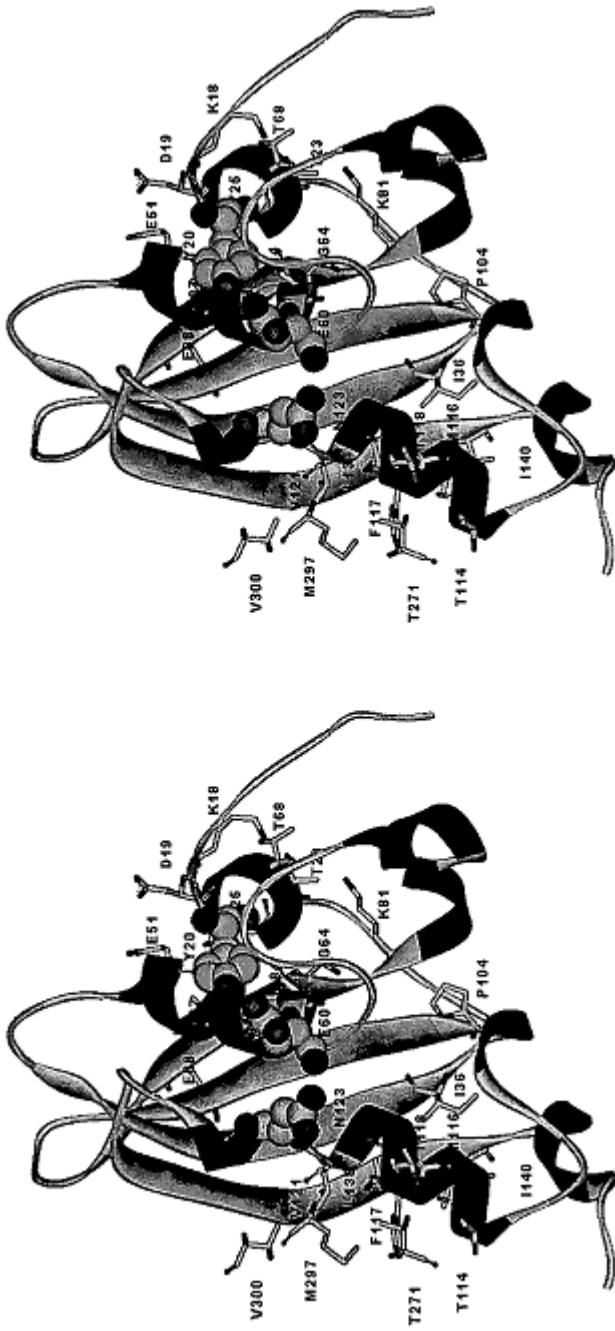


Figura 13

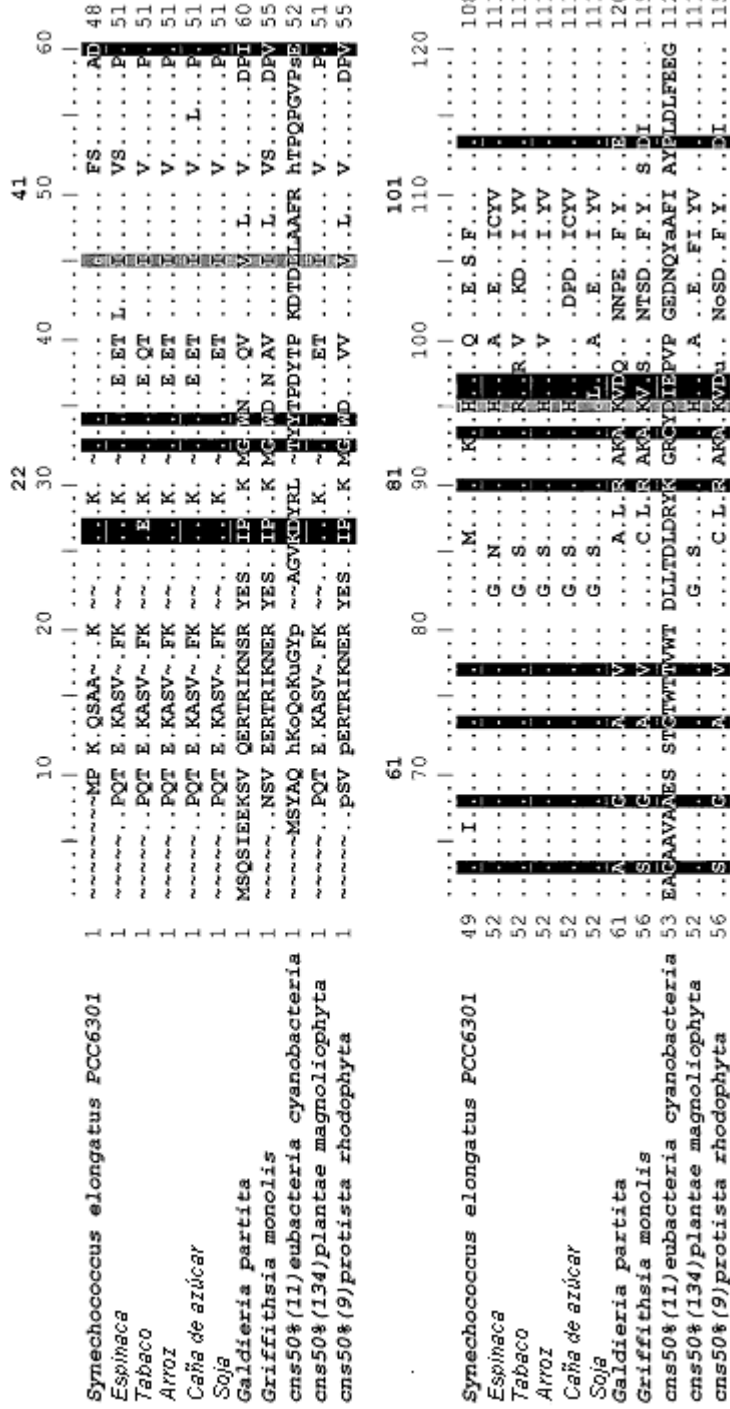


Figura 14A

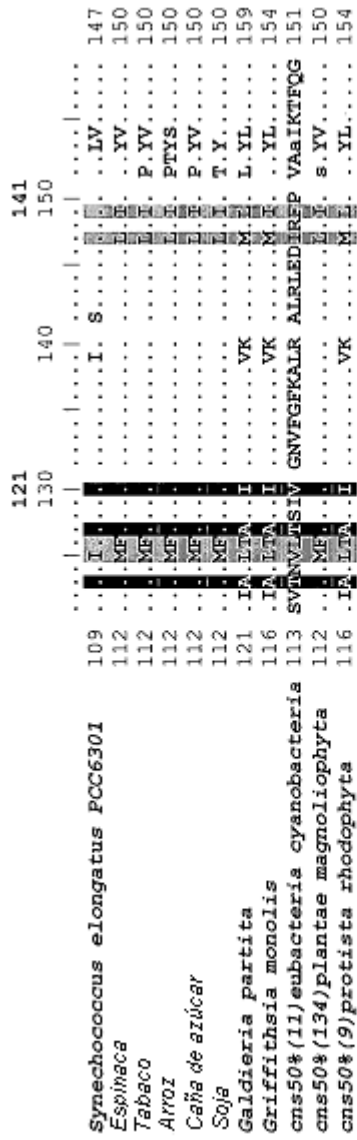


Figura 14B

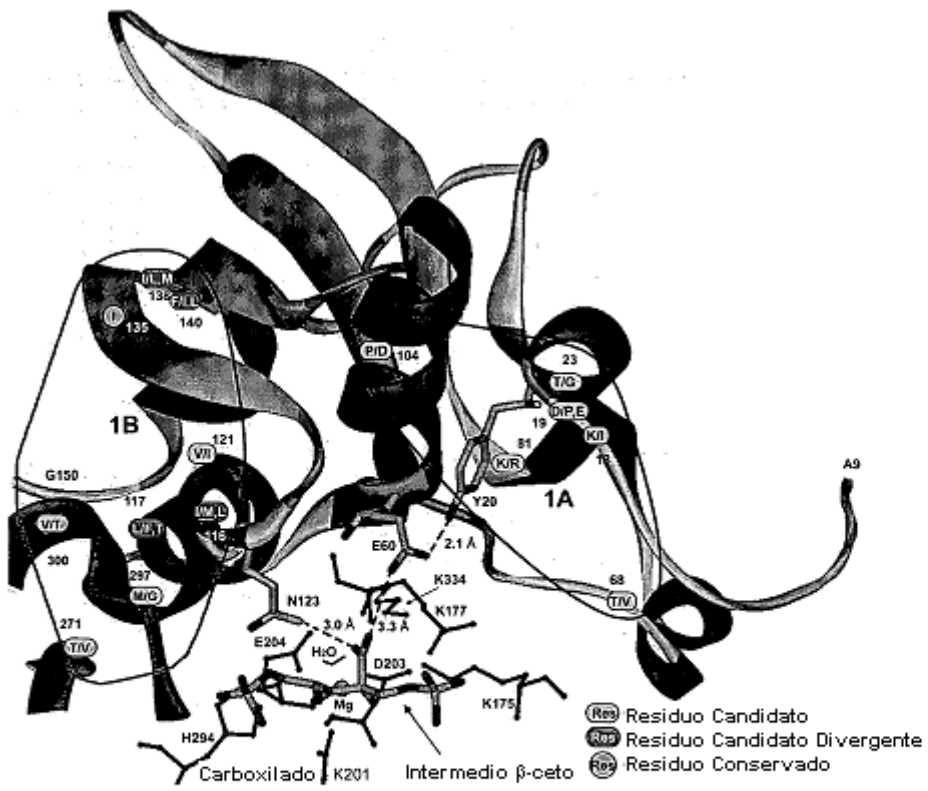


Figura 15

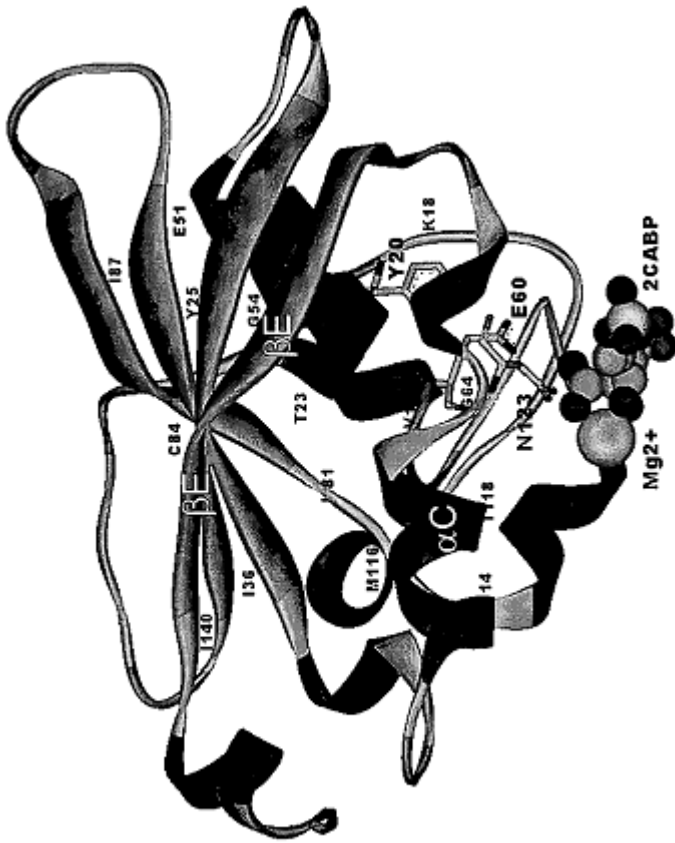


Figura 16 A

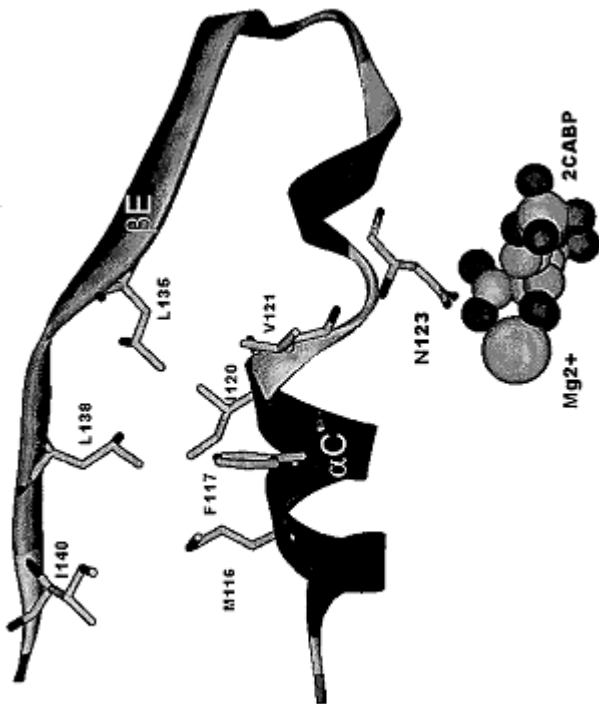


Figura 16 B

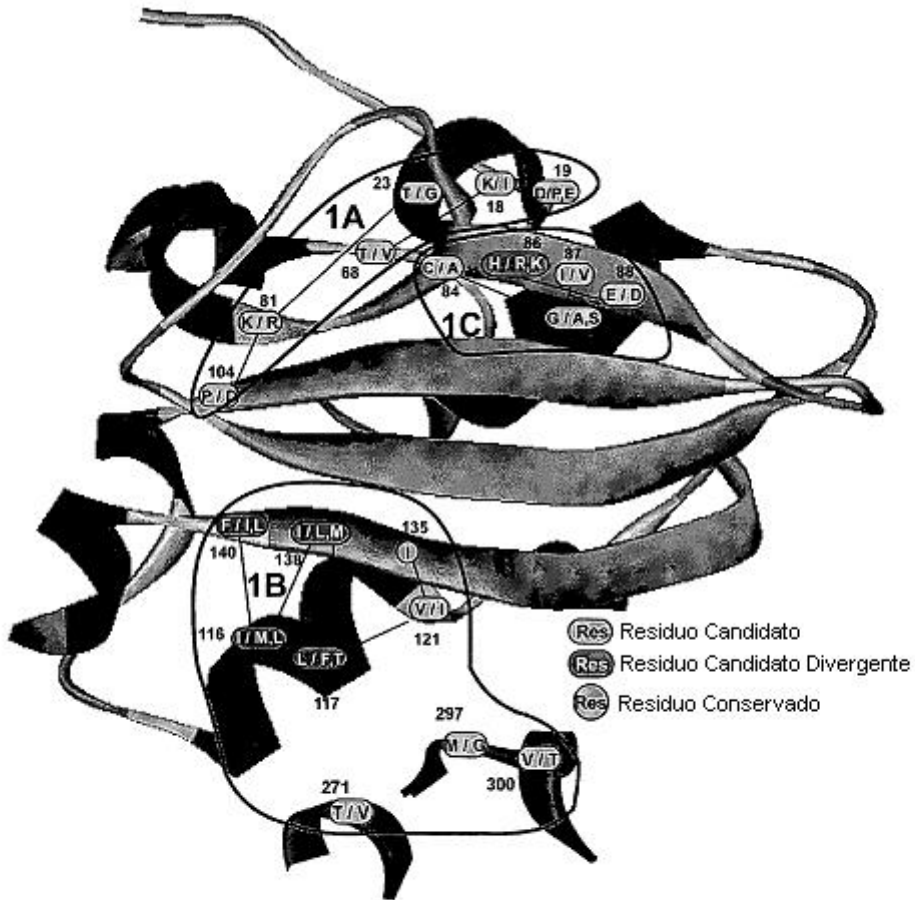
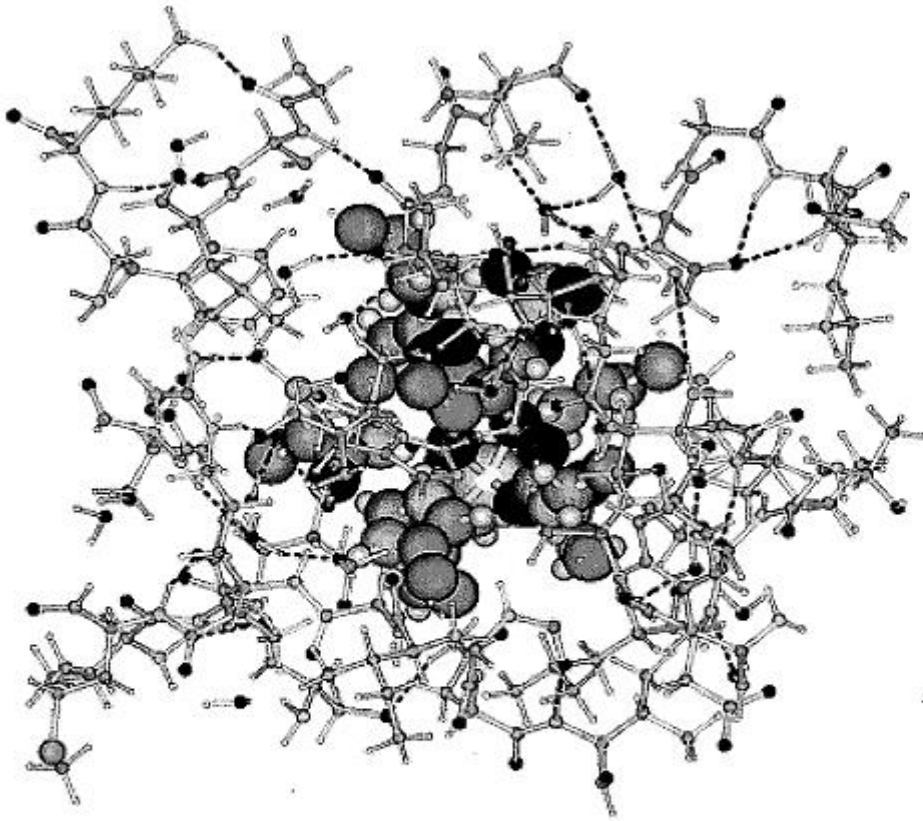


Figura 17



**Figura 18**

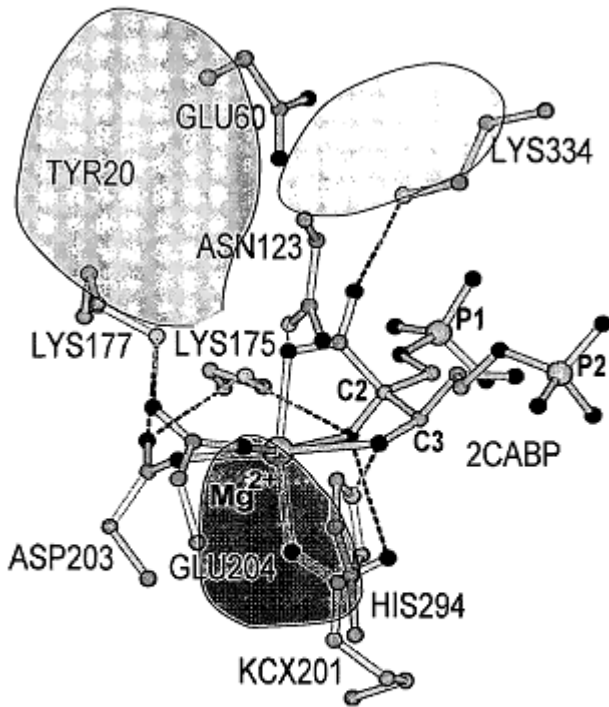


Figura 19

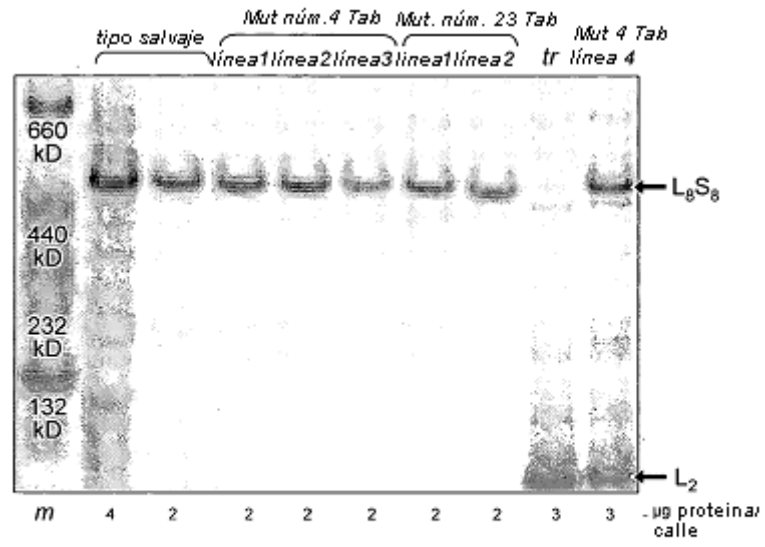


Figura 20