

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 585 054**

21 Número de solicitud: 201530877

51 Int. Cl.:

G06F 19/14 (2011.01)

C12N 15/10 (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

19.06.2015

43 Fecha de publicación de la solicitud:

03.10.2016

Fecha de concesión:

04.07.2017

45 Fecha de publicación de la concesión:

11.07.2017

73 Titular/es:

**UNIVERSIDAD DE GRANADA (100.0%)
HOSPITAL REAL. AVDA. HOSPICIO S/N
18071 GRANADA (Granada) ES**

72 Inventor/es:

**SÁNCHEZ RUIZ, José Manuel;
MARTÍNEZ RODRÍGUEZ, Sergio;
RISSO, Valeria Alejandra y
IBARRA MOLERO, Beatriz**

54 Título: **PROCEDIMIENTO PARA DISEÑAR POLIPÉPTIDOS PSEUDOANCESTRALES CON CARACTERÍSTICAS MEJORADAS**

57 Resumen:

Procedimiento para diseñar polipéptidos pseudoancestrales con características mejoradas. La presente invención tiene por objeto un procedimiento para la obtención de secuencias que codifican polipéptidos capaces de capturar propiedades extremas ancestrales tales como hiperestabilidad y promiscuidad de sustrato. Asimismo, la invención se refiere a un procedimiento de producción de tales polipéptidos y a los polipéptidos obtenidos.

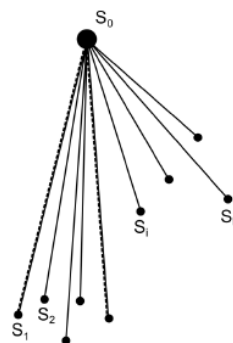


Figura 2

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP 11/1986.

ES 2 585 054 B1

DESCRIPCION**PROCEDIMIENTO PARA DISEÑAR POLIPÉPTIDOS PSEUDOANCESTRALES CON CARACTERÍSTICAS MEJORADAS****CAMPO DE LA INVENCION**

5

La presente invención se enmarca en el campo de la ingeniería de proteínas. En particular es útil para la obtención de polipéptidos modificados con características mejoradas (en términos de estabilidad y función) con vistas a sus aplicaciones biotecnológicas, directamente o como punto de partida de diseño racional o de evolución dirigida en laboratorio. Dado que la gran mayoría de las aplicaciones biomédicas e industriales de las proteínas requieren variantes con características mejoradas, la invención puede encontrar aplicación en sectores diversos (detergentes, industria textil, industria farmacéutica, industria de alimentación, etc.).

15

ESTADO DE LA TÉCNICADiseño de proteínas ancestrales

El gran número de secuencias de proteínas disponibles a partir de la era genómica ha provocado un gran interés en el uso de la diversidad natural de secuencias en ingeniería de proteínas [Cole MF, Cox VE, Gratton KL, Gaucher EA (2013) *Reconstructing evolutionary adaptive paths for protein engineering. Methods in Molecular Biology* 978:115-125.]. Uno de los enfoques más prometedores en este sentido es la reconstrucción de proteínas ancestrales, una metodología que, con carácter general, consta de las siguientes etapas [Fernandez JM, Gaucher E, Kosuri P, Perez-Jimenez R (2012) *Ancestral proteins* WO 2012009693 A1]:

- Etapa 1. Se obtiene un cierto número de secuencias correspondientes a una proteína en diferentes organismos actuales a partir de una búsqueda en una base de datos de secuencias.
- Etapa 2. Se lleva a cabo un análisis filogenético a partir de esas secuencias (análisis de las relaciones evolutivas de un grupo determinado de secuencias aminoacídicas) para obtener el correspondiente árbol filogenético y/o evolutivo.
- Etapa 3. Se utiliza un modelo evolutivo simple para obtener aproximaciones razonables a las secuencias en los nodos internos del árbol, es decir se reconstruyen las secuencias ancestrales.

- Etapa 4. Finalmente, se usan las técnicas habituales de biología molecular para preparar en el laboratorio las proteínas codificadas por las secuencias ancestrales reconstruidas, lo que se conoce habitualmente en el campo como “resucitar” las proteínas ancestrales.

5

Enzimas y sus aplicaciones

Las enzimas son polipéptidos que presentan numerosas aplicaciones biotecnológicas (industria textil, industria de alimentos, industria farmacéutica, detergentes, piensos para animales, procesamiento de papel, etc.) y biomédicas (anticuerpos monoclonales, hormonas, etc.).

10

La inmensa mayoría de estas aplicaciones usan enzimas modificadas ya que las enzimas naturales no suelen presentar las características adecuadas, en términos por ejemplo de estabilidad y propiedades catalíticas, para su aplicación práctica [Kirk O, Borchert TV, Fuglsang CC (2002) *Industrial enzyme applications*. Current Opinion in Biotechnology 13:345-351; Lazar GA, Marshall SA, Plecs JJ, Mayo SL, Dejarlais JR (2003) *Designing proteins for therapeutic applications*. Current Opinion in Structural Biology 13:513-518; Maynard J, Georgiu G (2000) Antibody Engineering Annual Review of Biomedical Engineering 2:339-376]. Las modificaciones introducidas son en muchos casos mutaciones y el campo en general se conoce como “Ingeniería de proteínas”.

20

Enzimas “ancestrales”

Varios estudios recientes en los que se han resucitado proteínas ancestrales correspondientes a nodos muy antiguos (nodos precámbricos del orden de miles de millones de años de antigüedad) han producido enzimas con muy alta estabilidad e incluso con promiscuidad catalítica (capacidad de degradar una variedad de sustratos, por ejemplo) [Gaucher EA, Govindarajan S, Ganesh OK (2008) *Palaeotemperature trend for Precambrian life inferred from resurrected proteins*. Nature 451:704-707; Perez-Jimenez R, Ingles-Prieto A, Zhao ZM, Sanchez-Romero I, Alegre-Cebollada J. Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A, Sanchez-Ruiz JM, Gaucher EA, Fernandez JM (2011) *Single-molecule paleoenzymology probes the chemistry of resurrected enzymes*; Risso VA, Gavira JA, Mejia-Carmona D, Gaucher EA, Sanchez-Ruiz JM (2013) *Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β -lactamases*. Journal of the American Chemical Society 135:2899-2902; Fernandez JM, Gaucher E, Kosuri P, Perez-Jimenez R (2012) *Ancestral proteins* WO 2012009693 A1]

35

Hay que resaltar que tanto la estabilidad como la promiscuidad contribuyen a la “evolvabilidad” (en inglés “*evolvability*”) de la enzima [Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) *Protein stability promotes evolvability*. Proceedings of the National Academy of Sciences of the U.S.A. 103:5869-5874; Risso VA, Gavira JA, Sanchez-Ruiz JM (2014) *Thermostable and promiscuous Precambrian proteins*. Environmental Microbiology 16:1485-1489], entendiendo por “evolvabilidad” la capacidad de desarrollar una función nueva mediante evolución dirigida en el laboratorio o mediante diseño racional [Nobeli I, Favia AD, Thornton JM (2009) *Protein promiscuity and its implications for biotechnology*. Nature Biotechnology 27:157-167]. La alta evolvabilidad es, de hecho, una característica altamente deseable en enzimas a utilizar como punto de partida en proyectos de ingeniería de proteínas. En este sentido, el potencial biotecnológico de la resurrección de proteínas ancestrales precámbricas es reconocido [ver, por ejemplo, los comentarios del Faculty 1000 al artículo de J AM Chem Soc. “*Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β -lactamases*”: <http://f1000.com/prime/717979625>. Último acceso 16/06/2015]. Sin embargo, tal y como se detalla debajo, el enfoque presenta un serio inconveniente metodológico que limita su aplicación práctica en un contexto biotecnológico.

Problemas metodológicos de la reconstrucción de secuencias ancestrales

El procedimiento de reconstrucción de secuencias ancestrales comentado anteriormente comprende una etapa que consiste en la derivación de un árbol filogenético creíble a partir de las secuencias modernas de partida. Esto, sin embargo, presenta varios problemas:

1. Las transferencias genéticas horizontales ocurren con mucha frecuencia y pueden distorsionar fuertemente el análisis filogenético [Koonin EV, Makarova KS, Aravind L (2001) *Horizontal gene transfer in prokaryotes: quantification and classification*. Annual Review of Microbiology 55:709-742.].
2. Igualmente, para muchas proteínas, aparecen procesos de duplicación genética durante la evolución [Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ (2012) *Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication*. PLoS Biology 10:e1001446.] y pueden resultar difíciles de modelar adecuadamente. Nótese que las duplicaciones genéticas hacen que el árbol derivado a partir de las secuencias de una

proteína concreta no tenga porqué parecerse a la filogenia de los organismos, con lo que el test más directo de la plausibilidad del árbol filogenético de la proteína no puede aplicarse.

3. Los programas que se consideran en el campo como más fiables para la obtención del árbol filogenético de una proteína, en particular los análisis bayesianos [Yang Z, Rannala B (1997) *Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method*. Molecular Biology and Evolution 14:717-724; Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) *Bayesian inference phylogeny and its impact on evolutionary biology*. Science 294:2310-2314; Kolaczkowski B, Thornton JW (2007) *Effects of branch length uncertainty on Bayesian posterior probabilities for phylogenetic hypotheses*. Molecular Biology and Evolution 249:2108-2118], son típicamente lentos y sólo son eficaces con un número limitado de secuencias (del orden de 100) y con un número limitado de residuos (del orden de unos cientos). Ese número de secuencias debe seleccionarse a partir de las depositadas en las bases de datos habituales de acuerdo con criterios que no se encuentran sistematizados en la literatura y que, en algunos casos, pueden ser arbitrarios o subjetivos hasta cierto punto. Por otra parte, el basar el análisis filogenético en un número limitado de secuencias implica que no se usa toda la información disponible en las bases de datos, lo que es claramente indeseable, sobre todo cuando las bases de datos de secuencias están creciendo exponencialmente.

Resulta claro de lo arriba expuesto que la construcción de un árbol filogenético para una proteína a partir de un conjunto de secuencias actuales puede llegar a ser un proceso extremadamente tedioso, incierto y poco flexible. Algunos programas comprendidos en el estado de la técnica que se basan en análisis bayesianos que emplean la metodología MCMC, (del inglés "*Markov Chain Monte Carlo*") o Métodos de Monte Carlo basados en Cadenas de Markov, pueden requerir tiempos de ejecución de incluso varios días en clusters de 8 procesadores de última generación (proporcionando un aumento sustancial de la capacidad de cálculo), incluso para números relativamente bajos de secuencias de longitud moderada. Hay que tener en cuenta, además, que, el árbol resultado de la ejecución puede no ser satisfactorio, debiendo examinarse para identificar inconsistencias con las relaciones evolutivas aceptadas [Cole MF, Cox VE, Gratton KL, Gaucher EA (2013) *Reconstructing evolutionary adaptive paths for protein engineering*. Methods in Molecular Biology 978:115-125] y decidir si tales inconsistencias son reales (debidas a eventos de

- transferencia horizontal o duplicación genética) o son debidas a errores (insuficiente representación de algunos grupos taxonómicos en las secuencias de partida). En este último caso, debe repetirse la ejecución [Cole MF, Cox VE, Gratton KL, Gaucher EA (2013) *Reconstructing evolutionary adaptive paths for protein engineering*. Methods in Molecular Biology 978:115-125] con parámetros modificados o incluso con un conjunto diferente de secuencias de partida para minimizar distorsiones debidas a transferencias horizontales, duplicaciones, etc. Todos estos ajustes, además, deben realizarse sin criterios claros que los dirijan. El resultado es que el análisis filogenético puede requerir muchas semanas, incluso para un número limitado de secuencias de una proteína de pequeño tamaño. El análisis para proteínas de tamaño medio o grande (la situación habitual en el contexto de una aplicación biotecnológica) y/o usando todas las secuencias depositadas en las bases de datos, puede simplemente no ser posible en un tiempo razonable.
- Los inventores de la presente invención desconocen otros procedimientos que permitan obtener variantes de polipéptidos con propiedades ancestrales extremas y que no incluyan la etapa de construcción del árbol filogenético salvo el llamado “*enfoque de consenso*” que se basa en la introducción en cada posición del aminoácido que aparece con más frecuencia en el alineamiento de las secuencias actuales. Sin embargo, se ha demostrado recientemente [Risso VA, Gavira JA, Gaucher EA, Sanchez-Ruiz JM (2014) *Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins*. Proteins 82:887-896] que el método de consenso captura las propiedades extremas ancestrales (particularmente la promiscuidad enzimática y la termoestabilidad) de forma muy limitada, limitando o disminuyendo el interés del uso de dichas variantes consenso ante una posible aplicación industrial al ser comparadas con las respectivas variantes ancestrales.

En resumen, resulta altamente deseable disponer de un procedimiento que permita acceder a las propiedades ancestrales de las proteínas (típicamente extremas y de claro interés biotecnológico) pero que evite todos los inconvenientes asociados a la derivación de un árbol filogenético de la proteína.

BREVE DESCRIPCIÓN DE LA INVENCION

35

La presente invención tiene por objeto un procedimiento para la obtención de secuencias que codifican polipéptidos pseudoancestrales, en adelante "procedimiento de la invención" que no utiliza un análisis filogenético, donde por "polipéptido pseudoancestral" (del inglés "*ancestral-like*") se entenderá a aquellos polipéptidos
 5 obtenidos mediante el procedimiento de la invención sin necesidad de realizar un análisis filogenético, y que con carácter general, capturan propiedades extremas ancestrales tales como hiperestabilidad y promiscuidad de sustrato.

Un segundo objeto de la invención es un procedimiento de síntesis/producción de polipéptidos que comprende la obtención de la secuencia polipeptídica mediante el
 10 primer procedimiento de la invención, en adelante "procedimiento de producción de polipéptidos de la invención" o "procedimiento de producción de la invención".

Un tercer objeto de la invención se refiere a los polipéptidos pseudoancestrales obtenidos con el procedimiento de producción de la invención, que no se encuentran descritos en el estado de la técnica y presentan propiedades ancestrales mejoradas que los hacen interesantes para su aplicación industrial.
 15

Un cuarto objeto de la invención es un programa de ordenador que comprende instrucciones para hacer que un ordenador lleve a cabo el procedimiento de la invención.
 20

Otro objeto de la invención es un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo el procedimiento de la invención.
 25

Otro objeto de la invención se refiere a una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo el procedimiento de la invención.
 30

DESCRIPCIÓN DE LAS FIGURAS

Figura 1.- Representación del árbol evolutivo / filogenético de la secuencia S_0' , correspondiente a la proteína ancestral de la que ha evolucionado el conjunto de proteínas con secuencias $S_1, S_2, \dots, S_N$. Los nodos intermedios corresponden a
 35

proteínas con secuencias $S'_1, S'_2, \dots, S'_T$.

Figura 2.- Representación del árbol en estrella de la secuencia pseudoancestral o central, S_0 , correspondiente al polipéptido pseudoancestral del que ha evolucionado el conjunto de polipéptidos con secuencias $S_1, S_2, \dots, S_N$.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Definiciones

En la presente descripción, dado un conjunto de secuencias de polipéptidos actuales, se entenderá por “secuencia de polipéptido ancestral” de ese conjunto a cualquier secuencia polipeptídica a partir de la que han evolucionado todos los elementos de dicho conjunto. Se entenderá por secuencia de polipéptido pseudoancestral a aquellas secuencias obtenidas mediante la metodología descrita en esta patente a partir de secuencias de polipéptidos actuales. Se entenderá por “secuencias de polipéptido actuales” a las secuencias polipeptídicas utilizadas para el cálculo de las secuencias pseudoancestrales. A modo de ejemplo, secuencias que pertenecen a organismos modernos que existen aún en la naturaleza, o habiendo sido extinguidos, sus secuencias se encuentran disponibles en la actualidad.

A lo largo de la presente invención se entenderá como “reconstrucción de secuencias de polipéptidos ancestrales o pseudoancestrales” a los diferentes cálculos matemáticos mediante los cuales se determinan dichas secuencias. Se entenderá como “resurrección de polipéptidos ancestrales o pseudoancestrales” a la producción o síntesis artificial en el laboratorio de las secuencias polipeptídicas reconstruidas mediante diferentes cálculos matemáticos, en términos similares a los descritos en WO 2012009693 A1.

En ese sentido, el término “polipéptido resucitado” se refiere a un polipéptido obtenido por dicho método.

Se entenderá por “árbol filogenético” (Figura 1) a un diagrama en forma de árbol que muestra las relaciones evolutivas entre varias especies u otras entidades que se cree que tienen una ascendencia común.

Se entenderá por “árbol en estrella” (Figura 2) a un diagrama en forma de árbol que tiene un único origen (denominada “secuencia central”) al que están conectados el resto de elementos del diagrama. A diferencia del árbol filogenético, a la hora de representar un polipéptido pseudoancestral del que han evolucionado un conjunto de polipéptidos, en esta representación se prescinde de todas las secuencias intermedias que formarían el árbol. Esto es, cada secuencia actual estará conectada directamente con un hipotético pseudoancestro común, evitándose así la necesidad de incluir los nodos intermedios en cada trayectoria, que complican el análisis.

Alineamientos de secuencias: Pasos previos al procedimiento de la invención

Se entenderá por “alineamiento de secuencias” a la forma de representar y comparar dos o más secuencias primarias de polipéptidos para resaltar sus zonas de similitud, que podrían indicar relaciones funcionales o evolutivas entre los genes o polipéptidos consultados. Las secuencias alineadas se escriben con las letras (representando aminoácidos o nucleótidos) en filas de una matriz en las que, si es necesario, se insertan espacios (en inglés, “*gaps*”) para que las zonas con idéntica o similar estructura se alineen.

Los alineamientos globales, que intentan alinear cada residuo de cada secuencia, son más útiles cuando las secuencias problema iniciales son similares y aproximadamente del mismo tamaño (no quiere decir que los alineamientos globales no puedan terminar en huecos). A modo de ejemplo, una estrategia general de alineamiento global es el algoritmo Needleman-Wunsch basado en programación dinámica.

Los alineamientos locales son más útiles para secuencias diferenciadas en las que se sospecha que existen regiones muy similares o motivos de secuencias similares dentro de un contexto mayor. A modo de ejemplo, el algoritmo Smith-Waterman es un método general de alineamiento local basado en programación dinámica. Con secuencias suficientemente similares, no existe diferencia entre alineamientos globales y locales.

Los métodos híbridos, conocidos como “*semiglobales*” o métodos “*glocales*” intentan encontrar el mejor alineamiento posible que incluya el inicio y el final de una u otra secuencia. Puede ser especialmente útil cuando la parte “*corriente arriba*” de una

secuencia se solapa con la parte "*corriente abajo*" de la otra. En este caso, ni el alineamiento global ni el local son completamente adecuados: un alineamiento global intentará forzar a la alineación a extenderse más allá de la región de solapamiento, mientras que el alineamiento local no cubrirá totalmente la región solapada.

5

El procedimiento de la invención parte de un alineamiento de las secuencias de los polipéptidos para los que se quiere calcular un polipéptido pseudoancestral.

De forma preferente, el procedimiento de la invención utilizará como punto de partida un alineamiento global de secuencias.

10

Pueden utilizarse como alineamientos de partida aquellos libremente disponibles en bases de datos públicas, tales como los descritos en [Lombard V, Camon EB, Parkinson HE, Hingamp P, Stoesser G, Redaschi N. (2002) EMBL-Align: a new public nucleotide and amino acid multiple sequence alignment database. Bioinformatics.18(5):763-764].

15

Alternativamente, se realiza un alineamiento, de forma preferente global, de las N secuencias de polipéptidos actuales: $S_1, S_2, \dots, S_N$, utilizando herramientas bioinformáticas libremente disponibles, tales como Clustal W [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 Research 22(22):4673-4680].

20

Todas las secuencias tras el alineamiento tienen la misma longitud, L , que se corresponden con el número de aminoácidos más los huecos o *gaps* generados en el alineamiento.

25

El aminoácido que se encuentran en la posición j de una secuencia se denominará A_j , por lo que una secuencia alineada se puede representar representará como $a_1, a_2, \dots, a_L$.

30

Así, si llamamos " A " a la matriz que contiene el alineamiento, el aminoácido presente en la posición j de la secuencia S_i se identificará como A_{ij} .

35

En particular, el aminoácido presente en la posición j de la secuencia pseudoancestral, S_0 , se identifica como A_{0j} .

- 5 Los procedimientos de cálculo de proteínas ancestrales conocidos necesitan una compleja edición del alineamiento inicial. El procedimiento de la invención evita tener que realizar modificaciones en el alineamiento.

10 **Procedimiento para la obtención de secuencias que codifican polipéptidos pseudoancestrales**

- 15 El primer objeto de la invención es la obtención de polipéptidos pseudoancestrales que presenten las propiedades extremas y de interés biotecnológico de los polipéptidos ancestrales, como por ejemplo la promiscuidad de sustrato y la termoestabilidad, pero evitando la etapa de derivación del árbol filogenético que incluyen los métodos de reconstrucción de secuencias ancestrales ya que dicha etapa constituye una limitación muy seria a la aplicación biotecnológica de la reconstrucción/resurrección de polipéptidos ancestrales.

- 20 Así, el procedimiento de la invención calcula la secuencia del polipéptido pseudoancestral, S_0 , o "secuencia central", a partir de una estimación máximo verosímil que parte de la siguiente hipótesis:

- 25 Se considera que la secuencia central, desconocida en principio, está unida directamente a las secuencias de partida actuales mediante ramas de longitudes $t(S_1)$, $t(S_2)$, ..., $t(S_N)$, siendo $t(S_i)$, la "longitud de la rama" que une S_0 y S_i ", o la "distancia entre S_0 y S_i ", el número de sustituciones de aminoácidos existentes entre S_0 y S_i .

- 30 A modo de ejemplo, la distancia entre las siguientes secuencias S_0 y S_1 es $t(S_1) = 4$:

```

S0=   FAGGTDVVLL
S1=   F-GGAD-VVLT
      . . . . .
      * * * * *
```

$$t(S_1) = 4$$

Gráficamente esta hipótesis se puede representar mediante un árbol en estrella (Figura 2) en el que todos los polipéptidos de partida comparten un pseudoancestro común del que parten todas las ramas del árbol. Evidentemente, las longitudes de cada rama son desconocidas en principio.

5

En este contexto, el procedimiento de la invención es un procedimiento que permite obtener la secuencia de un polipéptido pseudoancestral, S_0 , a partir del alineamiento, $A = [A_{ij}]$, $i = 1, \dots, N$; $j = 1, \dots, L$; de N polipéptidos (polipéptidos actuales) con secuencias alineadas $S_1, S_2, \dots, S_N$, que comprende la maximización de la función de verosimilitud $V'(S_0, \{t_i\})$

10

$$V'(S_0, \{t_i\}) = \prod_{i=1, \dots, N} \prod_{j=1, \dots, L} \left[P(A_{ij} | A_{0j}, t(S_i)) \right]^{w(S_i)} \quad (\text{Ec. 1})$$

donde $\{t_i\} = \{t(S_1), t(S_2), \dots, t(S_N)\}$, es el conjunto de las longitudes de rama que conectan S_0 con cada una de las secuencias alineadas de partida; $P(A_{ij} | A_{0j}, t(S_i))$ es la probabilidad de observar el aminoácido presente en la posición j de la secuencia S_i , fijados el aminoácido presente en esa posición en la secuencia central, S_0 , y la longitud de la rama que conecta S_0 con S_i , $t(S_i)$; y $w(S_1), w(S_2), \dots, w(S_N)$ son un conjunto de pesos estadísticos para las secuencias actuales que pueden usarse para corregir el hecho de que grupos de secuencias similares puede distorsionar el análisis.

15

20

El procedimiento no sólo calcula la secuencia pseudoancestral hipotética para unas secuencias de partida y longitudes de rama dadas, sino que explora el espacio de las longitudes de rama y encuentra la longitud de rama óptima (o sea, la que maximiza la verosimilitud) para cada secuencia actual dada.

25

En una realización preferente, las probabilidades se calculan a partir de matrices de sustitución de aminoácidos y frecuencias medias de éstos usando una cadena de Markov de tiempo continuo, tal y como se describe en la literatura [Whelan, S., and N. Goldman. 2001. *A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach*. Mol. Biol. Evol. 18:691–699; Huelsenbeck, Ronquist, *Bayesian analysis of molecular evolution using MrBayes*. In: Nielsen, R. (ed.), Statistical Methods in Molecular Evolution. Springer, New York,

30

2005]. Para ello se utiliza un modelo de Markov de tiempo continuo [Felsenstein J (1981) *Evolutionary trees from DNA sequences: a maximum likelihood approach*. Journal of Molecular Evolution 17:368-376] con una matriz de sustitución aminoacídica que permite obtener la probabilidad del aminoácido moderno en cada posición de cada
 5 secuencia actual dado el aminoácido ancestral y la longitud de la rama que conecta la secuencia actual y la ancestral hipotética.

Empleo de pesos

10 En la práctica, este planteamiento basado en un árbol de estrella podría sobrestimar la contribución de secuencias actuales similares que divergieron recientemente y cuyas ramas desde el ancestro no pueden considerarse independientes (como las trayectorias marcadas con líneas discontinuas en las Figuras 1 y 2). Para compensar esta distorsión se utiliza una función de verosimilitud modificada en que se introduce
 15 un peso estadístico para la probabilidad de cada secuencia actual que depende del número de secuencias parecidas a la secuencia que se considera.

Los pesos $w(S_1), w(S_2), \dots, w(S_N)$ se calculan de manera que el peso de una secuencia que forma parte de uno de esos grupos contribuya de forma distinta a la secuencia
 20 final. En una realización particular, los pesos de todas las secuencias son iguales a 1.

La secuencia central óptima, $\hat{S}_0$, es aquella que maximiza la verosimilitud definida en la ecuación 1.

Programa de ordenador, señal y dispositivo que implementa el procedimiento de la invención

Un cuarto objeto de la invención es un programa de ordenador que comprende instrucciones para hacer que un ordenador lleve a cabo el procedimiento de la
 30 invención.

La invención abarca programas de ordenador que pueden estar en forma de código fuente, de código objeto o en un código intermedio entre código fuente y código objeto, tal como en forma parcialmente compilada, o en cualquier otra forma adecuada para
 35 usar en la implementación de los procesos de acuerdo con la invención. En particular,

los programas de ordenador también abarcan aplicaciones en la nube que implementen el procedimiento de la invención.

Estos programas pueden estar dispuestos sobre o dentro de un soporte apto para su
 5 lectura, en adelante, "medio portador" o "portador". El medio portador puede ser cualquier entidad o dispositivo capaz de portar el programa. Cuando el programa va incorporado en una señal que puede ser transportada directamente por un cable u otro dispositivo o medio, el medio portador puede estar constituido por dicho cable u otro dispositivo o medio. Como variante, el medio portador podría ser un circuito integrado
 10 en el que va incluido el programa, estando el circuito integrado adaptado para ejecutar, o para ser utilizado en la ejecución de, los procesos correspondientes.

A modo de ejemplo, los programas podrían estar incorporados en un medio de almacenamiento, como una memoria ROM, una memoria CD ROM o una memoria
 15 ROM de semiconductor, una memoria USB, o un soporte de grabación magnética, por ejemplo, un disco flexible o un disco duro. Alternativamente, los programas podrían estar soportados en una señal portadora transmisible. Por ejemplo, podría tratarse de una señal eléctrica u óptica que podría transportarse a través de cable eléctrico u óptico, por radio o por cualesquiera otros medios.

20 En este sentido, otro objeto de la invención es un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo el procedimiento de la invención.

25 Finalmente, un último objeto de la invención se refiere a una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo el procedimiento de la invención.

Procedimiento de producción de polipéptidos

30 Otro objeto de la invención es un procedimiento, en adelante "procedimiento de producción de la invención", de síntesis o producción de un polipéptido, polipéptido pseudoancestral, del que han ha evolucionado un conjunto de N polipéptidos determinados, que comprende la obtención previa de la secuencia de dicho polipéptido
 35 pseudoancestral mediante el procedimiento de la invención.

En una realización particular, dado un conjunto de N polipéptidos actuales, el procedimiento para sintetizar un polipéptido pseudoancestral de los que habrían evolucionado las anteriores, comprende los siguientes pasos:

- a) Obtener la secuencia óptima del polipéptido pseudoancestral, S_0 , obtenida a partir del alineamiento de secuencias de los N polipéptidos.
- b) Sintetizar el gen que codifica la secuencia óptima del polipéptido obtenido en el paso anterior
- c) Obtener el polipéptido pseudoancestral mediante sobreexpresión heteróloga.

10 **Polipéptidos obtenidos por el procedimiento de síntesis de la invención**

Otro objeto de la invención se refiere a los polipéptidos pseudoancestrales obtenidos por el procedimiento de síntesis de la invención, en adelante “*polipéptidos de la invención*”. En particular, son objeto de la invención las β -lactamasas pseudoancestrales Seclet000, Seclet001, Seclet006 y Seclet012 (con secuencias SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 y SEQ ID NO: 4, respectivamente), no descritas en el estado de la técnica, y que presentan una alta promiscuidad de sustrato e hiperestabilidad:

SEQ ID NO: 1 (Seclet000):

20 MQAEALEALAELEERQSGGRLGVAVLDTASGRRIAYRADERFPMCSTFKVMLAAAVLARVDQGQE
QLDRRITYYSKSDLVNSPVTEKHVGEGMTVAELCAAIIQYSDNTAANLLKQIGGPAAVTAFLR
SIGDNTTRLDRWEPELNTAIPGDPRDTTTPAAMAASLRLLLLGDALSPASRAQLVAVMMGNKTG
DNRIRAGLPAGWRVGDKTGTGDYGTNDIAVIWPPNRAPIVLAVYFTQSQADAKARDDVIAAVA
RIVVAAFLEHHHHHH

SEQ ID NO: 2 (Seclet001):

25 MAAEALEALAELEERQSGGRLGVAVLDTASGRRIAYRADERFPMCSTFKVMLAAAVLARVDQGKE
QLDRRITYYSKSDLVNSPVTEKHVGNGMTVAELCAAIIQYSDNTAANLLKQIGGPAAVTAFLR
SIGDDTTRLDRWEPELNTAAPGDPRDTTTPAAMAATLRLLLLGDALSPASRAQLVEWMMGNKTG
30 DNRIRAGLPAGWRVGDKTGTGDYGTNDIAVIWPPNRAPIVLAVYFTQSQADAKARDDVIAAVA
RIVVAAFLEHHHHHH

SEQ ID NO: 3 (Seclet006):

35 MAAEALEQLAELEKRS GGRLGVAVLDTATGRRIGYRADERFPMCSTFKVLLAAAVLARVDQGKE
QLDRRISYSKSDLVNSPVTEKHVGNGMTVAELCAAITMSDNTAANLLKQIGGPAAVTAFLR
SIGDDTTRLDRREPELNNAAPGDPRDTTTPAAMAATLRLLLLGDALSPASRAQLMEWMMGNKTG

DNRIRAGLPAGWRVGDKTGSGGYGTTNDIAVIWPPNRAPIIIIAVYFTQSQADAEARDDVIAAVA
RIVVAAFLEHHHHHH

SEQ ID NO: 4 (Seclet012):

5 MAAEALEQLAELEKQSGGRLGVAVLDTATGRRIGYRGDERFPMCSTFKVLLAAVLAARVDQGKE
QLDRRISYSKSDLVSYSPVTEKHVGKGMTVAELCAAITMSDNTAANLLKQIGGPAAVTAFLR
SIGDNTTRLDRREPELNNAAPGDPRDTTTPAAMAATLRRLLLGDALSPASRAQLMEWMMANKTG
DNRLRAGLPAGWRIGDKTGTGGHGTTNDIAVIWPPNRAPIIIIAVYFTQSQADAAERDEVIAEVA
RAVVEAFLEHHHHHH

10

También son objeto de la presente invención polipéptidos con una identidad de al menos un 80%, un 85%, un 90%, un 95%, un 98% o un 99% con cualquiera de las secuencias 1 a 4 y en las que dichos polipéptidos poseen similar actividad y características estructurales que las β -lactamasas, en particular, polipéptidos con similar promiscuidad de sustrato e hiperestabilidad.

15

Estas β -lactamasas pseudoancestrales presentan aplicación en la descontaminación de antibióticos presentes en diferentes medios líquidos relacionados con el consumo humano (por ejemplo, en la leche de vaca [Lee MZ, Richardson T. (1987) *Preparation and characterization of immobilized beta-lactamase for destruction of penicillin in milk.* J Dairy Sci. 70(10):2032-2039]).

20

Un cuarto objeto de la invención se refiere a polinucleótidos (de aquí en adelante "polinucleótidos de la invención") que codifican los polipéptidos de la invención. En una realización particular, son objeto de la invención los polinucleótidos que codifican los polipéptidos con una identidad de al menos un 80%, un 85%, un 90%, un 95%, un 98% o un 99% con cualquiera de las secuencias 1 a 4 y en las que el polipéptido codificado por dichos polinucleótidos posee similar actividad y características estructurales que las β -lactamasas, en particular, similar promiscuidad de sustrato e hiperestabilidad.

25

30

En una realización más particular, dicho polipéptido tiene actividad enzimática.

Un quinto objeto de la invención se refiere a anticuerpos o fragmentos de los mismos que pueden unirse a polipéptidos con una cualquiera de las secuencias peptídicas mencionadas anteriormente.

35

La invención también se refiere a un vector recombinante, tal como un vector de expresión, que comprende un polinucleótido de la invención operativamente unido a una secuencia reguladora, por ejemplo un promotor; una célula huésped que se transforma con un polinucleótido de la invención; y un procedimiento de producción de un polipéptido con actividad enzimática, que comprende mantener una célula huésped transformada con un polinucleótido de la invención en condiciones para proporcionar la expresión del péptido.

MODO DE REALIZACIÓN DE LA INVENCION

Implementación del procedimiento

Una realización particular de la invención se ha llevado a cabo implementando el procedimiento de la invención en un programa de ordenador compilado en QB64 (Quick Basic 64, un compilador de lenguaje BASIC que es un emisor C++). El programa no sólo calcula la secuencia ancestral hipotética para unas secuencias de partida y longitudes de rama dadas, sino que explora el espacio de las longitudes de rama y encuentra la longitud de rama óptima (o sea, la que maximiza la verosimilitud) para cada secuencia actual dada.

Procedimiento para calcular las secuencias de β -lactamasas pseudoancestrales

Para validar el programa, así como la noción en la que se basa, se ha aplicado el procedimiento a un conjunto de 302 secuencias correspondientes a β -lactamasas de diversas proteobacterias ($N = 302$). Estas secuencias fueron localizadas utilizando el algoritmo BLAST, usando la secuencia de la β -lactamasa de *Escherichia coli* como molde (Número de Acceso en Genbank 1BTL). Se genera un alineamiento de las 302 secuencias utilizando Clustal W, y tras el curado del alineamiento se obtiene una longitud $L=262$.

Utilizando el procedimiento de la invención, se han calculado varias secuencias pseudoancestrales usando diferentes pesos estadísticos (valores comprendidos entre 0 y 1). Entre las variantes calculadas destacan las variantes pseudoancestrales Seclet000 (SEQ ID NO: 1), Seclet001 (SEQ ID NO: 2), Seclet006 (SEQ ID NO: 3), y Seclet012 (SEQ ID NO: 4).

Posteriormente, se han preparado en el laboratorio las correspondientes proteínas, adquiriendo en compañías privadas los plásmidos de expresión codificando las correspondientes secuencias génicas optimizadas para su expresión en *Escherichia coli* (TopGene Technologies). Las enzimas fueron producidas mediante un sistema de expresión heterólogo y purificadas mediante cromatografía de afinidad en un solo paso gracias a la inclusión de una cola de polihistidinas en el extremo carboxilo terminal.

Finalmente, se han comparado sus propiedades con la de la resurrección de laboratorio de la lactamasa correspondiente al ancestro común de las proteobacterias (GNCA, con secuencia SEQ ID NO: 5) que se obtuvo previamente mediante el análisis habitual que implica la construcción de un árbol filogenético:

SEQ ID NO: 5 (GNCA):

MAAQLSEQLAELEKRS GGRLGVAVLDTATGRRIAYRGDERFPMCSTFKALLAAV LARVDQGKE
RLDRRITYGKEDLDVYSPVTEKHVGDGMTVAELCEAAITLSDNTAANLLLEALGGPAALTAFRLR
SIGDEVTRLDRWEPELNEAAPGDPRDTTTPAAMAATLRTL LLDALSPASRQQQLVDWL VANKTG
DKRLRAGLPADWRVGDKTGTGGHGTTNDIAVIWPPGRAPIVVTVYLTESQVDADARDAVIAEVG
RLVVEAF

Esta proteína ancestral resucitada tras construir su árbol filogenético (Número de Acceso en Genbank 4B88) presenta una estabilidad enormemente incrementada y capacidad de catalizar eficazmente la degradación de varios antibióticos (Risso et al., JACS 2013). Es decir, presentó hiperestabilidad y promiscuidad de sustrato, cualidades altamente deseables desde un punto de vista biotecnológico.

Tal y como se recoge en la Tabla 1, las proteínas preparadas con la metodología propuesta aquí presentan unas propiedades similares, tanto en términos de estabilidad como de catálisis, a la de la proteína ancestral resucitada.

Sus secuencias se han obtenido sin embargo, utilizando un procedimiento que no implica la construcción de un árbol filogenético y que no requiere selección arbitraria de un subconjunto de secuencias.

Enzima	Eficiencia catalítica Bencilpenicilina ($\text{s}^{-1} \mu\text{M}^{-1}$)	Eficiencia catalítica Cefotaxina ($\text{s}^{-1} \mu\text{M}^{-1}$)	T _m (°C)

Tem-1	26±4.7	2.6x10 ⁻³ ±1x10 ⁻³	51.8
GNCA	0.3±0.1	1.2±0.3	90.3
Seclet000	1.2±0.4	0.6±0.1	84.4
Seclet001	5.4±1.9	1.0±0.0	85.0
Seclet006	2.0±0.9	1.6±0.1	83.9
Seclet012	0.8±0.3	1.1±0.3	75.5

Tabla 1.- Eficiencia catalítica de Las diferentes proteínas analizadas. T_m indica la temperatura media de desnaturalización.

REIVINDICACIONES

1. Procedimiento para obtener la secuencia, S_0 , de un polipéptido pseudoancestral a partir de un alineamiento $A = [A_{ij}]$, $i = 1, \dots, N$; $j = 1, \dots, L$ de N polipéptidos conocidos, $S_1, S_2, \dots, S_N$, que comprende calcular el vector, S_0 , que maximiza la función de verosimilitud:

$$V'(S_0, \{t_i\}) = \prod_{i=1, \dots, N} \prod_{j=1, \dots, L} \left[P(A_{ij} | A_{0j}, t(S_i)) \right]^{w(S_i)} \quad (\text{Ec. 1})$$

Donde:

- $\{t_i\} = \{t(S_1), t(S_2), \dots, t(S_N)\}$, siendo $t(S_i)$ el número de sustituciones de aminoácidos existentes entre S_0 y S_i .
- $P(A_{ij} | A_{0j}, t(S_i))$ es la probabilidad de observar el aminoácido presente en la posición j de la secuencia S_i , fijados el aminoácido presente en esa posición en la secuencia central, S_0 ;
- y $w(S_1), w(S_2), \dots, w(S_N)$ son un conjunto de pesos estadísticos.

2. Procedimiento según reivindicación anterior, caracterizado porque las probabilidades, $P(A_{ij} | A_{0j}, t(S_i))$, se calculan a partir de matrices de sustitución de aminoácidos y frecuencias medias de éstos usando una cadena de Markov de tiempo continuo.

3. Procedimiento según cualquiera de las reivindicaciones anteriores en el que los pesos son iguales a 1.

4. Programa de ordenador que comprende instrucciones de programa para hacer que un ordenador lleve a la práctica el procedimiento según cualquiera de las reivindicaciones 1 a 3.

5. Programa de ordenador de acuerdo con la reivindicación anterior, que está incorporado en unos medios de almacenamiento.

6. Señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador según reivindicación 4.

7. Procedimiento de síntesis de un polipéptido a partir del cual ha evolucionado un conjunto de N polipéptidos determinados que comprende la obtención previa de la secuencia de dicho polipéptido mediante el procedimiento de la invención.

8. Procedimiento de producción de un polipéptido a partir del cual ha evolucionado un conjunto de N polipéptidos determinados que comprende los siguientes pasos:

a) Obtener la secuencia óptima del polipéptido pseudoancestral, S_0 , obtenida a partir del alineamiento de secuencias de los N polipéptidos utilizando el procedimiento según reivindicaciones 1 a 3.

b) Sintetizar el gen que codifica la secuencia óptima del polipéptido obtenido en el paso anterior

c) Obtener el polipéptido pseudoancestral mediante sobreexpresión heteróloga.

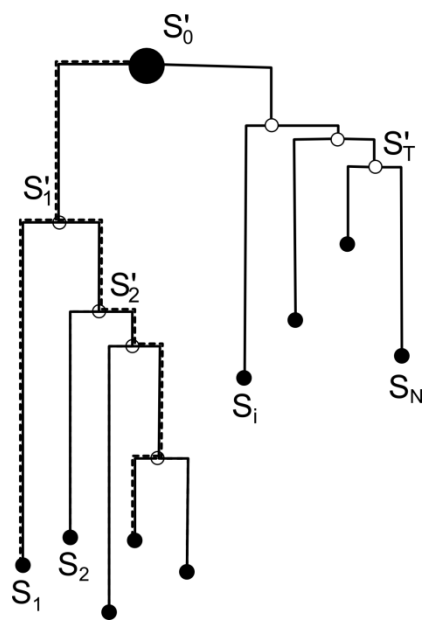


Figura 1

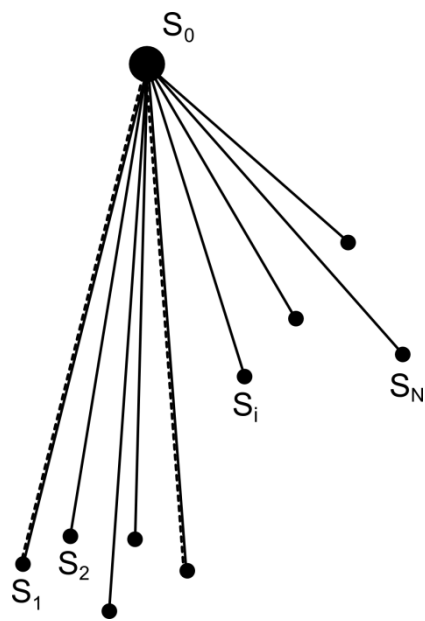


Figura 2

Listado de Secuencias

<110> Universidad de Granada
 <120> PROCEDIMIENTO PARA DISEÑAR POLIPÉPTIDOS PSEUDOANCESTRALES CON CARACTERÍSTICAS MEJORADAS
 <130> IPR-557
 <160> 5
 <170> PatentIn version 3.5
 <210> 1
 <211> 271
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> seclet000
 <400> 1

Met Gln Ala Glu Ala Leu Glu Ala Leu Ala Glu Leu Glu Arg Gln Ser
 1 5 10 15

Gly Gly Arg Leu Gly Val Ala Val Leu Asp Thr Ala Ser Gly Arg Arg
 20 25 30

Ile Ala Tyr Arg Ala Asp Glu Arg Phe Pro Met Cys Ser Thr Phe Lys
 35 40 45

Val Met Leu Ala Ala Ala Val Leu Ala Arg Val Asp Gln Gly Gln Glu
 50 55 60

Gln Leu Asp Arg Arg Ile Thr Tyr Ser Lys Ser Asp Leu Val Asn Tyr
 65 70 75 80

Ser Pro Val Thr Glu Lys His Val Gly Glu Gly Met Thr Val Ala Glu
 85 90 95

Leu Cys Ala Ala Ala Ile Gln Tyr Ser Asp Asn Thr Ala Ala Asn Leu
 100 105 110

Leu Leu Lys Gln Ile Gly Gly Pro Ala Ala Val Thr Ala Phe Leu Arg
 115 120 125

Ser Ile Gly Asp Asn Thr Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu
 130 135 140

Asn Thr Ala Ile Pro Gly Asp Pro Arg Asp Thr Thr Thr Pro Ala Ala
 145 150 155 160

Met Ala Ala Ser Leu Arg Arg Leu Leu Leu Gly Asp Ala Leu Ser Pro
 165 170 175

Ala Ser Arg Ala Gln Leu Val Ala Trp Met Met Gly Asn Lys Thr Gly
 180 185 190

ES 2 585 054 B1

Asp Asn Arg Ile Arg Ala Gly Leu Pro Ala Gly Trp Arg Val Gly Asp
195 200 205

Lys Thr Gly Thr Gly Asp Tyr Gly Thr Thr Asn Asp Ile Ala Val Ile
210 215 220

Trp Pro Pro Asn Arg Ala Pro Ile Val Leu Ala Val Tyr Phe Thr Gln
225 230 235 240

Ser Gln Ala Asp Ala Lys Ala Arg Asp Asp Val Ile Ala Ala Val Ala
245 250 255

Arg Ile Val Val Ala Ala Phe Leu Glu His His His His His His
260 265 270

<210> 2
<211> 271
<212> PRT
<213> Artificial Sequence

<220>
<223> Seclet001

<400> 2

Met Ala Ala Glu Ala Leu Glu Ala Leu Ala Glu Leu Glu Arg Gln Ser
1 5 10 15

Gly Gly Arg Leu Gly Val Ala Val Leu Asp Thr Ala Ser Gly Arg Arg
20 25 30

Ile Gly Tyr Arg Ala Asp Glu Arg Phe Pro Met Cys Ser Thr Phe Lys
35 40 45

Val Met Leu Ala Ala Ala Val Leu Ala Arg Val Asp Gln Gly Lys Glu
50 55 60

Gln Leu Asp Arg Arg Ile Thr Tyr Ser Lys Ser Asp Leu Val Ser Tyr
65 70 75 80

Ser Pro Val Thr Glu Lys His Val Gly Asn Gly Met Thr Val Ala Glu
85 90 95

Leu Cys Ala Ala Ala Ile Gln Tyr Ser Asp Asn Thr Ala Ala Asn Leu
100 105 110

Leu Leu Lys Gln Ile Gly Gly Pro Ala Ala Val Thr Ala Phe Leu Arg
115 120 125

Ser Ile Gly Asp Asp Thr Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu
130 135 140

Asn Thr Ala Ala Pro Gly Asp Pro Arg Asp Thr Thr Thr Pro Ala Ala

ES 2 585 054 B1

145	150										155				160	
Met	Ala	Ala	Thr	Leu 165	Arg	Arg	Leu	Leu	Leu 170	Gly	Asp	Ala	Leu	Ser 175	Pro	
Ala	Ser	Arg	Ala 180	Gln	Leu	Val	Glu	Trp 185	Met	Met	Gly	Asn	Lys 190	Thr	Gly	
Asp	Asn	Arg 195	Ile	Arg	Ala	Gly	Leu 200	Pro	Ala	Gly	Trp	Arg 205	Val	Gly	Asp	
Lys	Thr 210	Gly	Thr	Gly	Asp	Tyr 215	Gly	Thr	Thr	Asn	Asp 220	Ile	Ala	Val	Ile	
Trp 225	Pro	Pro	Asn	Arg	Ala 230	Pro	Ile	Val	Leu	Ala 235	Val	Tyr	Phe	Thr	Gln 240	
Ser	Gln	Ala	Asp	Ala 245	Lys	Ala	Arg	Asp	Asp 250	Val	Ile	Ala	Ala	Val 255	Ala	
Arg	Ile	Val	Val 260	Ala	Ala	Phe	Leu	Glu 265	His	His	His	His	His 270	His		
<210>	3															
<211>	271															
<212>	PRT															
<213>	Artificial Sequence															
<220>																
<223>	Seclet006															
<400>	3															
Met 1	Ala	Ala	Glu	Ala 5	Leu	Glu	Gln	Leu	Ala 10	Glu	Leu	Glu	Lys	Arg 15	Ser	
Gly	Gly	Arg	Leu 20	Gly	Val	Ala	Val	Leu 25	Asp	Thr	Ala	Thr	Gly 30	Arg	Arg	
Ile	Gly	Tyr 35	Arg	Ala	Asp	Glu	Arg 40	Phe	Pro	Met	Cys	Ser 45	Thr	Phe	Lys	
Val	Leu 50	Leu	Ala	Ala	Ala	Val 55	Leu	Ala	Arg	Val	Asp 60	Gln	Gly	Lys	Glu	
Gln 65	Leu	Asp	Arg	Arg	Ile 70	Ser	Tyr	Ser	Lys	Ser 75	Asp	Leu	Val	Ser	Tyr 80	
Ser	Pro	Val	Thr	Glu 85	Lys	His	Val	Gly	Asn 90	Gly	Met	Thr	Val	Ala 95	Glu	
Leu	Cys	Ala	Ala 100	Ala	Ile	Thr	Met	Ser 105	Asp	Asn	Thr	Ala	Ala 110	Asn	Leu	

ES 2 585 054 B1

Leu Leu Lys Gln Ile Gly Gly Pro Ala Ala Val Thr Ala Phe Leu Arg
115 120 125

Ser Ile Gly Asp Asp Thr Thr Arg Leu Asp Arg Arg Glu Pro Glu Leu
130 135 140

Asn Asn Ala Ala Pro Gly Asp Pro Arg Asp Thr Thr Thr Pro Ala Ala
145 150 155 160

Met Ala Ala Thr Leu Arg Arg Leu Leu Leu Gly Asp Ala Leu Ser Pro
165 170 175

Ala Ser Arg Ala Gln Leu Met Glu Trp Met Met Gly Asn Lys Thr Gly
180 185 190

Asp Asn Arg Ile Arg Ala Gly Leu Pro Ala Gly Trp Arg Val Gly Asp
195 200 205

Lys Thr Gly Ser Gly Gly Tyr Gly Thr Thr Asn Asp Ile Ala Val Ile
210 215 220

Trp Pro Pro Asn Arg Ala Pro Ile Ile Ile Ala Val Tyr Phe Thr Gln
225 230 235 240

Ser Gln Ala Asp Ala Glu Ala Arg Asp Asp Val Ile Ala Ala Val Ala
245 250 255

Arg Ile Val Val Ala Ala Phe Leu Glu His His His His His His
260 265 270

<210> 4
<211> 271
<212> PRT
<213> Artificial Sequence

<220>
<223> seclet012

<400> 4

Met Ala Ala Glu Ala Leu Glu Gln Leu Ala Glu Leu Glu Lys Gln Ser
1 5 10 15

Gly Gly Arg Leu Gly Val Ala Val Leu Asp Thr Ala Thr Gly Arg Arg
20 25 30

Ile Gly Tyr Arg Gly Asp Glu Arg Phe Pro Met Cys Ser Thr Phe Lys
35 40 45

Val Leu Leu Ala Ala Ala Val Leu Ala Arg Val Asp Gln Gly Lys Glu
50 55 60

Gln Leu Asp Arg Arg Ile Ser Tyr Ser Lys Ser Asp Leu Val Ser Tyr
65 70 75 80

ES 2 585 054 B1

Ser Pro Val Thr Glu Lys His Val Gly Lys Gly Met Thr Val Ala Glu
85 90 95

Leu Cys Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu
100 105 110

Leu Leu Lys Gln Ile Gly Gly Pro Ala Ala Val Thr Ala Phe Leu Arg
115 120 125

Ser Ile Gly Asp Asn Thr Thr Arg Leu Asp Arg Arg Glu Pro Glu Leu
130 135 140

Asn Asn Ala Ala Pro Gly Asp Pro Arg Asp Thr Thr Thr Pro Ala Ala
145 150 155 160

Met Ala Ala Thr Leu Arg Arg Leu Leu Leu Gly Asp Ala Leu Ser Pro
165 170 175

Ala Ser Arg Ala Gln Leu Met Glu Trp Met Met Ala Asn Lys Thr Gly
180 185 190

Asp Asn Arg Leu Arg Ala Gly Leu Pro Ala Gly Trp Arg Ile Gly Asp
195 200 205

Lys Thr Gly Thr Gly Gly His Gly Thr Thr Asn Asp Ile Ala Val Ile
210 215 220

Trp Pro Pro Asn Arg Ala Pro Ile Ile Ile Ala Val Tyr Phe Thr Gln
225 230 235 240

Ser Gln Ala Asp Ala Ala Glu Arg Asp Glu Val Ile Ala Glu Val Ala
245 250 255

Arg Ala Val Val Glu Ala Phe Leu Glu His His His His His His
260 265 270

<210> 5
<211> 263
<212> PRT
<213> Artificial sequence

<220>
<223> GNCA

<400> 5

Met Ala Ala Gln Leu Ser Glu Gln Leu Ala Glu Leu Glu Lys Arg Ser
1 5 10 15

Gly Gly Arg Leu Gly Val Ala Val Leu Asp Thr Ala Thr Gly Arg Arg
20 25 30

Ile Ala Tyr Arg Gly Asp Glu Arg Phe Pro Met Cys Ser Thr Phe Lys
35 40 45

ES 2 585 054 B1

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



- ②① N.º solicitud: 201530877
②② Fecha de presentación de la solicitud: 19.06.2015
③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. Cl.: **G06F19/14** (2011.01)
C12N15/10 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	SEMBA, Y., et al., Ancestral amino acid substitution improves the thermal stability of recombinant lignin peroxidase from white-rot fungi, <i>Phanerochaete chrysosporium</i> strain UAMH 3641, <i>Protein Engineering, Design & Selection</i> , 2015, Vol. 28, páginas. 221–230. DOI: 10.1093/protein/gzv023, en línea 8/04/2015. Resumen, y apartado: "Materials and methods".	1-8
A	RISSE, V.A., et al., Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins, <i>Proteins</i> , 2014, Vol.82, páginas: 887-896. Resumen, y apartado: "Materials and methods".	1-8
A	TÖPFER, A., et al., Probabilistic Inference of Viral Quasispecies Subject to Recombination, <i>Journal of Computational Biology</i> , 2013, Vol.20, páginas: 113–123. DOI: 10.1089/cmb.2012.0232. Apartado: "Methods".	1-8
A	KLEINMAN, C.L., et al., A maximum likelihood framework for protein design, <i>BMC Bioinformatics</i> , 2006, Vol.7:326. DOI: 10.1186/1471-2105-7-326. Apartado: "Methods".	1-8

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
22.09.2016

Examinador
M. d. García Poza

Página
1/5



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

②① N.º solicitud: 201530877

②② Fecha de presentación de la solicitud: 19.06.2015

③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. Cl.: **G06F19/14** (2011.01)
C12N15/10 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	WHELAN, S., et al., A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach, 2001, Mol. Biol. Evol., Vol. 18, páginas: 691–699. Apartado: “Models and amino acid replacement”.	1-8

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
22.09.2016

Examinador
M. d. García Poza

Página
2/5

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06F, C12N

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, BIOSIS, NPL, XPESP

Fecha de Realización de la Opinión Escrita: 22.09.2016

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones 1-8	SI
	Reivindicaciones	NO
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones 1-8	SI
	Reivindicaciones	NO

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	SEMBA, Y., et al., Ancestral amino acid substitution improves the thermal stability of recombinant lignin peroxidase from white-rot fungi, <i>Phanerochaete chrysosporium</i> strain UAMH 3641, <i>Protein Engineering, Design & Selection</i> , 2015, Vol. 28, páginas . 221–230.	8/04/2015
D02	RISSO, V.A., et al., Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins, <i>Proteins</i> , 2014, Vol.82, págs: 887-896.	
D03	TÖPFER, A., et al., Probabilistic Inference of Viral Quasispecies Subject to Recombination, <i>Journal of Computational Biology</i> , 2013, Vol.20, páginas: 113–123.	
D04	KLEINMAN, C.L., et al., A maximum likelihood framework for protein design, <i>BMC Bioinformatics</i> , 2006, Vol.7:326.	
D05	WHELAN, S., et al., A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach, 2001, <i>Mol. Biol. Evol.</i> , Vol. 18, páginas:691–699.	

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

Ninguno de los documentos citados, o cualquier combinación relevante de ellos, revela un procedimiento para obtener la secuencia de un polipéptido pseudoancestral a partir de un alineamiento de N polipéptidos conocidos, utilizando una función de verosimilitud que depende de dicha secuencia y de las distancias entre dicha secuencia y las secuencias dadas, tal y como se recoge en la reivindicación 1.

En el estado de la técnica, el procedimiento utilizado, principalmente, para la obtención de un polipéptido ancestral se basa en la obtención del árbol filogenético y en modelos evolutivos. De este modo se obtienen proteínas ancestrales.

Tal es el caso del documento D01 que divulga un procedimiento de preparación de una enzima con propiedades mejoradas a partir de una sustitución de aminoácidos ancestrales. Para ello, partiendo de secuencias de aminoácidos conocidas de interés, se construye el árbol filogenético, a partir de cual se obtuvieron residuos de aminoácidos ancestrales. Utilizando estos aminoácidos se prepararon las enzimas con propiedades mejoradas.

Los documentos D02 a D05 divulgan procedimientos de preparación de proteínas (o virus en el caso de D03) a partir de la construcción del árbol filogenético.

En consecuencia, la invención, recogida en la reivindicación 1, y en las dependientes 2 a 8, es nueva y se considera que implica actividad inventiva (Arts. 6.1 y 8.1 LP).