

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 371 793**

51 Int. Cl.:
C07K 14/00 (2006.01)
C07K 14/47 (2006.01)
C07K 19/00 (2006.01)
G01N 1/28 (2006.01)
H01J 37/20 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **08737092 .0**
96 Fecha de presentación: **23.04.2008**
97 Número de publicación de la solicitud: **2167529**
97 Fecha de publicación de la solicitud: **31.03.2010**

54 Título: **CAPAS PROTEICAS Y SU USO EN MICROSCOPIA ELECTRÓNICA.**

30 Prioridad:
30.05.2007 US 807922

45 Fecha de publicación de la mención BOPI:
10.01.2012

45 Fecha de la publicación del folleto de la patente:
10.01.2012

73 Titular/es:
ISIS INNOVATION LIMITED
EWERT PLACE, EWERT HOUSE
SUMMERTOWN, OXFORD OX2 7SG, GB

72 Inventor/es:
SINCLAIR, John, Charles y
NOBLE, Martin, Edward, Mäntylä

74 Agente: **Ungría López, Javier**

ES 2 371 793 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

Capas proteicas y su uso en microscopia electrónica

5 La presente invención se refiere a capas proteicas que se repiten con regularidad en dos dimensiones. En un aspecto, las capas proteicas se basan en ensamblajes de oligómeros simétricos capaces de autoensamblarse a partir de los monómeros del ensamblaje de oligómeros. Las capas pueden tener poros con dimensiones del orden de nanómetros a cientos de nanómetros. Las capas proteicas son nanoestructuras que tienen muchos posibles usos, por ejemplo como matriz para soporte de entidades moleculares para microscopia electrónica o cristalografía de rayos X. En otro aspecto, la invención se refiere al uso de capas proteicas para realizar microscopia electrónica.

15 El documento WO-00/68248 divulga estructuras proteicas regulares basadas en ensamblajes de oligómeros simétricos capaces de autoensamblarse. En concreto, el documento WO-00/68248 divulga estructuras formadas a partir de protómeros proteicos (denominados "proteína de fusión" en el documento WO-00/68248), que comprende al menos dos monómeros (denominados "dominios de oligomerización" en el documento WO-00/68248), que son, cada uno, monómeros de un correspondiente ensamblaje de oligómero simétrico. El autoensamblaje de los monómeros en el ensamblaje oligomérico produce el ensamblaje de las propias estructuras regulares. Se divulgan varios tipos de estructuras, incluidas estructuras pequeñas y estructuras que se extienden en una, dos y tres dimensiones.

20 En el documento WO-00/68248, las orientaciones relativas de los monómeros dentro de los protómeros se seleccionan de modo que proporcionen la estructura regular deseada tras el autoensamblaje. Los monómeros se condensan a través de un rígido grupo de enlace que se selecciona cuidadosamente de modo que proporcione la orientación relativa necesaria de los monómeros en el protómero. Por ejemplo, en la producción del laboratorio publicada en el documento WO-00/68248, se realizó la selección del protómero usando un programa informático para modelar los monómeros conectados por un grupo de enlace en forma de un segmento de alfa-hélice intermedio continuo sobre un intervalo de longitudes que aumentan de forma creciente. Por tanto, por ejemplo, las matrices sugeridas en el documento WO-00/68248 que tienen una estructura regular que se repite en tres dimensiones se forman a partir de protómeros que comprenden dos monómeros de los respectivos ensamblajes oligoméricos diméricos o triméricos que son simétricos alrededor de un único eje de rotación. La orientación relativa de los dos monómeros se selecciona de modo que proporciona un ángulo específico de intersección entre el eje de simetría de rotación de los dos ensamblajes oligoméricos. Por tanto, hay una única fusión entre los dos ensamblajes oligoméricos y la orientación relativa de los ensamblajes oligoméricos está controlada mediante una cuidadosa selección del grupo de enlace que proporciona la fusión. El documento WO-00/68248 sólo publica la producción de laboratorio de estructuras proteicas de una pequeña jaula y un filamento que se extiende en una dimensión.

35 Cabe esperar que la aplicación de las enseñanzas del documento WO-00/68248 a las capas proteicas que se repiten en dos dimensiones se encontrara con las dificultades siguientes. En primer lugar, cabe esperar que haya una dificultad en el diseño debido al requisito de seleccionar la orientación relativa de los monómeros dentro del protómero adecuado para construir una capa. Esto probablemente reduciría el número de tipos de ensamblaje oligomérico disponible para formar una capa proteica y, por tanto, dificultaría la identificación de proteínas adecuadas. En segundo lugar, cabe esperar que durante el ensamblaje se encuentren dificultades prácticas. Las estructuras divulgadas en el documento WO-00/68248 dependen de la rigidez de la fusión entre monómeros en los protómeros que forma la fusión sencilla entre ensamblajes oligoméricos. El documento WO-00/68248 enseña que la orientación relativa de los monómeros en los protómeros controla la orientación relativa de los ensamblajes oligoméricos en la estructura resultante, por tanto cabe esperar que la flexión de la fusión lejos de la orientación relativa deseada reduzca la fiabilidad del autoensamblaje. Cabe esperar que dicho problema sea más agudo a medida que el tamaño de la unidad de repetición aumenta, de modo que se proporciona una restricción práctica en la producción fiable de matrices con un tamaño de poro relativamente grande.

50 Sería deseable proporcionar capas proteicas que tengan un tipo diferente de estructura en la que estos problemas previstos se podrían aliviar.

55 Se ha apreciado que un uso particularmente ventajoso de una capa proteica que se repite con regularidad en dos dimensiones es realizar microscopia electrónica de una entidad molecular. Por tanto, de acuerdo con la presente invención se proporciona un procedimiento de realizar microscopia electrónica de una entidad molecular, que comprende:

60 proporcionar una capa proteica que tiene una estructura que se repite con regularidad en dos dimensiones y que soporta entidades moleculares cada una fijada en una posición predeterminada en la estructura de repetición de la capa proteica; y realizar microscopia electrónica de la capa proteica como soporte de las entidades moleculares para obtener una imagen

65 Por tanto, la capa proteica actúa como soporte de entidades moleculares. Como las entidades moleculares, cada una en una posición predeterminada en la estructura de repetición de la capa proteica, las entidades moleculares son soportadas en una matriz regular. Esto proporciona ventajas significativas en la microscopia electrónica porque

permite la obtención de imágenes de un gran número de las entidades moleculares individuales en posiciones conocidas. Esto facilita varias formas de análisis de datos de la imagen derivada, de modo que se permite la investigación de la estructura de la entidad molecular.

5 El procedimiento se puede aplicar a cualquier capa proteica que se repite con regularidad en dos dimensiones, pero tiene aplicación concreta a una capa proteica que se repite regularmente en dos dimensiones, comprendiendo la capa proteica protómeros de proteínas cada uno de los cuales comprende al menos dos monómeros condensados genéticamente, siendo cada uno de los monómeros monómeros de un respectivo ensamblaje oligomérico, en los que los protómeros comprenden: Un primer monómero que es un monómero de un primer ensamblaje oligomérico que pertenece a un grupo puntual diédrico del orden O , donde O es igual a 3, 4 o 6, y que tiene un grupo de ejes de simetría rotacional O de orden 2 que se extiende en dos dimensiones; y un segundo monómero condensado genéticamente a dicho primer monómero en el que el segundo monómero es un monómero de un segundo ensamblaje oligomérico que tiene un eje de simetría rotacional de orden 2, los primeros monómeros de los protómeros están ensamblados en dicho primer ensamblaje oligomérico y los segundos monómeros de los protómeros están ensamblados en dichos segundos ensamblajes oligoméricos, estando dicho eje de simetría rotacional de dichos segundos ensamblajes oligoméricos de orden 2 alineado con uno de dicho grupo de ejes de simetría rotacional del orden 2 de uno de dichos primeros ensamblajes oligoméricos con dos protómeros dispuestos simétricos a su alrededor.

20 Como resultado del uso de un segundo ensamblaje oligomérico que tiene un eje de simetría rotacional del mismo orden 2 como el grupo de O ejes de simetría rotacional de dicho primer ensamblaje oligomérico, los ensamblajes oligoméricos se condensan con los ejes de simetría alineados y con 2 protómeros dispuestos simétricamente alrededor. Esto significa que hay una fusión por dos entre el primero y el segundo ensamblaje oligomérico.

25 Además, el patrón de repetición de la capa proteica deriva de la disposición de los ejes de simetría rotacional del primer ensamblaje oligomérico y no depende de la orientación relativa de los monómeros dentro del protómero. Como el primer ensamblaje oligomérico es diédrico, el grupo de O ejes de simetría de orden 2 son coplanares. Por tanto, los protómeros se ensamblan en una capa que tiene la misma simetría que el grupo de O ejes de simetría.

30 Por tanto, dichas capas proteicas se pueden diseñar seleccionando ensamblajes de oligómeros con una simetría adecuada para construir una capa que se repita en dos dimensiones. Los protómeros se producen de modo que comprendan monómeros de los ensamblajes de oligómeros seleccionados condensados entre sí. Posteriormente, se deja que los protómeros se autoensamblen en condiciones adecuadas.

35 Para ayudar a la comprensión, se hace referencia a la Figura 1, que ilustra un ejemplo concreto de dicha capa proteica 1. La Figura 1 sólo muestra una parte de la capa proteica 1 que se repite de forma indefinida en dos dimensiones. La capa proteica 1 se ensambló a partir de los protómeros 2. La capa proteica 1 tiene un primer ensamblaje oligomérico 3 que, en este ejemplo, pertenece a un grupo puntual diédrico de orden 4 y, por tanto, tiene un grupo de 4 ejes de simetría rotacional de orden 2 (además de un solo eje de simetría rotacional de orden 4).
40 Cada uno de los monómeros 5 del primer ensamblaje oligomérico 3 está condensado a un segundo monómero 6 de un segundo ensamblaje oligomérico 4 que, en este ejemplo, pertenece al grupo puntual diédrico de orden 2, de modo que tiene un eje de simetría rotacional de orden 2. Como resultado, los segundos monómeros 6 están ensamblados en los segundos ensamblajes oligoméricos 4 dispuestos con sus ejes de simetría rotacional de orden 2 alineados a lo largo de los ejes de simetría rotacional de orden 2 del primer ensamblaje oligomérico 3 y con una fusión por 2 veces entre el primero y el segundo ensamblaje oligomérico 3 y 4. Por tanto, la simetría de la capa proteica 1 es la misma que la simetría del grupo de cuatro ejes de simetría rotacional de orden 2, en este caso simetría rotacional de orden 4.

50 De acuerdo con esto, dichas capas proteicas implican el uso de una clase diferente de ensamblajes oligoméricos de la usada en el documento WO-00/682478. Esto proporciona el beneficio de que uno no está restringido por la necesidad de controlar la orientación relativa de los monómeros dentro del protómero. Por tanto, el diseño de la estructura proteica está asistido en que la orientación relativa de los monómeros dentro del protómero es una restricción menos crítica. De forma similar, es posible un ensamblaje más fiable de la capa proteica, como se describe con mayor detalle a continuación.

55 La presente invención se describirá a continuación con mayor detalle a modo de ejemplo no limitativo en referencia a las figuras adjuntas en las que:

60 La Figura 1 es un diagrama esquemático de una capa proteica
La Figura 2 es un diagrama esquemático de una capa proteica que incluye un ensamblaje oligomérico heterólogo;
La Figura 3 es una microfotografía electrónica de una capa proteica específica que se ha preparado;
La Figura 4 es un diagrama esquemático de un microscopio electrónico de transmisión; y
La Figura 5 es un diagrama de flujo de un procedimiento de realización de microscopía electrónica.

65 Las capas proteicas se pueden diseñar seleccionando ensamblajes oligoméricos que, cuando se condensan entre sí

con ejes de simetría rotacionales de orden 2 alineados entre sí, producen una unidad de repetición que es capaz de repetirse en dos dimensiones. Como la simetría de la unidad de repetición y, por tanto, la capa proteica como un todo, depende de la simetría de los ensamblajes oligoméricos, esto implica una selección de ensamblajes oligoméricos que tienen una estructura cuaternaria que proporciona simetrías adecuadas. Esto es una tarea sencilla, ya que las simetrías de los ensamblajes oligoméricos están, en general, disponibles en la literatura científica sobre proteínas, de, por ejemplo, The Protein Data Bank; H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov & P. E. Bourne; Nucleic Acids Research, 28 pág. 235-242 (2000) que es el único archivo mundial de datos de estructura de macromoléculas biológicas, también disponible a través de sitios web como <http://www.rcsb.org>.

En algunos casos, la unidad de repetición se repite en la misma orientación a través de la capa. En otros casos, dos o más unidades de repetición adyacentes juntas forman una célula unitaria que se repite en la misma orientación a través de la capa, pero con las unidades de repetición dentro de una célula unitaria dispuesta en diferentes orientaciones.

Ejemplos de ensamblajes oligoméricos que producen estructuras que se repiten con regularidad en dos dimensiones se proporcionan a continuación.

El primer ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden O , en el que O es igual a 3, 4 o 6, y, por tanto, tiene una estructura cuaternaria con ejes de simetría rotacional que se extienden en dos dimensiones, incluyendo un grupo de O ejes de simetría rotacional de orden 2 que son coplanares, además un solo eje de simetría rotacional de orden O que es perpendicular.

El segundo ensamblaje oligomérico tiene una estructura cuaternaria con un eje de simetría rotacional del mismo orden 2 que el grupo de O ejes de simetría rotacional de dicho primer ensamblaje oligomérico. Por ejemplo, el segundo ensamblaje oligomérico puede pertenecer a un grupo puntual diédrico de orden 2 o a un grupo puntual cíclico de orden 2. El segundo ensamblaje oligomérico no tiene un eje de simetría rotacional de orden O .

En el primer ensamblaje oligomérico ensamblado, inevitablemente y por definición, hay grupos de primeros monómeros dispuestos simétricamente alrededor de cada uno del grupo de O ejes de simetría rotacionales de orden 2 del primer ensamblaje oligomérico. Esto se debe a que los resultados de simetría de los monómeros idénticos dispuestos de este modo alrededor de los ejes de simetría rotacional.

Como resultado de los segundos monómeros condensados con el primer ensamblaje oligomérico dispuesto simétricamente alrededor de un grupo de O ejes de simetría rotacional de orden 2 del primer ensamblaje oligomérico, se deduce que el segundo ensamblaje oligomérico se mantiene con el grupo de segundos monómeros condensados también mantenidos simétricamente alrededor de uno del grupo de O ejes de simetría rotacional de orden 2 del primer ensamblaje oligomérico.

Además, inevitablemente y por definición, los segundos monómeros también se ensamblan en el segundo ensamblaje oligomérico en una disposición simétrica alrededor del eje de simetría rotacional de orden 2 del segundo ensamblaje oligomérico. Por tanto, el resultado del segundo ensamblaje oligomérico que tiene un eje de simetría rotacional del mismo orden 2 que el grupo de O ejes de simetría rotacional del primer ensamblaje oligomérico es que el primero y el segundo ensamblaje oligomérico se ensamblan con sus ejes de simetría de orden 2 alineados uno con otro. A partir de la simetría de ambos ensamblajes oligoméricos se deduce que esta es la disposición más estable. Esto tiene como resultado una fusión por dos entre el primero y el segundo ensamblaje oligomérico. En cada uno del primero y el segundo ensamblaje oligomérico hay 2 monómeros dispuestos alrededor del eje de simetría rotacional, estando cada uno de los monómeros condensados dentro de un protómero correspondiente a un monómero del otro ensamblaje oligomérico.

Como se ha mencionado anteriormente, el grupo de ejes de simetría rotacional no incluye todos los ejes de simetría rotacional del primer ensamblaje oligomérico. En su lugar, el grupo comprende los ejes de simetría rotacional del primer ensamblaje oligomérico que son del mismo orden 2 que el eje de simetría rotacional del segundo ensamblaje oligomérico.

La elección concreta de las simetrías del primero y segundo ensamblaje oligomérico tiene como resultado, al ensamblar los protómeros en la capa, la integración de los ensamblajes oligoméricos en sus ejes de simetría rotacional alineados. Por tanto, la disposición relativa de los ensamblajes oligoméricos condensados y, por tanto, la capa proteica como un todo, derivan de las disposiciones de los ejes de simetría rotacional del primer ensamblaje oligomérico y del segundo ensamblaje oligomérico. En particular, la capa proteica tiene la misma simetría que el grupo de O ejes de simetría rotacional de orden 2. La simetría de la capa proteica no depende de la orientación relativa de los monómeros dentro del protómero. En otras palabras, la capa proteica proporciona la ventaja de que los dos patrones de repetición dimensional de la capa proteica se pueden basar únicamente en las disposiciones de los ejes de simetría rotacionales de los ensamblajes oligoméricos. Esto proporciona ventajas en el diseño de una capa proteica facilitando seleccionar los ensamblajes oligoméricos adecuados para uso en la capa proteica. Durante el diseño, la orientación relativa de los monómeros dentro de un protómero individual en su forma no ensamblada pasa a ser una restricción mucho menor de la que está presente en, por ejemplo, el documento WO-00/68248.

También hay ventajas durante el autoensamblaje de la capa. En particular, la formación de una fusión por 2 entre dos ensamblajes oligoméricos dados tiene como resultado que la unión entre los dos ensamblajes oligoméricos sea relativamente rígida. Esto reduce el movimiento relativo de los ensamblajes oligoméricos durante el proceso de ensamblaje y ayuda a la formación fiable de la capa con los ensamblajes oligoméricos en las posiciones relativas correctas.

A continuación se describirán la forma y producción de los protómeros. Aunque la capa proteica usa protómeros que son diferentes en cuanto a que comprenden diferentes monómeros del documento WO-00/68248, la forma y producción de los protómeros per se, así como el polinucleótido que codifica los protómeros, pueden ser las mismas que las descritas en el documento WO-00/68248.

A continuación se describirán la naturaleza de los propios monómeros.

Los monómeros son monómeros de ensamblajes oligoméricos que son capaces de autoensamblarse en condiciones adecuadas para producir una capa proteica. La estructura secundaria y terciaria de los monómeros no tiene importancia en si misma siempre que se ensamblen en una estructura secundaria con la simetría requerida. No obstante, es ventajoso si la proteína se expresa y pliega con facilidad en un sistema de expresión heterólogo (usando, por ejemplo, un vector de expresión plasmídica en *E. coli*).

Los monómeros pueden ser proteínas naturales o pueden estar modificados con elementos peptídicos ausentes, sustituidos o añadidos a una proteína natural, siempre que las modificaciones no afecten sustancialmente al ensamblaje de los monómeros en su respectivo ensamblaje oligomérico. Se conocen dichas modificaciones, en sí mismas, para una serie de fines diferentes que se pueden aplicar a monómeros de la capa proteica. En otras palabras, el monómero puede ser un homólogo y/o fragmento y/o proteína de fusión de una proteína natural.

El monómero puede estar químicamente modificado, por ejemplo modificado postraduccionalmente. Por ejemplo, puede estar glicosilado o comprender residuos de aminoácidos modificados.

Aunque los monómeros pueden condensarse directamente entre sí, preferentemente los monómeros se condensan mediante un grupo de enlace de elementos peptídicos o no peptídicos. En general, la unión de dos proteínas mediante un grupo de enlace se conoce para otros fines y dichos grupos de enlace se pueden aplicar a la capa proteica.

Otro factor en la selección de ensamblajes oligoméricos adecuados es la localización y orientación de (a) los extremos de los primeros monómeros cuando están dispuestos en el primer ensamblaje oligomérico en su forma natural (es decir, no condensado a un segundo ensamblaje oligomérico) y (b) los extremos de los segundos monómeros cuando están dispuestos en el segundo ensamblaje oligomérico en su forma natural (es decir, no condensado a un primer ensamblaje oligomérico). Dicha información sobre la disposición de los extremos en el ensamblaje oligomérico en su forma natural está, en general, disponible para ensamblajes oligoméricos, por ejemplo del Protein Data Bank al que se ha hecho referencia anteriormente. Idealmente, estos extremos tendrán la misma separación y orientación, porque se condensarán entre sí en la capa proteica ensamblada para constituir la fusión por 2 dispuesta simétricamente alrededor de un eje de simetría rotacional. Es decir, no es esencial que la separación y orientación sean las mismas, ya que cualquier diferencia se puede acomodar mediante la deformación de los monómeros cerca de la fusión por 2 y/o mediante el uso de un grupo de enlace. Por tanto, como punto general, deberán escogerse ensamblajes oligoméricos en los que los extremos de ambos ensamblajes oligoméricos que se han de condensar entre sí en una fusión por 2 permitan la formación de la fusión sin evitar el ensamblaje de los ensamblajes oligoméricos y, por tanto, la capa proteica.

Considerando la deformación de los monómeros cerca de la fusión por 2 mencionada anteriormente, es deseable minimizar dicha deformación que tenderá a reducir la fiabilidad del proceso de ensamblaje. No obstante, si un grupo de enlace se condensa entre monómeros, el propio grupo de enlace puede capturar dicha deformación, al menos parcialmente. Esto reduce la deformación de los monómeros, de modo que se incrementa la fiabilidad del autoensamblaje porque el grupo de enlace no forma parte del proceso de ensamblaje en lo que respecta a que no forma parte de la proteína natural. Existe una ventaja concreta del uso de un grupo de enlace.

Además, el grupo de enlace puede diseñarse específicamente de modo que esté orientado respecto al primero y al segundo monómeros en el protómero en su forma normal, antes del ensamblaje, para reducir dichas diferencias en la posición y/u orientación de los extremos del primero y al segundo monómeros. Usando la posición y la orientación de los extremos del primero y al segundo monómeros en el primero y el segundo ensamblaje oligomérico en su forma natural que generalmente está disponible para los ensamblajes oligoméricos, como se ha tratado anteriormente, es posible diseñar un grupo de enlace adecuado usando técnicas de modelación convencionales.

Normalmente, los monómeros están condensados por sus extremos. Como alternativa, los monómeros pueden estar condensados en una localización alternativa en la cadena polipeptídica, siempre que el plegamiento nativo y la simetría del ensamblaje oligomérico natural sigan siendo iguales. Por ejemplo, uno de los monómeros se puede insertar en una porción estructuralmente tolerante del otro monómero, por ejemplo en un bucle que se extiende hacia fuera del ensamblaje oligomérico. Asimismo, es factible un truncamiento de un monómero y puede estimarse

mediante examen estructural.

Algunos ejemplos de simetrías para que los ensamblajes oligoméricos produzcan una capa proteica que se repite en dos dimensiones son los siguientes:

5 En estos ejemplos, el primer ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden O , en el que O es igual a 3, 4 o 6. Por tanto, el primer ensamblaje oligomérico tiene un eje de simetría rotacional principal de orden O y, también, O ejes de simetría rotacional de orden 2 que se extienden perpendiculares al eje de simetría rotacional principal. Con el fin de desarrollar una capa que se extienda en dos dimensiones, el segundo ensamblaje oligomérico se escoge de modo que tenga un eje de simetría rotacional de orden 2 para alinearse con los O ejes de simetría rotacional de orden 2 del primer ensamblaje oligomérico con una fusión por 2 entre el primero y el segundo ensamblaje oligomérico. Por tanto, en este caso, los O ejes de simetría rotacional de orden 2 constituyen el grupo de ejes de simetría rotacional del primer ensamblaje oligomérico, es decir N es igual a O .

15 En algunas clases de capa proteica, los protómeros son homólogos con respecto a los monómeros, es decir hay un solo tipo de protómero dentro de la capa proteica. En este caso, el segundo ensamblaje oligomérico puede pertenecer a un grupo puntual diédrico de orden 2.

Por ejemplo, la Tabla 1 representa algunos protómeros homólogos simples capaces de formar una capa proteica.

20

Tabla 1- Protómeros homólogos

Protómero	M	N	Simetría de la capa
d3d2	6	2	P622
d4d2	8	2	P422
d6d2	12	2	P622

25 En la Tabla 1, cada protómero se identifica con letras que representan los ensamblajes oligoméricos a los que pertenecen los respectivos monómeros del protómero. En concreto, la letra d representa un grupo puntual diédrico y el número siguiente identifica el orden del grupo puntual diédrico. En las siguientes dos columnas de la Tabla 1, se proporciona el número M de los primeros monómeros en el primer ensamblaje oligomérico y el orden N del grupo de ejes de simetría rotacional del primer ensamblaje oligomérico, que en este caso es 2. La columna final da la simetría de la capa proteica resultante. En cada uno de estos casos, el segundo ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden 2.

30

Por tanto, es fácil visualizar las capas proteicas. En concreto, el primer ensamblaje oligomérico se puede visualizar como un nódulo a partir del cual el grupo de O ejes de simetría rotacional de orden 2 se extienden hacia fuera en un plano común, perpendicular al principal ejes de simetría rotacional de orden O . Los segundos ensamblajes oligoméricos se pueden visualizar como enlaces lineales que se extienden desde el nódulo alineado con los respectivos del grupo de O ejes de simetría rotacional de orden 2 de los primeros ensamblajes oligoméricos. De este modo, es fácil visualizar la formación de la capa con poros en los espacios entre los ensamblajes oligoméricos. Por tanto, se verá que la simetría de la capa deriva de la disposición simétrica del grupo de O ejes de simetría rotacional de orden 2 de los primeros ensamblajes oligoméricos.

35

40

En un tipo de capa proteica en la que los protómeros son homólogos con respecto a los monómeros, el segundo ensamblaje oligomérico es un ensamblaje oligomérico homólogo. En este caso, la capa proteica consta únicamente de los protómeros.

45

En otro tipo de tal capa proteica en la que los protómeros son homólogos con respecto a los monómeros, el segundo ensamblaje oligomérico es un ensamblaje oligomérico heterólogo. En este caso, la capa proteica consiste en los protómeros y, además, los terceros monómeros se ensamblaron con dichos segundos monómeros en dicho segundo ensamblaje oligomérico.

50

Por tanto, el propio protómero no se puede ensamblar en la totalidad la capa proteica. Los segundos monómeros del ensamblaje oligomérico heterólogo no puede autoensamblarse en todo el ensamblaje oligomérico heterólogo en ausencia de los terceros monómeros de dicho ensamblaje oligomérico heterólogo. Esto proporciona ventajas durante la fabricación de las capas proteicas porque los primeros ensamblajes oligoméricos se pueden ensamblar sin el ensamblaje de una capa proteica entera que, de otro modo, podría alterar la producción del protómero. Esto permite la producción en un proceso de dos etapas.

55

Un ensamblaje oligomérico heterólogo concreto que se puede usar como ventaja como el segundo ensamblaje oligomérico es uno que comprende monómeros que tienen un sitio de unión capaz de unir la biotina o un péptido, y aptámeros que son capaces de unirse a dicho sitio de unión, preferentemente de forma no covalente. Los aptámeros se usan como segundo monómero del protómero. Los monómeros que tienen un sitio de unión capaz de unirse a

60

biotina son un tercer monómero de la capa proteica que no está genéticamente condensado dentro de un protómero. Al ensamblar el segundo ensamblaje oligomérico, los terceros monómeros se ensamblan entre sí y los aptámeros se ensamblan en el segundo ensamblaje oligomérico uniéndose cada uno a un correspondiente tercer monómero.

5 Esto se muestra esquemáticamente en la Figura 2, que muestra un ejemplo de una parte de la matriz proteica 1 que incluye un único segundo ensamblaje oligomérico 4 de este tipo, repitiéndose, de otro modo, la matriz proteica del mismo modo que en el ejemplo que se muestra en la Figura 1. En este ejemplo, el primer ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden 4 y., por tanto, tiene un grupo de cuatro ejes de simetría rotacional de orden 2. Cada uno de los monómeros 5 del primer ensamblaje oligomérico 3 está condensado a un segundo monómero 6 que es un aptámero. La matriz proteica 1 comprende además terceros monómeros 7 que están ensamblados entre sí como parte del segundo ensamblaje oligomérico 4. Los segundos monómeros 6 se ensamblan en el segundo ensamblaje oligomérico 4 uniéndose cada uno a un respectivo tercer monómero 7. Por tanto, en el segundo ensamblaje oligomérico 4, los segundos monómeros 6 se sujetan con la misma simetría que los terceros monómeros 7, pero los segundos monómeros 6 no están ensamblados entre sí.

15 Esto proporciona ventajas como ayuda para la formación de la matriz proteica. La matriz proteica 1 sigue teniendo una fusión por 2 entre un primer ensamblaje oligomérico 3 y un segundo ensamblaje oligomérico 4, debido a que ambos ensamblajes oligoméricos 3 y 4 tienen un eje de simetría de orden 2, como se ha tratado anteriormente. No obstante, esto se consigue sin que los propios segundos monómeros 6 se tengan que ensamblar entre sí. Esto ayuda al ensamblaje del primer ensamblaje oligomérico 4, en contraste con la matriz proteica 1 mostrada en la Figura 1 en la que tanto el primero como el segundo ensamblaje oligomérico 3 y 4 tienen que ensamblarse de forma simultánea. En su lugar, los terceros monómeros 7 se ensamblan y los segundos monómeros, 6, cada uno de forma individual, se ensamblan con un respectivo tercer monómero 7.

25 Normalmente, los terceros monómeros comprenden un sitio de unión. Dicho sitio de unión puede ser capaz de unirse a restos peptídicos o no peptídicos. En una realización preferida, el sitio de unión es capaz de unirse a biotina. En este caso, el tercer monómero puede ser de cualquier tipo que tenga dicho sitio de unión, por ejemplo estreptavidina, avidina o neutravidina.

30 Los términos “estreptavidina”, “avidina” o “neutravidina”, como se usan en el presente documento, abarcan variantes de estas moléculas, a menos que el contexto requiera lo contrario. Dichas variantes son, normalmente, homólogos de las secuencias originales, es decir normalmente son homólogos de las secuencias mostradas e la SEC ID N° 4 o la SEC ID N° 5. Las variantes pueden ser fragmentos de las secuencias originales o de homólogos de las secuencias originales. Las proteínas variantes pueden comprender secuencias adicionales (normalmente no estreptavidina, no avidina o no neutravidina) y, por tanto, ser proteínas de fusión que comprenden dichas secuencias originales, homólogos o fragmentos.

35 Preferentemente, las secuencias variantes conservan las propiedades estructurales de las secuencias originales, tal como cualquier propiedad estructural mencionada en el presente documento. Además, las secuencias variantes generalmente conservan la capacidad para unirse a biotina y/o un péptido (tal como el péptido de la SEC ID N° 3). En una realización, la secuencia variante puede ser reconocida por un anticuerpo que es capaz de reconocer la secuencia original. Por supuesto, las secuencias variantes conservarán la propiedad de formar una capa proteica como se ha descrito en el presente documento.

45 Los segundos monómeros que son aptámeros capaces de unirse al sitio de unión pueden ser cualquiera de una serie de marcadores peptídicos, incluidos, sin limitaciones, streptag I, streptag 11 o nanotag. Aptámeros preferidos son péptidos que tienen una longitud de 7 a 20 aminoácidos, por ejemplo de 9 a 15 aminoácidos de longitud. El aptámero puede tener homología con la SEC ID N° 3, que tiene, por ejemplo, al menos 6, 7 u 8 aminoácidos en común (es decir igual que) con la SEC ID N° 3.

50 En general, el primer ensamblaje oligomérico 1 puede ser de cualquier tipo que tenga la simetría requerida. Un ejemplo posible es ADAL (ácido delta-aminolevulínico deshidrogenasa) de E. coli. Más adelante se exponen otros criterios para la selección del primer ensamblaje oligomérico.

55 El aptámero se puede condensar a un extremo del primer monómero. Cuando se usa el extremo, el primer ensamblaje oligomérico debería poseer, preferentemente, un extremo cercano a un eje de simetría de orden 2 (normalmente dentro de los 15 Å).

60 Como alternativa, el aptámero puede, en general, está condensado en una posición distinta al extremo siempre que la estructura cuaternaria de las propiedades del primer ensamblaje oligomérico permanezca sustancialmente sin afectar y siempre que el aptámero sea uno que no requiera condensarse a un extremo. Por ejemplo, Streptag I requiere un extremo sin C con el fin de unirse a estreptavidina. De nuevo, es preferible que el aptámero se condense en una posición dentro de la secuencia peptídica del primer monómero que tiene como resultado el aptámero localizado en el ensamblaje oligomérico ensamblado en una posición cercana a un eje de simetría de orden 2 (normalmente dentro de los 15 Å).

65

Opcionalmente puede haber un grupo de enlace en el protómero entre el primer monómero y el segundo monómero que es un aptámero. Normalmente, el grupo de enlace podría tener una longitud dentro del intervalo de 1 a 10 aminoácidos. El grupo de enlace podría proporcionar, de forma ventajosa, flexibilidad que ayuda en el ensamblaje de la matriz. Por ejemplo, en el caso en el que el primer monómero es ALAD de E. coli y el segundo monómero es streptag I, se prefiere proporcionar un grupo de enlace de uno o dos aminoácidos de longitud.

Opcionalmente, las fusiones proteicas adicionales pueden condensarse genéticamente con cualquier extremo libre del primer monómero o tercer monómero. Esto podría realizarse para permitir la funcionalización de la matriz. Ejemplos no limitativos específicos de proteínas adicionales adecuadas son marcadores de hexa-histidina, péptidos de afinidad específicos, repeticiones de anquirina y calmodulina, cada uno de los cuales se ha demostrado en la literatura que es capaz de sufrir fusión genética con el extremo N de ALAD de E. coli-streptag I sin afectar a la capacidad de ese ensamblaje para autoensamblarse en matrices.

En otras clases de capa proteica, los protómeros son heterólogos con respecto a los monómeros, es decir hay dos o más tipos de protómero dentro de la capa proteica.

Para conseguir el ensamblaje de dos tipos de protómero, los dos tipos de protómero incluyen diferentes monómeros del mismo ensamblaje oligomérico heterólogo que puede pertenecer a un grupo puntual cíclico de orden 2. Por tanto, el primer tipo de protómero comprende un primer monómero que es un monómero de dicho primer ensamblaje oligomérico que pertenece a un grupo puntual diédrico de orden O, en el que O es igual a 3, 4 o 6, condensado genéticamente con un segundo monómero que es un monómero del segundo ensamblaje oligomérico, que es el ensamblaje oligomérico heterólogo que pertenece a un grupo de punto cíclico de orden 2. Además, el segundo tipo de protómero comprende un tercer monómero que es un monómero de dicho segundo ensamblaje oligomérico. En el segundo tipo de protómero, el tercer monómero está genéticamente condensado con un cuarto monómero que es un monómero de un tercer ensamblaje oligomérico, perteneciendo el tercer ensamblaje oligomérico a un grupo puntual diédrico de orden 1 o O.

Por tanto, cuando los protómeros de los diferentes tipos se dejan ensamblar, los ensamblajes oligoméricos heterólogos se ensamblan, de modo que unen los protómeros de los dos tipos. No obstante, un solo tipo de protómero no puede por sí mismo ensamblarse en la totalidad la capa proteica. Los monómeros individuales del ensamblaje oligomérico heterólogo no pueden autoensamblarse en todo el ensamblaje oligomérico heterólogo en ausencia de los otros monómeros diferentes de dicho ensamblaje oligomérico heterólogo. Esto proporciona ventajas durante la fabricación de las capas proteicas, ya que cada tipo de protómero puede producirse por separado y ensamblarse en un respectivo componente pequeño de la célula unitaria del patrón de repetición, como resultado de los monómeros del autoensamblaje del primer ensamblaje oligomérico homólogo, pero sin ensamblaje de una capa proteica completa. Esto es una ventaja de los protómeros heterólogos, porque el ensamblaje de la capa se puede evitar hasta que los componentes se juntan. Por otro lado, el ensamblaje de la capa podría dificultar la producción de los propios protómeros. Esto permite la producción en un proceso de dos etapas.

En los tipos más sencillos de capa proteica, el primer ensamblaje oligomérico de ambos tipos de protómero es un monómero de un ensamblaje oligomérico homólogo perteneciente a un grupo puntual diédrico. Por tanto, los tipos individuales de protómero pueden. Por ejemplo, la Tabla 2 representa algunos protómeros heterólogos simples capaces de formar una capa proteica.

Tabla 2- Protómeros heterólogos

Protómero	Componentes	1er protómero		2º protómero		Simetría de la capa
		M	N	M	N	
d3c2A + d3c2A*	D3/D3	6	2	6	2	P622
d4c2A + d4c2A*	D4/D4	8	2	8	2	P422
d6c2A + d6c2A*	D6/D6	12	2	12	2	P622
d3c2A + d2c2A*	D3/D2	6	2	4	2	P622
d4c2A + d2c2A*	D4/D2	8	2	4	2	P422
d6c2A + d2c2A*	D6/D2	12	2	4	2	P622

En la Tabla 2, la primera columna identifica los dos tipos de protómero. Cada protómero se identifica con letras que representan los ensamblajes oligoméricos a los que pertenecen los respectivos monómeros del protómero. En concreto, la letra d representa un grupo puntual diédrico y la letra c representa un monómero de un ensamblaje oligomérico heterólogo que pertenece a un grupo puntual cíclico. El subíndice representa, de nuevo, el orden del grupo puntual. Las letras A y A* en subíndice se usan para identificar los dos monómeros diferentes del mismo ensamblaje heterólogo.

En la Tabla 2, la segunda columna identifica los grupos de punto a los que pertenecen los componentes resultantes del ensamblaje de cada tipo de protómero. Una notación similar se usa como para los monómeros del protómero, excepto que las letras en mayúscula se usan para indicar al grupo puntual del componente al que se hace referencia. Por tanto, la letra mayúscula D indica que el componente pertenece a un grupo puntual diédrico y el número proporciona el orden del grupo puntual.

En las siguientes cuatro columnas de la Tabla 2, se proporciona el número M de los primeros monómeros en el primer ensamblaje oligomérico y cuatro oligómeros en el tercer ensamblaje oligomérico, así como el orden N (= 2) del grupo de O ejes de simetría rotacional del primer ensamblaje oligomérico y el tercer ensamblaje oligomérico. La última columna indica la simetría de la capa proteica resultante.

En todos los ejemplos de la Tabla 2, el primer ensamblaje oligomérico del primer tipo de protómero pertenece a un grupo de punto diédrico de orden O, en el que O es igual a 3, 4 o 6.

En los primeros tres ejemplos de la Tabla 2, el primer ensamblaje oligomérico del segundo tipo de protómero pertenece a un grupo puntual diédrico de orden L, en el que L es igual a O. Por tanto, estos tres ejemplos tienen espacialmente la misma disposición que los tres ejemplos de los correspondientes protómeros homólogos de la Tabla 1. En los primeros tres ejemplos de la Tabla 2, los primeros ensamblajes oligoméricos de los dos tipos de protómero pueden ser el mismo ensamblaje oligomérico o pueden ser un ensamblaje oligomérico diferente.

En los segundos tres ejemplos de la Tabla 2, el primer ensamblaje oligomérico del segundo tipo de protómero pertenece a un grupo puntual diédrico de orden L, en el que L es igual a 2. Estos tres ejemplos tienen espacialmente la misma disposición que los tres ejemplos de los correspondientes protómeros homólogos de la Tabla 1, excepto en lo siguiente. En lugar de la unión de los dos ensamblajes oligoméricos diédricos de orden O mediante un único ensamblaje oligomérico cíclico, el enlace entre los ensamblajes oligoméricos diédricos de orden O se extiende para formarse mediante una cadena que comprende dos ensamblajes oligoméricos cíclicos de orden 2 en ambos lados de un ensamblaje oligomérico diédrico de orden 2. Por tanto, se verá que la unidad de repetición del ensamblaje oligomérico heterólogo extiende de forma eficaz la longitud de los enlaces de la unidad de repetición entre los ensamblajes oligoméricos diédricos de orden O que se pueden considerar nódulos en la capa proteica. Por tanto, el tamaño de los poros dentro de la capa proteica también aumenta respecto al uso de los correspondientes protómeros homólogos.

Se cree que los ejemplos anteriores de capas proteicas representan la forma más simple de protómeros capaz de formar una capa proteica y por este motivo son los preferidos. No obstante, se apreciará que otros protómeros formados a partir de monómeros de ensamblajes oligoméricos que tienen simetrías adecuadas serán capaces de formar una capa proteica. Por ejemplo, otros protómeros homólogos que tienen un número más grande de monómeros que los indicados en la Tabla 1 serán capaces de formar una capa proteica. De un modo similar, otros protómeros heterólogos serán capaces de formar una capa proteica. Estos pueden incluir dos tipos de protómero que tienen un número más grande de monómeros que en los ejemplos de la Tabla 2 o pueden incluir más de dos tipos de protómeros.

Para cada uno de los monómeros hay una gran elección de ensamblajes oligoméricos que tienen la simetría requerida. La capa proteica no está limitada a ensamblajes oligoméricos concretos, porque, en principio, no se puede usar ningún ensamblaje oligomérico que tiene una estructura cuaternaria con la simetría requerida. No obstante, como ejemplos, la Tabla 3 indica algunas posibles elecciones de ensamblajes oligoméricos de varios grupos de punto, incluidos los de las Tablas 1 y 2.

Tabla 3 – Ejemplo de ensamblajes oligoméricos

Grupo puntual	Fuente	Nombre del ensamblaje oligomérico	Código PDB
P ₃ (T, 32)	<i>E. coli</i>	Dps	1DPS
	<i>S. epidermis</i>	EpiD	1G63
P ₄ (O, 432)	Ser humano	Cadena pesada de ferritina	2FHA
	<i>E. coli</i>	Dihidrolipoamida succiniltransferasa	1E2O
	<i>A. vinelandii</i>		1EAB
D ₂	Ser humano	Mn superóxido dismutasa	1AP5
	<i>P. falciparum</i>	Lactato deshidrogenasa	1CEQ

(continuación)

Grupo puntual	Fuente	Nombre del ensamblaje oligomérico	Código PDB
D ₃	Rata <i>E. coli</i>	6-piruvilo tetrahidropterin sintasa Aminoácido aminotransferasa	1B66 1/1L
D ₄	<i>E. coli</i> Gusano sipuncúlido	PurE Hemeritrina	1QCZ 2HMQ
D ₆	<i>S. typhimurium</i>	Glutamina sintetasa	1F1H
C _{2A} + C _{2A}	Ser humano	Cadenas alfa y beta de la caseína quinasa	1JWH
C _{3A} + C _{3A}	Colifago T4 VIH <i>Pseudomonas putida</i>	gp5 + gp27 N36 + C24 Naftaleno 1,2-dioxigenasa	1K28 1AIK 1NDO
C _{4A} + C _{4A}	<i>Eraquiópodo</i>	Hemeritrina	N/A

Por tanto, un protómero proteico o varios protómeros proteicos son capaces de ensamblarse en una capa proteica. Los monómeros del protómero pueden tener cualquier longitud, pero normalmente tienen una longitud de 5 a 1000 aminoácidos, preferentemente al menos 20 aminoácidos y/o, preferentemente, como máximo 500 aminoácidos.

Los polinucleótidos pueden codificar los protómeros proteicos. Normalmente, el polinucleótido también comprenderá una secuencia adicional más allá de los extremos 5 y/o 3 de la secuencia de codificación. Normalmente, el polinucleótido tiene una longitud de al menos tres veces la longitud del protómero codificado. El polinucleótido puede ser ARN o ADN, incluido ADN genómico, ADN sintético o ADNc. El polinucleótido puede ser bicatenario o monocatenario.

Los polinucleótidos pueden comprender nucleótidos sintéticos o modificados, tales como estructuras de metilfosfonato y fosforotioato o la adición de cadenas de acridina o polilisina en los extremos 3' y/o 5' de la molécula.

Dichos polinucleótidos se pueden producir y usar usando técnicas estándar. Por ejemplo, los comentarios realizados en el documento WO-00/68248 sobre ácidos nucleicos y sus usos se aplican igualmente a los polinucleótidos descritos en el presente documento.

Normalmente, los monómeros se combinan para formar protómeros mediante fusión de los respectivos genes a nivel genético (p. ej., eliminando el codón de terminación del gen en 5' y dejando una lectura dentro del marco hasta el gen en 3'). En este caso, el gen recombinante se expresa como un único polipéptido. Como alternativa, los genes se pueden condensar en una posición distinta a la del extremo terminal, siempre que la estructura cuaternaria de las propiedades del ensamblaje oligomérico permanezcan sustancialmente sin afectar. En particular, se puede insertar un gen dentro de una región estructuralmente tolerante de un segundo gen para producir una fusión dentro del marco.

Los vectores de expresión pueden comprender polinucleótidos de la invención y que son capaces de expresar un protómero proteico. Dichos vectores pueden también comprender iniciadores, promotores, potenciadores adecuados, además de otros elementos, tales como, por ejemplo, señales de poliadenilación que pueden ser necesarias y que están en la orientación correcta con el fin de permitir la expresión de proteínas.

Por tanto, la secuencia de codificación en el vector está unida operablemente a dichos elementos de modo que proporcione la expresión de la secuencia de codificación (normalmente en una célula). La expresión "operablemente unido" se refiere a una yuxtaposición en la que los componentes descritos están en una relación que les permite funcionar en el modo previsto.

El vector puede ser, por ejemplo, un vector plásmido, virus o fago. Normalmente, el vector tiene un origen de replicación. El vector puede comprender uno o más genes marcadores seleccionables, por ejemplo un gen de resistencia a ampicilina en el caso de un plásmido bacteriano, o un gen de resistencia para un vector fúngico.

Los promotores y otras señales de regulación de la expresión pueden seleccionarse de modo que sean compatibles con la célula huésped para la que se ha diseñado su expresión. Por ejemplo, los promotores de levadura incluyen los promotores GAL4 y ADH de *S. cerevisiae*, el promotor *nmt1* y *adh* de *S. pombe*. Los promotores de mamífero

incluyen el promotor de metalotioneína, que puede inducirse en respuesta a metales pesados tales como cadmio. También se pueden usar promotores virales, tales como el promotor de antígeno R grande de SV40 o promotores de adenovirus.

5 Se pueden usar promotores de mamífero, tales como promotores de b-actina. Especialmente preferidos son los promotores específicos de tejido. También se pueden usar promotores virales, por ejemplo la repetición terminal larga del virus de la leucemia murina Molones (MMLV LTR), el promotor LTR del virus de sarcoma de rous (RSV), el promotor de SV40, el promotor IE del citomegalovirus humano (CMV), adenovirus, promotores de HSV (tales como los promotores IE de HSV) o promotores de HPV, particularmente la región reguladora en 5' de HPV (URRR).

10 Otro procedimiento que se puede usar para la expresión de los protómeros proteicos es la expresión acelular, por ejemplo bacterianos, de levaduras o de mamíferos.

15 Las células se pueden modificar para expresar los protómeros de la invención. Dichas células incluyen líneas de células eucariotas superiores transitorias o, preferentemente, estables tales como células de mamífero o células de insecto, usando, por ejemplo, un sistema de expresión en baculovirus, células eucariotas inferiores, tales como células de levadura o procariotas tales como células bacterianas. Ejemplos concretos de células que se pueden modificar mediante inserción de vectores que codifican un polipéptido incluyen las células de mamífero HEK293T, CHO, HeLa, 293 y COS. Preferentemente, la línea celular seleccionada será una que no sólo sea estable, sino que también permita la glicosilación madura de un polipéptido. La expresión se puede conseguir en oocitos transformados.

20 Los protómeros proteicos, polinucleótidos, vectores o células pueden estar presentes en una forma sustancialmente aislada. Pueden estar en forma sustancialmente purificada, en cuyo caso en general comprenderán al menos 90 %, por ejemplo al menos 95 %, 98 % o 99 %, de las proteínas, polinucleótidos, células o masa seca de la preparación.

25 Los protómeros se pueden preparar usando los vectores y células huésped usando técnicas estándar. Por ejemplo, los comentarios realizados en el documento WO-00/68248 sobre procedimientos de preparación de protómeros (denominados "proteínas de fusión" en el documento WO-00/68248) se aplican igualmente a la preparación de los protómeros descritos en el presente documento.

30 La ensamblaje de la capa proteica a partir de los protómeros se puede realizar simplemente introduciendo los protómeros en condiciones adecuadas para el autoensamblaje de los monómeros de los ensamblajes oligoméricos. Normalmente, esto se realizará introduciendo los protómeros en solución, preferentemente en solución acuosa. Normalmente, las condiciones adecuadas corresponderán a aquéllas en las que la proteína natural se autoensambla en la naturaleza. Las condiciones adecuadas pueden ser aquéllas que se divulgan en el documento WO-00/68248.

En el caso de los protómeros homólogos, esto tiene como resultado un ensamblaje directo de la capa proteica.

40 En el caso de protómeros heterólogos, el ensamblaje se realiza, preferentemente, en varias etapas. En una primera etapa, cada tipo de protómero se ensambla por separado en un respectivo componente pequeño. En una segunda etapa, los componentes pequeños se juntan y ensamblan en la capa proteica. Cuando se usan varios protómeros heterólogos puede haber etapas adicionales intermedias entre la primera y la segunda etapa en las que los respectivos componentes pequeños se juntan y ensamblan en componentes intermedios más grandes.

45 A continuación se describirá un procedimiento por el cual se ha preparado una capa proteica específica que es un ejemplo del tipo que se muestra en la Figura 2.

50 Los protómeros consistían en un primer monómero que es ALAD de *E. coli* y un segundo monómero que es streptag 1. El tercer monómero fue estreptavidina.

Los protómeros se prepararon en un vector plasmídico de *E. coli* usando técnicas estándar. El vector plasmídico de *E. coli* fue un derivado de pUC19 que tiene la secuencia SEC ID N° 1. La secuencia del protómero es:

**MTMGSM TDLIQRPRRLRKSPALRAMFEETLSLNDLVLPFVVEEIDDYKAVEAPGVMRIPEK
HLAREIERIANAGIRSVMTFGISHHTDETGSDAWREDGLVARMSRICKQTVPEMIVMSDTCFC
EYTSHGHCGLVCEHGVNDATLENLKGQAVVAAAAGAXFIAPSAAMDGQVQAIRQALDAA
GFKDTAIMSYSTKFASSFYGPFREAAGSALKGDRKSYQMNPMNRREAIRESLLDEAQQANCL
MVKPAGAYLDIVRELRRTELPIGAYQVSGEYAMIKFAALAGAIDEKVVLESLSIKRAGA
55 DLIFS YFALDLAEKKILRRSAWRHPQFGG (SEQ ID NO. 2)**

La secuencia de Streptag I es:

AWRHPQFGG (SEC ID N° 3)

La secuencia de estreptavidina (como se usa en el trabajo que se describe en el presente documento) es:

**MET GLU ALA GLY ILE THR GLY THR TRP TYR ASN GLN LEU
GLY SER THR PHE ILE VAL THR ALA GLY ALA ASP GLY ALA
LEU THR GLY THR TYR GLU SER ALA VAL GLY ASN ALA GLU
SER ARG TYR VAL LEU THR GLY ARG TYR ASP SER ALA PRO
ALA THR ASP GLY SER GLY THR ALA LEU GLY TRP THR VAL
ALA TRP LYS ASN ASN TYR ARG ASN ALA HIS SER ALA THR
THR TRP SER GLY GLN TYR VAL GLY GLY ALA GLU ALA ARG
ILE ASN THR GLN TRP LEU LEU THR SER GLY THR THR GLU
ALA ASN ALA TRP LYS SER THR LEU VAL GLY HIS ASP THR
PHE THR LYS VAL LYS PRO SER ALA ALA SER (SEQ ID NO:4).**

5

Para referencia, la secuencia de avidina es:

**ARKCSLTGKW TNDLGSNMTI GAVNSRGEFT GTYITAVTAT SNEIKESPLH GTQNTINKRT
QPTFGFTVNW KFSESTTVFT GQCFIDRNGK EVLKTMWLLR SSVNDIGDDW KATRVGINIF
TRLRTQKE (SEQ ID NO:5).**

10

El gen que codifica la ácido 5-aminolaevulínico deshidratasa (ALAD) se amplificó a partir de ADN genómico del DH5alfa y se inserta en el vector de expresión DsRed-Express-streptagl descrito anteriormente para sustituir el casete génico de DsRed-Express.

15

Después se preparó un protómero de ALAD-streptagl. En el medio de expresión se incluyó IPTG 0,1 mM. La inducción de la expresión fue del siguiente modo: Un cultivo de 10 ml durante la noche de la cepa de expresión (en caldo LB que contiene 30 µg/ml de kanamicina) se diluyó a 1:100 en caldo KB fresco que contiene 30 µg/ml de kanamicina), las células se cultivaron con agitación a 37 °C hasta una densidad correspondiente a una DO₆₀₀ de 0,6 y, después, se indujo la expresión de la proteína diana mediante la adición de IPTG hasta una concentración final de 1 mM. El cultivo se mantuvo a 37 °C con agitación durante 3 horas adicionales antes de recoger las células mediante centrifugación (5000 g, 10 minutos, 4 °C). El sedimento celular se resuspendió en 20 ml de tampón A (NaCl 300 mM, EDTA 1 mM, HEPES 50 mM, pH 7,5). Las células se lisaron mediante sonicación y la fracción insoluble se recogió mediante centrifugación (25.000 g, 30 minutos, 4 °C). Esta fracción se disolvió en urea 8M y se centrifugó (25.000 g, 30 minutos, 4 °C) para eliminar las partículas insolubles. El material de urea solubilizado se concentró hasta 16 mg/ml y se pasó a través de un filtro de 0,22 µm. Después, una gota de este material (1 µl) se inyectó directamente en una gota más grande (5 µl) del tampón A:

25

En general se dispone de muchas opciones de expresión y purificación. Otro protocolo con éxito repetido es el siguiente:

30

1. Una única colonia de BL21 (DE3)Star de E.coli se transfirió desde una placa de agar Luria-Bertani a 500 ml de medio de Luria-Bertani que contiene 75 µg/ml de ampicilina e isopropiltio-beta-D-galactopiranósido (IPTG) 0,1 mM.

2. Este cultivo se incubó a 37°C con agitación durante 18 horas.

3. El cultivo se recogió mediante centrifugación (5.000 g, 5 minutos) y se resuspendió en 10 ml de tampón "GF" (NaCl 150 Mm, Tris-Hcl 50 mM, EDTA 1 mM, azida sódica 0,02 %, pH 8,0).

35

4. Las células se lisaron usando sonicación, congelación-descongelación, reactivos de lisis celular (p. ej., "Bugbuster"), o lisozima y tratamiento con ADNasa. Estas son técnicas estándar en la materia.

5. La fracción insoluble se eliminó mediante centrifugación (30.000 g, 30 minutos).

6. La proteína de fusión se purificó a partir de la fracción soluble usando sefarosa Strep-tactin (IBA GmbH) de acuerdo con las instrucciones de los fabricantes.

40

7. La proteína eluida se separó de la contaminación con destiobiotina que es el resultado de la columna de Strep-tactin, por medio de cromatografía de exclusión por tamaño usando una matriz de superosa 6 y tampón GF.

8. La proteína purificada se pudo almacenar a 4 °C durante al menos 6 meses.

El protómero ALAD-streptag I purificado (~1 mg/ml) se mezcló con estreptavidina nuclear comercialmente disponible en cantidades equimolares. El autoensamblaje comenzó inmediatamente y las matrices proteicas resultantes se visualizaron por medio de microscopia electrónica de transmisión. La Figura 3 muestra una micrografía electrónica de transmisión teñida negativamente de la matriz proteica, siendo la célula unitaria de 13 nm x 13 nm. Se realizó procesamiento de imágenes de las micrografías electrónicas para potenciar la calidad de la imagen. En particular, la micrografía electrónica se sometió a transformación de Fourier, se filtró usando un filtro derivado de grupo espaciador y se promedió, y después se reconstruyó.

Las capas proteicas tienen numerosos usos diferentes. En general, dichos usos aprovecharán la estructura de repetición regular y/o los poros que están presentes dentro de la estructura. Las capas proteicas pueden diseñarse de modo que tengan poros con dimensiones previstas del orden de nanómetros a cientos de nanómetros. Las capas se pueden diseñar con un tamaño de poro adecuado para un uso deseado.

Los tamaños de poro altamente definidos, de tamaño inusual y finamente controlados de las matrices o capas proteicas, junto con la estabilidad de sus estructuras, las convierten en ideales para las aplicaciones que requieren materiales microporosos con tamaños de poro en el intervalo que se ha mencionado. Como ejemplo, se espera que las matrices o capas sean útiles como elemento de filtro o tamiz molecular para procedimientos de filtración o separación. En este uso, los tamaños de poro que se pueden conseguir y la capacidad para diseñar el tamaño de un poro son particularmente ventajosos.

En otra clase de uso, las entidades moleculares se fijarían a la capa proteica. Dicha fijación se puede realizar usando técnicas convencionales. Las entidades moleculares pueden ser cualquier entidad de un tamaño adecuado, normalmente una entidad macromolecular, por ejemplo proteínas, polinucleótidos, tales como ADN, o entidades no biológicas. Las entidades moleculares pueden ser una única molécula o un complejo de varias moléculas. Como tales, cabe esperar que las capas proteicas sean útiles como matrices biológicas para transportar entidades moleculares, por ejemplo para usar en la liberación de fármacos o para cristalizar entidades moleculares.

La fijación de las entidades moleculares a la capa proteica se puede efectuar mediante varias formas.

Algunos abordajes implican "marcar" cualquiera o ambos protómeros proteicos (u otro componente de la capa) o las entidades moleculares de interés. En este contexto, el marcaje es la adición covalente a cualquiera o ambos protómeros proteicos (u otro componente de la capa) o a las entidades moleculares diana, de una estructura conocida como marcador o marcador de afinidad que forma fuertes interacciones con una estructura diana. Normalmente, para este fin se usan cortos motivos peptídicos (p. ej., hélices enrolladas heterodiméricas, tales como el ácido "velero" y péptidos base). En el caso del protómero proteico (u otro componente de la capa), un una entidad molecular que es una proteína, esto se puede conseguir condensando genéticamente el marcador con un componente de la capa proteica o la entidad molecular, que es la expresión de una versión genéticamente modificada de la proteína para que porte una secuencia adicional de elementos peptídicos que constituyen la marcador, por ejemplo en uno de sus extremos o en una región bucle. Procedimientos alternativos de añadir un marcador incluyen la modificación covalente de una proteína después de que se ha expresado, mediante técnicas tales como la tecnología de interna.

En un abordaje, la estructura diana puede ser otro marcador unido al otro del protómero proteico o de la entidad molecular diana, es decir ambos componentes de la capa y de la entidad molecular diana incluyen marcadores de afinidad complementarios para que se unan entre sí.

En otro abordaje, la estructura diana puede formar parte del protómero proteico (u otro componente de la capa) o de la entidad molecular diana, es decir uno de un componente de la capa y de la entidad molecular diana tiene marcadores de afinidad que tienen afinidad por el otro componente de la capa y de la entidad molecular diana. Por tanto, para fijar la entidad molecular a la capa proteica, un componente de la capa puede incluir, en una posición predeterminada en los protómeros, un marcador de afinidad fijado a la entidad molecular de interés. Como alternativa, la entidad molecular puede tener en una posición predeterminada en la entidad molecular, un marcador de afinidad fijado a un componente de la capa.

Cuando se sabe que un componente de la capa proteica forma fuertes interacciones con una secuencia peptídica conocida, dicha secuencia peptídica puede usarse como marcador a añadir a la entidad molecular diana. Cuando no se conoce dicha pareja de unión estrecha, se pueden identificar marcadores adecuados por medio de detección selectiva. Los tipos de detección selectiva posibles son técnicas de visualización en fagos o enfoques de bibliotecas químicas redundantes para producir un gran número de diferentes péptidos cortos (p. ej., 3-50 aminoácidos). Los elementos peptídicos para unión más estrecha se pueden identificar usando técnicas estándar, por ejemplo amplificación y secuenciación en el caso de bibliotecas de expresión en fagos o por medio de secuenciación peptídica en el caso de bibliotecas redundantes,

Un abordaje alternativo es cuando la propia entidad molecular diana se expresa como una fusión genética directa con un componente de la capa.

Otro abordaje alternativo es realizar modificaciones químicas específicas de la matriz con el fin de proporcionar medios de fijación alternativos covalentes o basados en la afinidad. Por ejemplo, la derivación específica de sitio de los grupos sulfhidrilo accesibles en la matriz se puede usar para la incorporación de grupos de nitrilo-ácido triacético (NTA) que a su vez se puede usar para la unión de iones metálicos y, por tanto, proteínas diana ricas en histidina.

5 Para fijar la entidad molecular a la capa proteica usando un marcador de afinidad sobre la capa o la entidad molecular, se puede dejar que la entidad molecular se difunda en y, por tanto se fije, una capa proteica preformada, por ejemplo se puede hibridar la entidad molecular unida en sus configuraciones de energía menor en la capa proteica usando enfriamiento controlado en una criocorriente de nitrógeno líquido. Como alternativa, las entidades moleculares se pueden mezclar con los protómeros durante la formación de la capa proteica para ensamblarse con la capa.

En otra clase de usos, se podrían incorporar proteínas con propiedades útiles como uno de los protómeros.

15 Un uso en el que una entidad se fija a la capa proteica es realizar una cristalografía de rayos X de las entidades moleculares. En este caso, la estructura regular de la capa proteica permitir mantener las entidades moleculares en una posición predeterminada con respecto a una estructura de repetición, de modo que se mantienen en una matriz regular en una orientación regular. La cristalografía de rayos X es importante en investigación bioquímica y el diseño razonado de fármacos.

20 La capa proteica que tiene una matriz de entidades moleculares sobre ella se puede estudiar usando técnicas cristalográficas de rayos X estándar. Cabe esperar que el uso de la capa proteica como soporte de la cristalografía de rayos X proporcione numerosas y significativas ventajas sobre la tecnología y el protocolo actuales para la cristalografía de rayos X, incluidas las siguientes:

- 25 (1) Serán necesarias cantidades significativamente menores de la molécula (probablemente en el orden de microgramos y no en miligramos). Esto permitirá la determinación de algunas dianas previamente no tratables.
 (2) El uso de marcadores de afinidad permitirá determinar la estructura sin el requisito típico de una serie de etapas de purificación.
 30 (3) No habrá necesidad de cristalizar la entidad molecular. Esta es una etapa difícil y, en ocasiones infranqueable, en la determinación tradicional de la estructura mediante rayos X.
 (4) No habrá necesidad de obtener derivados cristalinos para cada nueva estructura de cristal para obtener la información de fase necesaria. Dado que la mayoría de la materia de dispersión será capa proteica conocida en cada caso, la determinación de la estructura puede ser automática y un usuario de ordenado podrá realizarlo rápidamente con poca o ninguna experiencia en cristalografía.
 35 (5) Los complejos de una proteína con sustancias químicas (sustratos/fármacos) y con otras proteínas se pueden analizar sin requerir condiciones totalmente nuevas de cristalización.
 (6) Cabe esperar que el procedimiento sea extremadamente rápido y aplicable universalmente, lo que proporcionará enormes ahorros en tiempo y costes.

40 Para uso en catalización de biotransformaciones, las enzimas se pueden fijar a la capa proteica o incorporar en la capa proteica.

45 Para uso en almacenamiento de datos, puede ser posible fijar una proteína que es óptica o electrónicamente activa. Un ejemplo es la bacteriorodopsina, pero se pueden usar muchas otras proteínas en esta capacidad. En este caso, la capa proteica mantiene la proteína fijada en una matriz altamente ordenada, de modo que se permite abordar la matriz. La capa proteica podría superar las limitaciones de tamaño de las matrices existentes para mantener las proteínas para usar en almacenamiento de datos..

50 Para uso en una visualización, puede ser posible fijar una proteína que sea fotoactiva o fluorescente. En este caso, la capa proteica mantiene la proteína fijada en una matriz altamente ordenada, de modo que se permite abordar la matriz para visualizar una imagen.

55 Para uso en separación de cargas, una proteína que es capaz de llevar a cabo un proceso de separación de cargas puede fijarse a la capa proteica o incorporarse en la capa proteica. Después, se puede inducir la proteína para que lleve a cabo la separación, por ejemplo "bioquímicamente" mediante un "combustible", tal como ATP, u ópticamente en el caso de un centro fotoactivo tal como clorofila o una proteína fotoactiva como rodopsina. De este modo se pueden realizar varios procesos de separación de cargas, por ejemplo bombeo iónico o desarrollo de una carga fotovoltaica.

60 Para uso como nanoalambre, una proteína que es capaz de conducción eléctrica puede fijarse a la capa proteica o incorporarse en la capa proteica. Usando una capa proteica anisotrópica podría ser capaz de proporcionar la capacidad para llevar la corriente en una dirección concreta.

65 Para uso como motor, se pueden incorporar en la capa proteica proteínas que son capaces de inducir expansión/contracción.

Se pueden usar matrices proteicas como molde. Por ejemplo, el silicio se podría difundir o, por el contrario, impregnar, en los poros de la matriz proteica, llenando de este modo parcial o completamente los intersticios de la matriz. El material proteico que comprende la matriz original puede, si es necesario, eliminarse mediante, por

5

Otro uso en el que una entidad se fija a la capa proteica es realizar microscopia electrónica de las entidades moleculares. Esto se puede realizar para determinar la estructura de las entidades. Las entidades pueden ser de cualquier tipo, incluida una macromolécula (p. ej., una proteína o ADN) o un complejo macromolecular (p. ej., un complejo de una macromolécula con una o más especies moleculares distintas).

10

A modo de antecedentes se describirá primero las técnicas conocidas de microscopia electrónica.

La Figura 4 muestra esquemáticamente un microscopio electrónico de transmisión 10 dispuesto del siguiente modo. Una fuente de electrones 11 produce electrones. Un sistema de lentes objetivas¹² dirige un haz de electrones desde la fuente 11 sobre la muestra 13. Un sistema de lentes de imagen 14 dirige los electrones transmitidos a través de la muestra 13 sobre un sensor 15 que produce una imagen. La imagen puede ser una imagen enfocada o puede ser un patrón de difracción, siendo este último útil cuando la entidad se presenta en una matriz regular (p. ej., tubos de moléculas, cristales 2D, o matrices helicoidales). La información de múltiples imágenes, correspondientes a múltiples visiones diferentes de las especies moleculares, pueden combinarse después para producir una reconstrucción en 3D.

15

20

La preparación y presentación de la muestra en el microscopio se realiza del siguiente modo.

25

En la práctica, las muestras 13 se presentan al haz de electrones dentro del soporte para muestras de un microscopio electrónico. En general, las muestras 13 se montan sobre una rejilla de cobre. Esta puede estar recubierta con una capa fina de carbono depositado que puede, a su vez, ser continua a través de los orificios de la rejilla o pueden estar deliberadamente incompletos, de modo que se dejan orificios en los que la muestra flota (una capa de carbono de tipo "encaje").

30

Los detalles del protocolo de montaje de la muestra dependen de si la muestra se va a visualizar o no en condiciones criogénicas.

35

Para las condiciones criogénicas, la muestra 13 puede introducirse en un medio potenciado con un agente crioprotector para minimizar la tendencia a formar hielo a temperaturas bajas. Ejemplos de agentes crioprotectores incluyen glucosa y trehalosa. Además, al entorno de la muestra 13 se puede añadir un constituyente potenciador del contraste. Un ejemplo de agente de potenciación de contraste es el tanino. Tras la crioprotección, la muestra 13 se introduce sobre la rejilla de cobre (posiblemente recubierta), el exceso de muestra y de medio de inclusión se extrae mediante transferencia para producir una muestra 13 de un espesor no superior a 1000 Å y la rejilla se introduce en un ambiente a temperaturas criogénicas (< -73 °C). La velocidad del enfriamiento es un factor importante para evitar la formación de hielo y el consiguiente daño en la muestra durante la congelación. El enfriamiento rápido se puede conseguir introduciendo la muestra 13 en nitrógeno líquido en una corriente de nitrógeno gaseoso a temperaturas inferiores a -153 °C o en un baño de líquido menos volátil (como propano) a temperaturas criogénicas. Se pueden usar etapas mecánicas para garantizar una introducción rápida y reproducible de la rejilla de cobre en el ambiente de congelación.

40

45

Cuando las muestras 13 no se van a presentar en solución vítrea congelada (es decir, en condiciones que no son criogénicas) se introduce una solución de la sustancia de la que se va a obtener una imagen sobre una rejilla de cobre recubierta por carbono, se deja un periodo de tiempo para que la muestra se adsorba sobre la capa de carbono y, después, se extrae el exceso de muestra y solución mediante transferencia. Para potencial el contraste de imágenes y minimizar las consecuencias perjudiciales de daños por radiación, la muestra 13 se puede teñir. Dado que las muestras biológicas demuestran una dispersión intrínsecamente baja, los tintes usados son, en general, densos a los electrones y, por tanto, dan una fuerte dispersión. Por tanto, los tintes usados son, en general, "tintes negativos". Las imágenes registradas son oscuras cuando el tinte lo es y son más claras en las regiones en las que el tinte es excluido por la presencia de la muestra. Un ejemplo de tinte negativo es acetato de uranilo.

50

55

La recolección de datos se realiza del siguiente modo.

En el caso de derivar una imagen enfocada, en realidad las imágenes se registran lejos de un enfoque perfecto. Aunque esto se realiza para generar contraste en la imagen, tiene como resultado una degradación de la imagen. Específicamente, los términos de Fourier calculados a partir de la imagen se modulan mediante una "Función de transferencia de contraste" (CTF), que modula las amplitudes de los términos de Fourier de un modo que sea una función del correspondiente ángulo de dispersión. En general, los términos de Fourier corregidos se pueden recuperar mediante escalado adecuado una vez que se ha caracterizado la extensión del desenfoque y astigmatismo. Para un desenfoque dado, la CTF adoptará un valor de cero para los términos de Fourier correspondientes a ángulos de dispersión concretos. Por tanto, estos términos no se pueden recuperar mediante posprocesamiento. Para llenar los

60

65

orificios correspondientes en espacios recíprocos, las imágenes se registran a un intervalo de desenfoques, de modo que los términos de Fourier que se han modulado a cero (o cerca de cero) en una imagen registrada a un desenfoque tendrán una amplitud mensurable en otro desenfoque.

- 5 La interacción inelástica de electrones con la muestra tiene como resultado el depósito de energía que, a su vez, causa daños en la muestra. Estos daños degradan la estructura de las moléculas dentro de la muestra. Por esta razón, las imágenes y los patrones de difracción se registran usando una dosis relativamente baja de electrones. Esta limitación experimental significa que la proporción entre la señal y el ruido es relativamente baja en las imágenes registradas de cada especie molecular capturada dentro del campo de visión de una imagen. Esto se traduce en que cada imagen porta una información de resolución sobre la estructura de la muestra relativamente
- 10 baja. En general, la potenciación de la proporción señal:ruido se consigue realizando un promedio eficaz de las imágenes de múltiples moléculas que se observan en las mismas (o similares) orientaciones con respecto al haz de electrones.
- 15 Cada imagen se aproxima a una protección de la distribución de la densidad de electrones (o, con mayor precisión, el potencial) de las especies moleculares. Por tanto, una única imagen de una única molécula no contiene suficiente información para deducir la estructura 3D de dicha molécula. Por tanto, las imágenes se tienen que registrar a partir de la muestra en múltiples orientaciones con respecto al haz.
- 20 Para las estructuras periódicas, los componentes de Fourier se pueden medir directamente registrando el patrón de difracción en lugar de una imagen de la muestra. Este abordaje evita la complicación de la modulación mediante la CTF, aunque todavía se tienen que corregir otras características del experimento y del instrumento en el posprocesamiento. Para muestras periódicas (p. ej., cristales 2D o matrices helicoidales), la dispersión se concentra en direcciones diferenciadas que son características del tamaño y de la forma de la unidad repetida (es decir, la célula unitaria), lo que da lugar a puntos de difracción en el patrón de dispersión en lugar de una función de dispersión continua. Este proceso de "amplificación Bragg" conlleva señales de fácil registro. Una ventaja adicional del registro de los patrones de difracción es que las intensidades del patrón disperso (es decir, la propiedad que se registra) son independientes de movimientos globales de la muestra durante la exposición. Dichos movimientos pueden deberse a fluctuación térmica, así como por el calentamiento específico y la carga de la muestra producidos
- 25 por el haz de electrones. Una desventaja de registrar el patrón de dispersión en lugar de los electrones enfocados (Es decir, un patrón de difracción en lugar de una imagen) es que el registro del patrón disperso pierde la información de la fase para los términos de Fourier. Al mismo tiempo, se pueden corregir las imperfecciones locales si se recoge una imagen de las mismas, pero no (trivialmente) si se recoge un patrón de difracción.
- 30 En el caso de la tomografía de electrones, se obtienen imágenes de un único ejemplo de la especie que se va a visualizar con una dosis extremadamente baja a un intervalo de orientaciones. Por tanto, se obtiene una imagen de una única especie molecular. Esto aborda una posible crítica de otros abordajes: Cada molécula representativa de una muestra podría ser sutilmente diferente, lo que hace que tanto el promedio de múltiples imágenes y la reconstrucción 3D sean inadecuados. Tiene la desventaja de que la dosis de electrones que puede tolerar una única especie se extiende sobre las imágenes en múltiples orientaciones: esto, en último término, limita la resolución de la reconstrucción en 3D que se puede conseguir.
- 35
- 40

El análisis de los datos se realiza del siguiente modo.

- 45 El protocolo usado para analizar los datos de la microscopía electrónica depende principalmente de si la muestra es periódica (es decir, cristalina 2D o presentada en una matriz helicoidal) o aperiódica, es decir presentada como partículas aisladas que pueden o no tener simetría rotacional local, pero que carece de una simetría traslacional significativa. En ambos casos, cuando se han recolectado imágenes (en lugar de datos de difracción), el desenfoque y el astigmatismo de la muestra se identifican mediante análisis de la distribución de la intensidad de las regiones transformadas de Fourier de la imagen. En base a estos valores, que pueden variar a lo largo de la imagen, se puede calcular una corrección adecuada para compensar los efectos de la CTF.
- 50

Un tipo de análisis de datos es la reconstrucción de una partícula. Esto permite la reconstrucción de una imagen tridimensional (3D) a partir de las imágenes de entidades individuales, del siguiente modo.

- 55 Para muestras no periódicas se registran imágenes (en lugar de patrones de difracción). El análisis comienza con la localización de muestras sobre la imagen registrada. Para las moléculas biológicas no teñidas, esto presenta un problema significativo: La proporción señal:ruido inherentemente baja significa que las moléculas pueden no ser evidentes frente al fondo. Incluso si son visibles, dichas moléculas pueden tener una imagen tan mala que impida la caracterización de su orientación en comparación con otras imágenes de la misma especie molecular. Este problema empeora cuando la especie que se va a visualizar es pequeña. En la práctica, no es fácilmente posible aplicar ME convencional a las muestras no cristalinas de macromoléculas (o complejos macromoleculares) con un peso molecular combinado inferior a 125 kDa.
- 60
- 65 Después de localizar múltiples especies moleculares para ensamblar un "grupo de datos" de imágenes (ruidosas), la siguiente etapa es la clasificación. En esta etapa, las imágenes de la especie molecular se agrupan, de modo que

las que representan visiones similares están asociadas entre sí. Particularmente, cuando se ha usado un soporte de carbono puede haber un grupo limitado de dichas visiones presente en el grupo de datos. Se efectúa el promedio de las imágenes de partículas que entran dentro de dichos conjuntos para proporcionar "promedios de clase": Las orientaciones relativas de un grupo de promedios de clase se determinan por medio de una "línea común" o un enfoque similar. En última instancia, esto permite que la información de múltiples visiones diferentes en un espacio recíproco para permitir la reconstrucción 3D.

Otro tipo de análisis de datos es el análisis cristalográfico bidimensional (2D). Esto es aplicable a muestras periódicas. Los datos se pueden recolectar como imágenes o como patrones de difracción.

En las imágenes de una matriz cristalina, el reconocimiento de la geometría y la localización de la matriz proporciona un medio fácilmente explotado de predecir la localización de las múltiples copias de la especie de la que se va a obtener la imagen. El promedio se puede obtener en espacio real (cuando se suman las imágenes de una célula unitaria individual) o en espacio recíproco. En este último abordaje, la imagen se transforma por Fourier para producir un grupo puntual de difracción que son el resultado de la dispersión por dicha parte de la imagen que tiene un carácter periódico, es decir mediante la matriz ordenada de moléculas. El resto de la dispersión (es decir, la intensidad que no entra en la posición de los puntos de difracción) procede del fondo y el ruido. Por tanto, se puede realizar el promedio de múltiples células unitarias en el campo de visión fijando todas las intensidades valle a cero y realizando otra transformación de Fourier. Este procedimiento se denomina filtración de Fourier. Tanto las imágenes de las que se ha realizado el promedio en espacio real como las filtradas por Fourier se pueden potenciar mediante un procedimiento de "enderezamiento". En este procedimiento se pueden identificar las distorsiones locales de la matriz (en general mediante un procedimiento de autocorrelación) y se usa para corregir la imagen para generar una imagen que prevalecerá si la matriz no está sujeta a ninguna distorsión local.

Los patrones de difracción de la matriz cristalina se pueden usar para medir directamente las amplitudes de los componentes de Fourier. Las fases para estos términos se pueden establecer únicamente usando procedimientos análogos a los usados en la cristalografía de proteínas. Estos incluyen reemplazo isomorfo (RI), reemplazo molecular (RM) y modificación de densidad (MD). Para el RI, la difracción se tiene que medir antes y después de la adición (o sustitución) de una parte de la estructura. Para el RM, se puede usar una estructura o distribución de densidad de electrones conocidas para calcular las fases para la estructura desconocida. Para la MD, deben estar disponibles las fases para los términos de resolución baja (p. ej., de análisis de imágenes como se indica anteriormente) y el escalonamiento de los términos de alta resolución se realiza mediante imposición iterativa de promediar y aplanar por disolvente, incluyendo los términos de resolución alta creciente, en el procedimiento a medida que se extiende la fase.

La disposición del cristal con respecto al haz se puede deducir a partir de la geometría aparente de los puntos de difracción, que pueden registrarse directamente o calcularse mediante transformación de Fourier de una imagen, siempre que se conozca la geometría de la unidad de repetición en el cristal (la geometría de la célula unitaria). Cuando se conoce la estructura de una parte significativa de la matriz, también se puede usar una imagen calculada de esta parte de la matriz para evaluar la orientación de la matriz en una imagen registrada de forma experimental. Por tanto, la información de múltiples imágenes en múltiples orientaciones, recogidas a múltiples ángulos de inclinación, se puede combinar fácilmente para llevar a cabo la reconstrucción de la imagen en 3D.

A continuación se describirá la aplicación a la obtención de imágenes de entidades moleculares sobre una matriz proteica de soporte. Se obtienen beneficios porque las entidades se soportan, cada una, en una posición predeterminada en la estructura de repetición de la capa proteica.

Se puede usar un microscopio electrónico de transmisión convencional 10, como se muestra en, por ejemplo, la Figura 4. La obtención de imagen se realiza usando el procedimiento que se muestra en la Figura 5.

Primero, en la etapa S1 se prepara una matriz proteica que tiene las entidades moleculares fijadas a la misma. Esto se realiza usando las técnicas descritas anteriormente. Con la matriz proteica se prepara una muestra 13 para el microscopio electrónico de transmisión 10 usando procedimientos estándar, como se ha tratado anteriormente.

Dos abordajes para fijar la entidad a la capa proteica son los siguientes.

En el primer abordaje, la entidad se añade a una solución (o suspensión) que contiene la capa proteica. Por tanto, las entidades se fijan a la capa en solución. Después, la capa resultante se somete a la preparación de la muestra como se ha descrito anteriormente para microscopía crioelectrónica o para microscopía no crioelectrónica, con o sin tinción.

En el segundo abordaje, la capa proteica se deposita primero sobre la capa de carbono de una rejilla de cobre recubierta para formar el soporte de muestras del microscopio electrónico 10. Posteriormente se introduce la entidad. En este caso, una suspensión de la matriz proteica se coloca sobre la rejilla recubierta de carbono, se deja que se produzca la adsorción, se eliminan el exceso de crisalina y de solución circundante y se introduce una solución de la especie diana. Después de una incubación en la que se produce la unión de la diana a la crisalina, se

retira el exceso de la diana y de la solución circundante. La posterior preparación de la muestra es como se ha descrito anteriormente para microscopia crioelectrónica o para microscopia no crioelectrónica.

5 Para una resolución óptima en la estructura de la entidad molecular, es preferible que las entidades moleculares se alineen con orientaciones idénticas con respecto a cada eje. En la etapa S2, que es opcional, las entidades moleculares se alinean con respecto a la matriz proteica.

Dos posibles procedimientos de alineación molecular que se pueden implementar, bien de forma independiente o en combinación, son los siguientes.

10 Un primer procedimiento de alineación es aplicar un campo eléctrico con un vector paralelo al eje de simetría principal del "primer" componente de la capa proteica con el fin de alinear las entidades moleculares en virtud de sus dipolos intrínsecos o inducidos.

15 Un segundo procedimiento de alineación aprovecha las interacciones polares y/o hidrófobas entre las entidades moleculares y la capa proteica mediante un procedimiento de hibridación térmica durante el cual las moléculas diana se enfrían lentamente hasta conformaciones de energía mínima idénticas.

20 En la etapa S3, se obtienen imágenes para derivar una imagen. Dicha recolección de datos se realiza usando protocolos estándar, por ejemplo como se ha descrito antes para la ME convencional. A modo de ejemplo, las imágenes se pueden recolectar a una serie de etapas de desenfoque y, también, empleando la etapa de inclinación del microscopio para obtener la imagen de la matriz mediante un intervalo de ángulos. Cuando ha tenido éxito la orientación de las moléculas diana, también se puede recolectar de forma útil una serie de imágenes de difracción de electrones.

25 En la etapa S4, se realiza el análisis de los datos de las imágenes. Se pueden aplicar diversas técnicas de análisis de datos del siguiente modo.

30 Cuando ha sido posible imponer una orientación aproximadamente común de cada molécula diana unida con respecto a la matriz subyacente se puede realizar un análisis de datos cristalográfico 2D, como se ha descrito anteriormente. Esto permite una reconstrucción 3D de la molécula diana que se va a derivar.

35 En teoría, las herramientas para la reconstrucción de imágenes de una partícula se pueden aplicar a una reconstrucción de imágenes de matrices periódicas en 2D y, cuando esto proporciona una reconstrucción de imágenes mejorada, dicho abordaje también se toma para obtener imágenes de capas proteicas y de entidades moleculares fijadas. Cuando es adecuado, también se usan procedimientos híbridos, en los que algunas técnicas computacionales de cristalografía 2D se combinan con técnicas computacionales de análisis de imagen de partícula única.

40 Cuando no ha sido posible imponer una orientación aproximadamente común de cada molécula diana unida se aplica una combinación de los procedimientos indicados anteriormente para la reconstrucción 3D de partícula única y cristalografía 2D. En esta combinación, se identifican los componentes de la propia matriz proteica y se restan de la imagen.

45 Los componentes de la matriz proteica se pueden derivar como se ha descrito anteriormente a partir de un análisis de una o más imágenes registradas de una capa proteica y de entidades moleculares fijadas. Como alternativa, los componentes de la matriz proteica se pueden derivar a partir de una imagen de referencia adquirida por separado o de una imagen almacenada adquirida anteriormente.

50 Esto permite la identificación de los componentes de la matriz de cada imagen para su eliminación. La imagen de diferencia resultante es una imagen de las entidades en aislamiento que se habrían registrado si las entidades estuvieran dispuestas en el espacio en posiciones con el mismo patrón de repetición como la estructura de la capa proteica, aunque en una orientación parcialmente aleatoria.

55 A partir de entonces se realiza una reconstrucción de partícula única, como se ha descrito anteriormente. Este procedimiento se agiliza por el hecho de que la capa proteica se encontrará en posiciones en la imagen fácilmente previstas, como consecuencia de su unión en ubicaciones conocidas sobre la capa proteica, cuya ubicación y orientación se identifican con facilidad. La resta de la imagen de referencia realiza con eficacia la primera etapa de la reconstrucción en 3D de una única partícula (selección de partícula) como se ha descrito anteriormente. De forma similar, es probable que se aplique un grado de alineación de las moléculas y ello contribuya a la clasificación de las partículas.

Variantes

65 En el presente documento se hace referencia a homólogos de secuencias de proteínas. Normalmente, dichos homólogos tienen una homología de al menos 70 %, preferentemente una homología de al menos 80, 90%, 95%,

97% o 99%, por ejemplo, sobre una región de al menos 15, 20, 30, 100 o más aminoácidos contiguos. La homología se puede calcular sobre la base de la identidad de los aminoácidos (en ocasiones denominada "homología difícil").

Por ejemplo, el paquete UWGCG proporciona el programa BESTFIT, que se puede usar para calcular la homología (por ejemplo usada en sus parámetros por defecto) (Devereux y col. (1984) Nucleic Acids Research 12, pág. 387-395). Los algoritmos PILEUP y BLAST se pueden usar para calcular la homología o secuencias alineadas (tal como identificar secuencias equivalentes o correspondientes (normalmente en sus parámetros predeterminados), por ejemplo como se ha descrito en Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S, F y col. (1990) J Mol Biol 215:403-10.

El software para realizar análisis BLAST está a disposición del público en el National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar primero pares de secuencia de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia problema que coinciden o satisfacen alguna puntuación T umbral de valor positivo cuando se alinea con una palabra de la misma longitud en una secuencia de la base de datos. T es el umbral de la puntuación de la palabra vecina (Altschul y col., anteriormente). Estas coincidencias de la palabra vecina inicial actúan como semillas para iniciar las búsquedas para encontrar los HSP que las contienen. Las coincidencias con la palabra se extienden en ambas direcciones a lo largo de cada secuencia todo lo que la puntuación de la alineación acumulada se pueda incrementar. Las extensiones para las coincidencias de la palabra en cada dirección se detienen cuando: La puntuación de la alineación acumulada se salga de la cantidad X a partir de su valor máximo alcanzado; la puntuación acumulada llega a cero o menor, debido a la acumulación de una o más alineaciones de residuos que puntúan negativo; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la velocidad de la alineación. El programa BLAST usa como valores predeterminados una longitud de palabra (W) de 11, las alineaciones (B) de la matriz de puntuación BLOSUM62 de 50 (véase Henikoff y Henikoff (1992) Proc. Natl. Acad. Sci. USA 89: expectativa (E) de 10, M= 5, N= 4, y una comparación de ambas hebras.

El algoritmo BLAST realiza un análisis estadístico de la similitud entre dos secuencias (véase, p. ej., Karlin y Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). Una medida de la similitud proporcionada por el algoritmo BLAST es la menor probabilidad de la suma ((P(N)), que proporciona una indicación de la probabilidad de que se produzca al azar una coincidencia entre dos secuencias de aminoácidos. Por ejemplo, una secuencia se considera similar a otra secuencia si la menor probabilidad de la suma en una comparación de la primera secuencia con la segunda secuencia es inferior a aproximadamente 1, preferentemente inferior a aproximadamente 0,1 más preferentemente inferior a aproximadamente 0,01 y, más preferentemente, inferior a aproximadamente 0,001.

Normalmente, la secuencia homóloga difiere en al menos 2, 5, 10, 20 o más mutaciones (que pueden ser sustituciones, deleciones o inserciones de aminoácidos). Normalmente, la secuencia homóloga difiere como máximo en 5, 10, 20 o más mutaciones (que pueden ser sustituciones, deleciones o inserciones de aminoácidos). Normalmente, hasta el 40 % de los aminoácidos de la secuencia está mutado. Estas mutaciones se pueden medir a través de cualquiera de las regiones mencionadas anteriormente en relación con el cálculo de la homología. Preferentemente, las sustituciones son sustituciones conservadoras. Estas se definen de acuerdo con la Tabla siguiente. Los aminoácidos del mismo bloque en la segunda columna y, preferentemente en la misma línea en la tercera columna, pueden sustituirse entre sí:

ALIFÁTICOS	No polar	G A P
		I L V
	Polar-sin carga	T M C S
		N Q
	Polar-con carga	D E
		K R
AROMÁTICOS		HFWY

ES 2 371 793 T3

LISTADO DE SECUENCIAS

<110> Isis Innovation Limited

5 <120> Capa proteica

<130> N101226A

<150> US 11/807.922

10 <151> 30/05/2007

<160> 5

<170> Patentín versión 3,4

15 <210> 1

<211> 3583

<212> ADN

<213> Secuencia artificial

20 <220>

<223> Derivado del vector pUC19

<220>

25 <221> misc_feature

<222> (709) .. (709)

<223> n e s a , c , g o t

<400> 1

30

```

agcgccaat acgcaaaccg cctctccccg cgcggtggcc gattcattaa tgcagctggc      60
acgacagggt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc      120
tactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa      180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgggatc catgacagac      240
ttaatccaac gcctcgctcg cctgcgcaaa tctcctgcgc tgcgcgctat gtttgaagag      300
acaacactta gccttaacga cctggtgttg ccgatctttg ttgaagaaga aattgacgac      360
tacaaagccg ttgaagccat gccaggcgtg atgcgcattc cagagaaaca tctggcacgc      420
gaaattgaac gcacgcgcaa cgccggtatt cgttccgtga tgacttttgg catctctcac      480
cataccgatg aaaccggcag cgatgcctgg cgggaagatg gactggtggc gcgtatgtcg      540
cgcatctgca agcagaccgt gccagaaatg atcgttatgt cagacacctg cttctgtgaa      600
tacacttctc acggtcactg cgggtgtgctg tgcgagcatg gcgtcgacaa cgacgcgact      660
ctggaaaatt taggcaagca agccgtggtt gcagctgctg caggtgcana cttcatcgcc      720
ccttccgccg cgatggacgg ccaggtagag gcgattcgtc aggcgctgga cgctgcggga      780
tttaaagata cggcgattat gtcgtattcg accaagttcg cctcctcctt ttatggcccc      840
ttccgtgaag ctgccggaag cgcattaaaa ggcgaccgca aaagctatca gatgaaccca      900
atgaaccgtc gtgaggcgat tcgtgaatca ctgctggatg aagcccaggg cgcaaactgc      960

```

ES 2 371 793 T3

ctgatggtta aacctgctgg agcgtaccto gacatcgtgc gtgagctgcg tgaacgtact 1020
 gaattgccga ttggcgcgta tcaggtgagc ggtgagtatg cgatgattaa gttcgccgcg 1080
 ctggcgggtg ctatagatga agagaaagtc gtgctcgaaa gcttaggttc gattaagcgt 1140
 gcggggtgcg atctgatttt cagctacttt gcgctggatt tggctgagaa gaagattctg 1200
 cgtagatctg cgtggcgcca tccgcagttt ggtgggtaag gccgcgactc tagaattcca 1260
 actgagcggc ggtcgtacc attaccaact tgtctgggtg caaaaataat aggcctacta 1320
 gtcggccgta cgggcccctt cgtctcgcgc gtttcgggta tgacggtgaa aacctctgac 1380
 acatgcagct cccggagacg gtcacagctt gtctgtaagc ggatgccggg agcagacaag 1440
 cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcccgat 1500
 cagagcagat tgtactgaga gtgcaccata tgcgggtgta aataccgcac agatgcgtaa 1560
 ggagaaaata ccgcatcagg cggccttaag ggcctcgtga tacgcctatt tttataggtt 1620
 aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcbc 1680
 ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 1740
 taacctgat aaatgcttca ataatttga aaaaggaaga gtatgagtat tcaacatttc 1800
 cgtgtcggcc ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcaccagaa 1860
 acgctgggta aagtaaaaga tgetgaagat cagttgggtg cacgagtggtt ttacatcgaa 1920
 ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg 1980
 atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga cgccgggcaa 2040
 gagcaactcg gtcgccgat aactattct cagaatgact tgggtgagta ctcaccagtc 2100
 acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc 2160
 atgagtgata aactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta 2220
 accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcggtg ggaaccggag 2280
 ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgctgtagc aatggcaaca 2340
 acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca acaattaata 2400
 gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc 2460
 tggtttattg ctgataaatc tggagccggt gagcgtgggt ctgcgggtat cattgcagca 2520
 ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg gagtcaggca 2580
 actatggatg aacgaaatag acagatcgtc gagataggtg cctcactgat taagcattgg 2640
 taactgtcag accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa 2700
 tttaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaact cccttaacgt 2760

ES 2 371 793 T3

gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat 2820
 cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg 2880
 gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg cttcagcaga 2940
 gcgcagatac caaatactgt ccttctagtg tagccgtagt taggccacca cttcaagaac 3000
 tctgtagcac cgcctacata cctcgtcttg ctaatcctgt taccagtggc tgctgccagt 3060
 ggcgataagt cgtgtcttac cgggttgac tcaagacgat agttaccgga taaggcgcag 3120
 cggtcgggct gaacggggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc 3180
 gaactgagat acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaaag 3240
 gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca 3300
 gggggaaaac cctggtatct ttatagtctt gtcgggtttc gccacctctg acttgagcgt 3360
 cgatttttgt gatgctcgtc agggggggcg agcctatgga aaaacgccag caacgcggcc 3420
 tttttacggt tcttgccctt ttgetggcct tttgctcaca tgttctttcc tgcggtatcc 3480
 cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc tcgccgcagc 3540
 cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aag 3583

<210> 2

<211> 339

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> Secuencia del protómero

<220>

<221> misc_feature

<222> (164) .. (164)

15 <223> Xaa puede ser cualquier aminoácido natural

<400> 2

Met Thr Met Gly Ser Met Thr Asp Leu Ile Gln Arg Pro Arg Arg Leu
 1 5 10 15

Arg Lys Ser Pro Ala Leu Arg Ala Met Phe Glu Glu Thr Thr Leu Ser
 20 25 30

Leu Asn Asp Leu Val Leu Pro Ile Phe Val Glu Glu Glu Ile Asp Asp
 35 40 45

Tyr Lys Ala Val Glu Ala Pro Gly Val Met Arg Ile Pro Glu Lys His
 50 55 60

ES 2 371 793 T3

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

ES 2 371 793 T3

305

310

315

320

Leu Ala Glu Lys Lys Ile Leu Arg Arg Ser Ala Trp Arg His Pro Gln
325 330 335

Phe Gly Gly

<210> 3

<211> 9

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Ligando artificial de estreptavidina

10

<400> 3

Ala Trp Arg His Pro Gln Phe Gly Gly
1 5

<210> 4

15 <211> 127

<212> PRT

<213> Streptomyces avidinii

<400> 4

20

ES 2 371 793 T3

Met Glu Ala Gly Ile Thr Gly Thr Trp Tyr Asn Gln Leu Gly Ser Thr
1 5 10 15

Phe Ile Val Thr Ala Gly Ala Asp Gly Ala Leu Thr Gly Thr Tyr Glu
20 25 30

Ser Ala Val Gly Asn Ala Glu Ser Arg Tyr Val Leu Thr Gly Arg Tyr
35 40 45

Asp Ser Ala Pro Ala Thr Asp Gly Ser Gly Thr Ala Leu Gly Trp Thr
50 55 60

Val Ala Trp Lys Asn Asn Tyr Arg Asn Ala His Ser Ala Thr Thr Trp
65 70 75 80

Ser Gly Gln Tyr Val Gly Gly Ala Glu Ala Arg Ile Asn Thr Gln Trp
85 90 95

Leu Leu Thr Ser Gly Thr Thr Glu Ala Asn Ala Trp Lys Ser Thr Leu
100 105 110

Val Gly His Asp Thr Phe Thr Lys Val Lys Pro Ser Ala Ala Ser

115

120

125

<210> 5
<211> 128
5 <212> PRT
<213> Gallus gallus

<400> 5

ES 2 371 793 T3

Ala Arg Lys Cys Ser Leu Thr Gly Lys Trp Thr Asn Asp Leu Gly Ser
 1 5 10 15

Asn Met Thr Ile Gly Ala Val Asn Ser Arg Gly Glu Phe Thr Gly Thr
 20 25 30

Tyr Ile Thr Ala Val Thr Ala Thr Ser Asn Glu Ile Lys Glu Ser Pro
 35 40 45

Leu His Gly Thr Gln Asn Thr Ile Asn Lys Arg Thr Gln Pro Thr Phe
 50 55 60

Gly Phe Thr Val Asn Trp Lys Phe Ser Glu Ser Thr Thr Val Phe Thr
 65 70 75 80

Gly Gln Cys Phe Ile Asp Arg Asn Gly Lys Glu Val Leu Lys Thr Met
 85 90 95

Trp Leu Leu Arg Ser Ser Val Asn Asp Ile Gly Asp Asp Trp Lys Ala
 100 105 110

Thr Arg Val Gly Ile Asn Ile Phe Thr Arg Leu Arg Thr Gln Lys Glu
 115 120 125

REIVINDICACIONES

1. Un procedimiento de realizar microscopia electrónica de una entidad molecular, que comprende:
- 5 proporcionar una capa proteica que tiene una estructura que se repite con regularidad en dos dimensiones y que soporta entidades moleculares cada una fijada en una posición predeterminada en la estructura de repetición de la capa proteica; y realizar microscopia electrónica de la capa proteica como soporte de las entidades moleculares para obtener una imagen.
- 10 2. Un procedimiento de acuerdo con la reivindicación 1, en el que dicha etapa de proporcionar una capa proteica que soporta entidades moleculares cada una fijada en una posición predeterminada en la estructura de repetición de la capa proteica; y realizar microscopia electrónica de la capa proteica como soporte de las entidades moleculares para derivar una imagen.
- 15 3. Un procedimiento de acuerdo con la reivindicación 1 o 2, que además comprende, antes de la etapa de realizar microscopia electrónica, alinear las entidades moleculares con respecto a la matriz proteica.
4. Un procedimiento de acuerdo con la reivindicación 3, en el que la etapa de alinear las entidades moleculares con respecto a la matriz proteica comprende aplicar un campo eléctrico a la matriz proteica o enfriar la matriz proteica hasta un estado de energía mínima.
- 20 5. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones precedentes, que además comprende realizar un análisis de datos cristalográfico en dos dimensiones de una imagen.
- 25 6. Un procedimiento de acuerdo con la reivindicación 5, en el que el análisis de datos comprende identificar los componentes de la matriz proteica y restarlos de la imagen derivada en dicha etapa de realizar microscopia electrónica para derivar una imagen de las entidades moleculares y realizar una reconstrucción de partícula única de la imagen de las entidades moleculares.
- 30 7. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que la capa proteica es una capa proteica que se repite con regularidad en dos dimensiones, comprendiendo la capa proteica protómeros cada uno de los cuales comprende al menos dos monómeros condensados genéticamente, siendo cada uno de los monómeros monómeros de un respectivo ensamblaje oligomérico, en los que los protómeros comprenden:
- 35 un primer monómero que es un monómero de un primer ensamblaje oligomérico perteneciente a un grupo puntual poliédrico de orden O , en el que O es igual a 3, 4 o 6, y que tiene un grupo de O ejes de simetría rotacional de orden 2 que se extiende en dos dimensiones; y
- 40 un segundo monómero genéticamente condensado con dicho primer monómero, en el que el segundo monómero es un monómero de un segundo ensamblaje oligomérico que tiene un eje de simetría rotacional de orden 2,
- 45 los primeros monómeros de los protómeros se ensamblan en dichos primeros ensamblajes oligoméricos y los segundos monómeros de los protómeros están ensamblados en dichos segundos ensamblajes oligoméricos, estando dicho eje de simetría rotacional de dichos segundos ensamblajes oligoméricos de orden 2 alineado con uno de dicho grupo de ejes de simetría rotacional de orden 2 de uno de dichos primeros ensamblajes oligoméricos con dos protómeros dispuestos simétricamente alrededor.
8. Un procedimiento de acuerdo con la reivindicación 7, en el que el segundo ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden 2 o a un grupo puntual cíclico de orden 2.
- 50 9. Un procedimiento de acuerdo con la reivindicación 7 u 8, en el que los protómeros son homólogos o heterólogos con respecto a los monómeros.
10. Un procedimiento de acuerdo con la reivindicación 9, en el que los protómeros son homólogos con respecto a los monómeros y el segundo ensamblaje oligomérico es un ensamblaje oligomérico heterólogo de dichos segundos monómeros y de terceros monómeros, comprendiendo además dicha capa proteica dichos terceros monómeros ensamblados con dichos segundos monómeros en dicho segundo ensamblaje oligomérico.
- 55 11. Un procedimiento de acuerdo con la reivindicación 10, en el que los terceros monómeros son monómeros que tienen un sitio de unión capaz de unir la biotina o un péptido, y dichos segundos monómeros son aptámeros que son capaces de unirse a dicho sitio de unión.
- 60 12. Un procedimiento de acuerdo con la reivindicación 11, en el que dichos terceros monómeros son estreptavidina y/o dichos segundos monómeros son Streptag I (SEC ID N° 3).
- 65 13. Un procedimiento de acuerdo con la reivindicación 9, en el que los protómeros son heterólogos con respecto a los monómeros y la capa proteica comprende protómeros proteicos de dos tipos,

- comprendiendo el primer tipo de protómero un primer monómero que es un monómero de dicho primer ensamblaje oligomérico que pertenece a un grupo puntual diédrico de orden O , en el que O es igual a 3, 4 o 6, condensado genéticamente con un segundo monómero que es un monómero de dicho segundo ensamblaje oligomérico, siendo dicho segundo ensamblaje oligomérico un ensamblaje oligomérico heterólogo perteneciente a un grupo puntual cíclico de orden 2, y
- 5 comprendiendo el segundo tipo de protómero un tercer monómero que es un monómero de dicho segundo ensamblaje oligomérico genéticamente condensado con un cuarto monómero que es un monómero de un tercer ensamblaje oligomérico, perteneciendo dicho tercer ensamblaje oligomérico a un grupo puntual diédrico de orden 2 o O .
- 10 14. Un procedimiento de acuerdo con la reivindicación 13, en el que dicho tercer ensamblaje oligomérico pertenece a un grupo puntual diédrico de orden O , siendo dicho tercer ensamblaje oligomérico el mismo que dicho primer ensamblaje oligomérico.
- 15 15. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que las entidades moleculares están fijadas por medio de uno de:
- siendo un componente de la capa proteica un marcador de afinidad, estando las entidades moleculares fijadas a los respectivos marcadores de afinidad;
- 20 comprendiendo la entidad molecular una proteína que tiene un marcador de afinidad peptídico fijado a un componente de la capa proteica;
- comprendiendo la entidad molecular una proteína y tanto un componente de la capa proteica y la entidad molecular tienen marcadores de afinidad respectivos fijados; o
- estando las entidades moleculares condensadas genéticamente dentro de un componente de la capa proteica.
- 25

Fig. 1

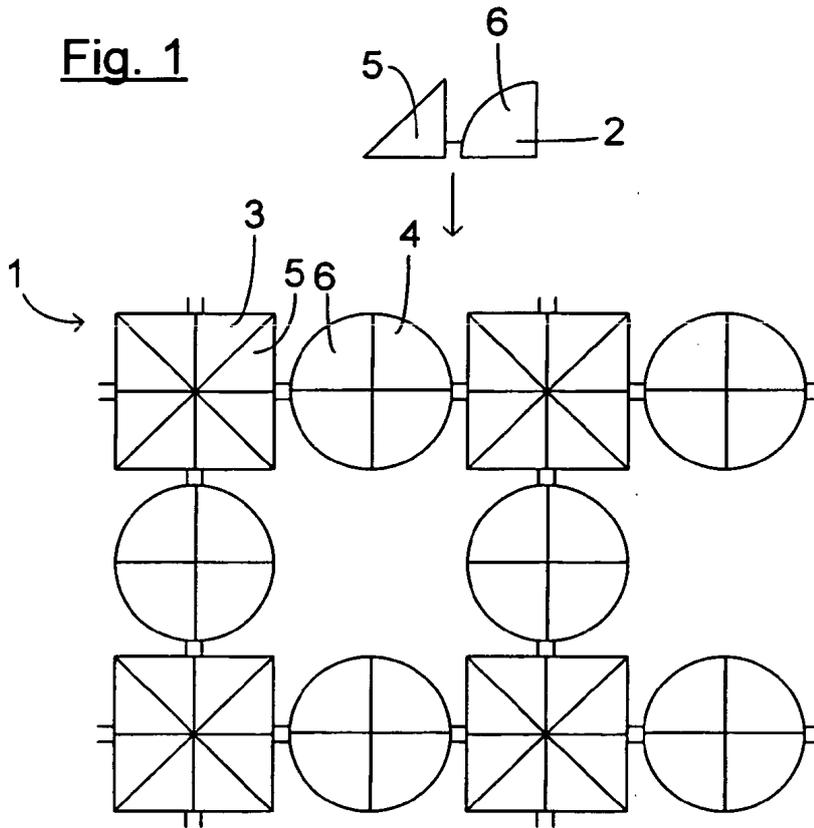


Fig. 2

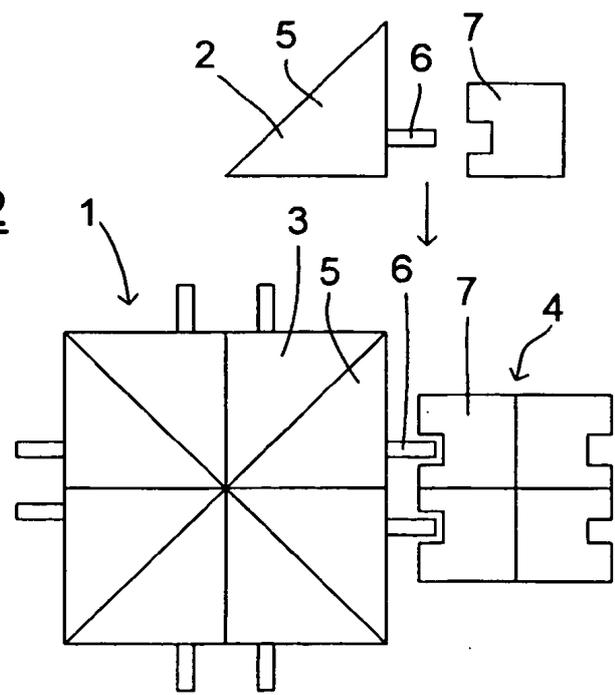


Fig. 3

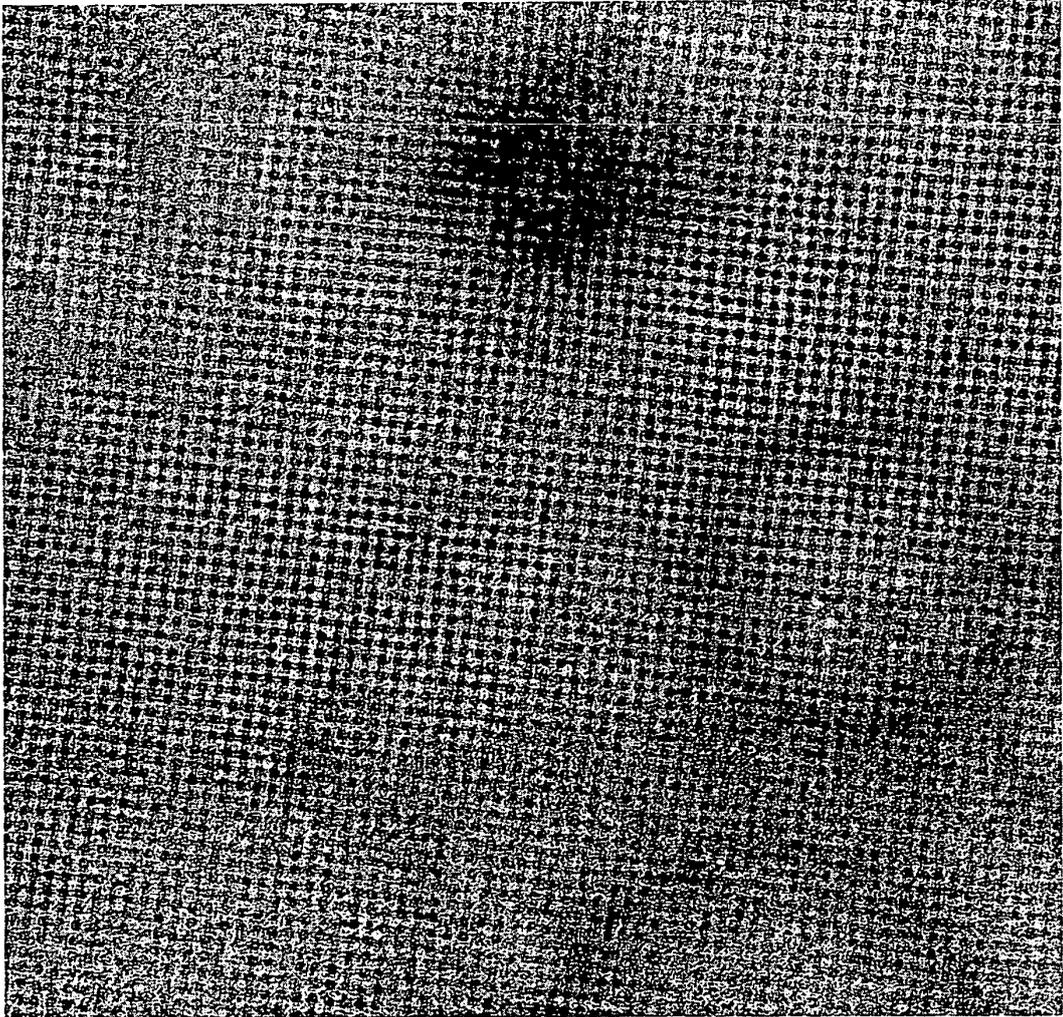


Fig. 4

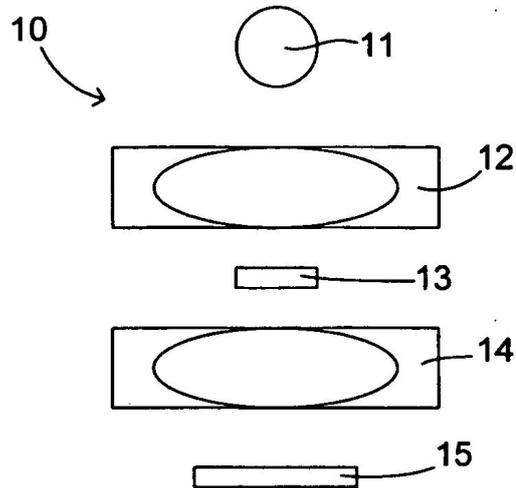


Fig. 5

