

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 811 698**

51 Int. Cl.:

**G16B 15/20** (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.11.2013 PCT/EP2013/074556**

87 Fecha y número de publicación internacional: **30.05.2014 WO14080005**

96 Fecha de presentación y número de la solicitud europea: **25.11.2013 E 13796039 (9)**

97 Fecha y número de publicación de la concesión europea: **27.05.2020 EP 2923291**

54 Título: **Determinación de novedosas funcionalidades enzimáticas utilizando nubes de puntos tridimensionales que representan propiedades fisicoquímicas de cavidades proteínicas**

30 Prioridad:

**26.11.2012 EP 12194206**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**15.03.2021**

73 Titular/es:

**INNOPHORE GMBH (100.0%)  
Am Eisernen Tor 3  
8010 Graz, AT**

72 Inventor/es:

**GRUBER, KARL;  
STEINKELLNER, GEORG y  
GRUBER, CHRISTIAN**

74 Agente/Representante:

**CARVAJAL Y URQUIJO, Isabel**

**ES 2 811 698 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

Determinación de novedosas funcionalidades enzimáticas utilizando nubes de puntos tridimensionales que representan propiedades fisicoquímicas de cavidades proteínicas

5 La invención se refiere a un método de selección para el catalizador de una reacción química predeterminada como se define en las reivindicaciones.

## Antecedentes

10 Las enzimas se utilizan cada vez más en procesos biotecnológicos y biocatalíticos. Sin embargo, a pesar de numerosas aplicaciones exitosas, es claro que no todas las reacciones químicas potencialmente interesantes están representadas en la cartera actual de enzimas. Además, las enzimas naturales generalmente no están completamente optimizadas para su utilización en procesos industriales. Por lo tanto, la identificación y/o diseño de nuevas funcionalidades enzimáticas y la mejora de las enzimas existentes son de suma importancia.

15 Se han descrito en la literatura varias metodologías para identificar sitios de unión a proteínas. Por ejemplo, se ha utilizado una infraestructura de computación en la nube basada en GPU para realizar de manera eficiente una comparación estructural de sitios de unión a proteínas (Leinweber et al., (2012) IEEE International Conference on Digital Ecosystems and Technologies, art. no. 6227926).

Se ha descrito que un modelo de nube de puntos etiquetado es adecuado para modelar biomoléculas como proteínas y sitios de unión a proteínas, donde una etiqueta puede representar un tipo de átomo o una propiedad fisicoquímica (Fober et al., (2011) IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8(6), art. no. 5722954, pp. 1653-1666).

20 Un método para la comparación de la estructura de la proteína en el cual la información sobre la geometría y las propiedades fisicoquímicas de tales estructuras están representada en forma de nubes de puntos marcadas, ha sido descrita por Fober et al. ((2009) ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications, art. no. 5364137, pp. 1251-1256).

25 El documento WO0023474 describe un método de manipulación de proteínas en donde se crea una base de datos de ordenador que permite búsquedas que comprenden entradas en forma de descripciones de una ubicación y orientación en el espacio 3D de las cadenas laterales de los residuos de aminoácidos constituyentes de una proteína como marco de referencia para identificar un acierto que corresponde a una proteína marco que tiene similitud estructural con una proteína de muestra.

## Breve descripción de la invención

30 La invención se refiere a los esfuerzos de manipulación y diseño de proteínas basados en una integración de estudios estructurales y mecanicistas y metodologías de evolución dirigida para lograr una estrategia óptima y eficiente que determine nuevas funcionalidades enzimáticas utilizando nubes de puntos tridimensionales que representan las propiedades fisicoquímicas de las cavidades de proteínas.

35 Además de su actividad catalítica original, se ha demostrado que algunas enzimas también catalizan reacciones químicas completamente diferentes como actividad secundaria. Esta llamada "promiscuidad catalítica" puede ser el punto de partida para diseñar una enzima competente que catalice una reacción química no natural. La promiscuidad catalítica en las enzimas es la capacidad de las enzimas para catalizar transformaciones químicas claramente diferentes. La transformación química puede diferir en el grupo funcional involucrado, es decir, el tipo de enlace formado o escindido durante la reacción y/o puede diferir en el mecanismo catalítico o la ruta de formación y ruptura del enlace. Hasta ahora, sin embargo, la identificación de la promiscuidad catalítica se basó en la casualidad. La selección de bases de datos estructurales con motivos funcionales "catalóforos", que se basan en propiedades fisicoquímicas, representa una metodología más racional para predecir la "promiscuidad catalítica" en las enzimas. Usando este método de última generación, recientemente los presentes inventores han tenido éxito en la identificación de dos enzimas que según se predice exhiben actividades interesantes de enoato reductasa pero que no muestran ninguna similitud (en la secuencia y en el nivel estructural) a enoato reductasas conocidas (las antiguas enzimas amarillas). En este caso, el diseño de los motivos funcionales utilizados para la selección de la base de datos se basó en una gran cantidad de información estructural y mecanicista disponible para esta clase de enzimas. Sin embargo, esta metodología también podría involucrar un motivo estructural teórico del sitio activo derivado de los requisitos mecanicistas de una reacción química particular.

50 Además de explotar la promiscuidad catalítica en las enzimas naturales existentes, se pueden crear nuevas funcionalidades enzimáticas a través del diseño de novo de sitios activos injertando residuos catalíticamente importantes en andamios de proteínas estables. Obviamente, esta metodología también requiere un conocimiento mecanicista detallado y puede beneficiarse enormemente de la gran cantidad de datos sobre catalizadores orgánicos y órgano metálicos no enzimáticos utilizados en la química de soluciones. Los ejemplos en la literatura indican que se ha logrado la prueba de principio para tal metodología. Hasta ahora, sin embargo, no se ha utilizado para diseñar un nuevo catalizador para biocatálisis industrial.

Muy frecuentemente, la actividad catalítica de una enzima se logra mediante un pequeño dominio de una proteína más grande que podría incluir otros dominios necesarios para la función fisiológica pero irrelevantes para la aplicación biotecnológica. La reducción del tamaño del catalizador -es decir, el diseño de una "enzima mínima"- tiene potencialmente enormes beneficios para la expresión de la proteína, la estabilidad del proceso o para la manipulación por evolución dirigida.

Descripción detallada de la invención

La presente invención se refiere a un método como se define en las reivindicaciones.

Por lo tanto, se desarrolló un método para encontrar catalóforos con geometrías específicas del sitio activo mediante el uso de una plantilla de distancia tridimensional que imita una constelación catalítica mínima del sitio activo.

10 En el contexto de esta invención, el término "catalóforo" se refiere a la descripción del sitio activo de una proteína, preferiblemente de una enzima. Un catalóforo también abarca una enzima mínima o parte o fragmento de una enzima que comprende el sitio activo de dicha enzima, incluidos los descriptores 3D del sitio activo.

De acuerdo con la invención, el sitio activo puede ser, pero no está limitado a, un sitio enzimático, catalítico, de unión al receptor o de expresión de proteínas.

15 Específicamente, en caso de que el catalóforo identificado por el método de la invención, sea un dominio catalóforo aislado, dicho dominio catalóforo tiene preferiblemente las mismas o similares características/actividad que el dominio catalóforo presente en la proteína original. Alternativamente, dicho dominio de catalóforo puede modificarse para tener una actividad catalítica y enzimática aumentada en comparación con el dominio original.

Aquí se describe un método para determinar un catalóforo de una proteína que incluye los pasos de

20 a) crear una base de datos de nube de puntos para estructuras de proteínas diana;  
b) crear una nube de puntos de consulta;  
c) buscar en dicha base de datos con dicha consulta para identificar de este modo uno o más catalóforos.

También se divulga en el presente documento el método descrito anteriormente, en donde se calcula una nube de puntos para cada propiedad fisicoquímica de las proteínas diana.

25 También se divulga en el presente documento el método descrito anteriormente, en donde se calculan nubes de puntos para interacciones adicionales como enlaces H para estructuras de proteínas diana.

También se divulga en el presente documento el método descrito anteriormente, en donde la nube de puntos de consulta es un conjunto de nubes de puntos.

30 También se divulga en el presente documento el método descrito anteriormente, en donde dicho conjunto de nubes de puntos se deriva de la actividad catalítica de proteínas diana o de los modelos de mecánica cuántica de reacciones de plantilla.

También se divulga en el presente documento el método descrito anteriormente, en donde dichas nubes de puntos de consulta y dicha base de datos de nubes de puntos se superponen.

35 También se divulga en el presente documento el método descrito anteriormente, en donde dicha nube de puntos de consulta se hace rotar y se traslada aleatoriamente.

También se divulga en el presente documento el método descrito anteriormente, en donde dichas estructuras de proteínas diana se homogeneizan.

También se divulga en el presente documento una enzima que comprende un catalóforo, en donde dicho catalóforo se obtiene mediante un método como el descrito anteriormente.

40 También se divulga en el presente documento un método para construir un conjunto de catalóforos con propiedades catalíticas y fisicoquímicas predefinidas, en donde dicho catalóforo se obtiene mediante un método como el descrito anteriormente.

También se divulga en el presente documento un conjunto de catalóforos con propiedades predefinidas obtenidas mediante un método como el descrito anteriormente.

45 También se divulga aquí el uso de un catalóforo como catalizador para biocatálisis industrial.

Los posibles resultados en bases de datos estructurales se probaron experimentalmente para determinar la actividad deseada y mejorar las propiedades para la aplicación industrial. Como ejemplo, la plantilla de búsqueda se derivó del sitio activo de antiguas enzimas amarillas (OYE). Se ha demostrado que los miembros de la conocida familia de enoato reductasa poseen un alto potencial como biocatalizadores en la reducción asimétrica de alquenos activados. Las proteínas de la familia OYE catalizan la reducción de los enlaces C=C activados en compuestos  $\alpha,\beta$ -

insaturados, y se ha demostrado que representan una alternativa biocatalítica eficiente a la síntesis orgánica estándar. Además, en el cofactor de mononucleótido de flavina (FMN), el sitio catalítico en sí mismo generalmente comprende un par de residuos (típicamente histidina/histidina o asparagina/histidina) que actúan como donantes de enlaces de H al grupo del sustrato que extrae electrones y un residuo de tirosina conservado, que es necesario para suministrar un protón al átomo C<sub>o</sub> durante el recambio.

Es un objeto de la invención detectar otras enzimas con funcionalidad enoato reductasa en bases de datos estructurales incluso dentro de esta familia intensamente investigada. Por lo tanto, se utilizó una imagen de sitio activo tridimensional relativamente simple, generada mediante la comparación de las características del sitio activo de las estructuras OYE representativas para buscar otras enzimas con esta constelación específica del sitio activo.

Como plataforma de búsqueda se utilizó la base de datos Relibase+ disponible comercialmente (v. 2.2), que contiene información estructural preprocesada. Además, se desarrolló una base de datos interna (CATALObase) de estructuras preprocesadas que se prepararon aplicando funciones de estructura YASARA. Para la búsqueda en Relibase+, la imagen del sitio activo se creó utilizando el constructor Relibase+ 2D/3D. La imagen del sitio activo en 3D para la búsqueda usando la base de datos interna se generó usando una secuencia de comandos YASARA. Con ambas metodologías de búsqueda se identificaron estructuras OYE típicas, así como dos proteínas "novedosas" con antiguas actividades enzimáticas amarillas predichas. También se utilizó la base de datos de bolsillo CavBase, que está incrustada en la base de datos Relibase+ y que contiene información sobre cavidades de sitios activos y pseudocentros abstractos. En el caso de las OYE y utilizando la configuración como se describió anteriormente, no se pudieron identificar las proteínas que se encontraron con la metodología geométrica pura dentro de los primeros 200 aciertos CavBase.

La búsqueda final de catalóforo resultó en una cantidad manejable de 46 resultados que podrían reducirse aún más mediante la aplicación de un filtro para eliminar los resultados obvios de OYE conocidas. Esto se realizó por secuencia (usando BLAST <http://blast.ncbi.nlm.nih.gov/>) y la comparación estructural usando PDBefold, así como mediante la clasificación de patrones de secuencia OYE, dejando solo las enzimas novedosas con una disposición de sitio activo similar. Por lo tanto, los resultados reducidos revelaron varias estructuras que consistían en los requisitos de constelación de sitio activo de la plantilla a 6 aciertos.

Mediante inspección visual se identificaron dos oxidorreductasas (Código PDB: 2i0k y 3fbs) donde el residuo de tirosina tiene la distancia requerida pero está en el lado opuesto del cofactor de flavina y la cavidad del sitio activo, por lo que no puede proporcionar el protón para un sustrato potencial. La subunidad de flavoproteína también identificada de una fumarato reductasa mutada (N204Y) se descartó en la primera ronda de investigación, debido al heterocomponente hemo que puede haber alterado los propios análisis de enzimas amarillas antiguas. La monooxigenasa putativa (Código PDB: 1usc y 1usf) y la proteína PH0856 no caracterizada (Código PDB: 2r6v) se seleccionaron para su investigación y caracterización. Las enzimas encontradas no son secuenciales ni, y este es un resultado más sorprendente, están relacionadas estructuralmente con cualquier estructura OYE conocida. La constelación del sitio activo mostró una simetría de imagen especular aproximada a las estructuras de la plantilla, lo que sugiere un resultado estereoquímico diferente. La monooxigenasa putativa de *Thermus thermophilus* y la proteína PH0856 no caracterizada de *Pyrococcus horikoshii* se encuentran en el siguiente texto referido a Tth y Pho, respectivamente, con base en los organismos donde fueron aislados. La actividad biocatalítica de estas enzimas ha sido evaluada experimentalmente. En conclusión, se han identificado con éxito dos enzimas con una función aún desconocida de las cuales se encontró que tienen una arquitectura de sitio activo similar en comparación con las enoato reductasas clásicas y no muestran ninguna similitud ni en secuencia ni a nivel estructural global con ninguna OYE conocida.

Como una realización adicional, dicho dominio puede introducirse en andamios de proteínas, en donde la estructura secundaria/terciaria del catalóforo se conserva para mantener su funcionalidad. Como una realización alternativa, el catalóforo puede modificarse mediante técnicas conocidas por la persona experta, por ejemplo mediante mutagénesis aleatoria para mejorar o cambiar su funcionalidad, específicamente en vista de la actividad enzimática, de unión a la diana o catalítica.

Breve descripción de las figuras

La Figura 1 representa la preparación de nubes de puntos de proteínas.

Figura 2 representa el flujo de trabajo del proceso de coincidencia de nube de puntos.

Figura 3 representa el proceso de coincidencia.

## Ejemplos

### Ejemplo 1 - Generación de nubes de puntos

Esta metodología para identificar "catalóforos" de motivos catalíticos activos se basa en la implementación de nuevos protocolos que reutilizan mapas de cuadrícula precalculados de estructuras de proteínas diana generadas con una versión modificada de AutoGrid4 que forma parte del paquete de software AutoDock4 (AutoDock, <http://http://autodock.scripps.edu>). Se calcula una nube de puntos para cada tipo de propiedad fisicoquímica, así

como nubes de puntos de propósito especial que representan interacciones adicionales como enlaces H (Tabla 1 y Figura 1).

Tabla 1: Propiedades fisicoquímicas

Entrada	Tipo	Ejemplos	
1	energías potenciales	átomos sin enlaces H	azufre, cloro, calcio, manganeso, hierro, zinc, bromo, yodo, flúor, magnesio, fósforo, carbono alifático, carbono aromático, nitrógeno, hidrógeno
		aceptadores de enlaces H	nitrógeno, azufre, oxígeno
		donantes de enlaces H	hidrógeno
2	potencial electrostático		
3	hidrofobicidad		
4	accesibilidad		

5 Estos mapas de cuadrícula diana se convierten en archivos de datos de nube de puntos para ser procesados dentro de la tubería del software de catalóforo. Las estructuras de proteínas se preparan siguiendo protocolos estándar para la adición de átomos de hidrógeno, la adición de átomos y residuos faltantes, seguido de la detección automática de cavidades (Figura 2). Las áreas identificadas como cavidades proteicas adecuadas para la unión de sustratos potenciales se caracterizan posteriormente y se almacenan como nubes de puntos de propiedad.

10 Se requiere un conjunto de nubes de puntos para describir la "consulta de búsqueda". Estas nubes de puntos pueden derivarse de enzimas que presentan la actividad catalítica deseada o de modelos de mecánica cuántica de reacciones de plantilla. Estas nubes de puntos se comparan con las nubes de puntos que se calcularon previamente para todas las estructuras de proteínas disponibles (Figura 2).

**Ejemplo 2 - Establecimiento de una base de datos de catalóforo**

15 Para homogeneizar las estructuras de proteínas diana y crear nubes de puntos para cada cavidad identificada, se recopilan y almacenan estructuras de diversos "reservorios" en un formato estandarizado en un sistema de base de datos relacional CATALObase (Tabla 2).

Tabla 2: Fuentes de datos

Entrada	Reservorio	Intervalo de actualización
1	RCSB PDB	diario
2	PDBe PISA	diario
3	Modelos de homología	continuamente
4	Estructuras propias	de acuerdo con lo solicitado
5	Mutantes y variantes	de acuerdo con lo solicitado

20 Para cada una de estas estructuras de proteínas, las nubes de puntos se generan y almacenan automáticamente para el siguiente proceso de coincidencia.

**Ejemplo 3 - Coincidencia de catalóforos de consulta con entradas de base de datos**

25 La consulta y las nubes de puntos de la base de datos se superponen y la desviación de los dos mapas de cuadrícula se evalúa comparando los puntos de cuadrícula más cercanos entre la diana y la plantilla. Para cada tipo de nube de puntos se genera una "puntuación de identidad" que representa la calidad de la coincidencia. Este puntaje individual (Ecuación 1) está ponderado por un factor seleccionable para permitir el ajuste de la influencia de los diferentes mapas de cuadrícula (por ejemplo, establecer el factor de puntaje para las nubes de puntos de enlace H en cero hace que esta interacción se ignore). Al evaluar la calidad de la coincidencia, la nube de puntos de consulta se hace rotar y se traslada aleatoriamente. Si el nuevo complejo obtenido es mejor que el anterior, se acepta este paso. El procedimiento se repite hasta que no se pueda obtener una mejor coincidencia (Figura 3). Los

30

algoritmos se implementaron en una versión independiente del software CATALOphore en C++ basado en el código fuente disponible libreme de AutoDock4 y PCL (The Point Cloud Library (PCL), <http://pointclouds.org>).

**Ecuación 1: Puntuación de identidad CATALOphore**

$$\begin{aligned}
 \text{puntuación} &= \sum_i \text{puntuación}_{\text{punto}} \\
 \text{puntuación}_{\text{punto}} &= \sum_{\text{tipo}=\text{A,C,d,e,HD,N,NA,OA,SA},\dots} \text{puntuación}_{\text{tipo}} \\
 \text{puntuación}_{\text{punto}}^{\text{tipo}} &= \left| E_{\text{consulta}} - E_{\text{diana}} \right|
 \end{aligned}$$

**Ejemplo 4 –**

- 5 Para probar el procedimiento descrito anteriormente, se extrajeron nubes de puntos para enzimas que pertenecen al grupo de las antiguas enzimas amarillas. Estas nubes de puntos de consulta se compararon con un subconjunto de prueba de CATALObase que contiene nubes de puntos que representan 10 estructuras proteicas diversas (enzimas amarillas antiguas, alcohol deshidrogenasas y lipasas). La Tabla 2 resume los resultados del proceso de coincidencia. La mejor coincidencia se obtuvo para la estructura de la proteína utilizada para el diseño de la consulta, como se esperaba, Clasificados mediante el presente "puntaje de identidad", los siguientes dos aciertos son enzimas que previamente han demostrado tener una actividad similar a la antigua enzima amarilla pero que no fueron anotadas ni caracterizadas bioquímicamente. Un control negativo y positivo se clasificó como se esperaba.
- 10

Tabla 3: Los 5 mejores resultados del ejemplo de antigua enzima amarilla.

Entrada	Proteína	Código PDB	Puntuación de identidad	Comentarios
1	Antigua enzima amarilla	1OYA	97%	utilizada para la creación de la nube de puntos de consulta
2	Antigua enzima amarilla	2GQ9	84%	estructuralmente similar a 1OYA, control positivo
3	Estireno monooxigenasa putativa	1USC	83%	actividad OYE observada experimentalmente
4	Proteína no caracterizada PH0856	2R6V	81%	actividad OYE observada experimentalmente
5	Lipasa	1TCA	14%	control negativo

**REIVINDICACIONES**

1. Un método para seleccionar un catalizador de una reacción química predeterminada, que comprende los pasos de:
- 5 a) determinar las propiedades fisicoquímicas en las ubicaciones de los vóxeles de la cuadrícula, en donde la cuadrícula incluye la cavidad del sitio activo de una primera enzima, para obtener nubes de puntos de consulta para cada tipo de propiedad fisicoquímica,
- en donde las propiedades fisicoquímicas son
- energías potenciales de átomos sin enlaces H, seleccionados entre azufre, cloro, calcio, manganeso, hierro, zinc, bromo, yodo, flúor, magnesio, fósforo, carbono alifático, carbono aromático, nitrógeno e hidrógeno,
- 10 energías potenciales de aceptores de enlaces H, seleccionados entre nitrógeno, azufre y oxígeno,
- energías potenciales de donantes de enlaces H, como hidrógeno, potencial electrostático,
- hidrofobicidad, y
- accesibilidad,
- 15 b) crear una base de datos de nube de puntos de nubes de puntos correspondientes para cavidades de sitios activos de proteínas predeterminadas;
- c) buscar en la base de datos con las nubes de puntos de consulta para identificar de ese modo una de las proteínas predeterminadas que tiene una cavidad del sitio activo que coincide con la cavidad del sitio activo de la primera enzima, identificando así una segunda enzima como catalizador, comparando las nubes de puntos de consulta con las nubes de puntos correspondientes de la base de datos de nubes de puntos,
- 20 en donde
- el paso d) comprende superponer las nubes de puntos de consulta y las nubes de puntos correspondientes de la base de datos de nubes de puntos,
- la comparación comprende comparar las propiedades fisicoquímicas en los vóxeles de cuadrícula de las nubes de puntos de consulta con las propiedades fisicoquímicas en los vóxeles de cuadrícula más cercanos de las nubes de puntos correspondientes de la base de datos de nubes de puntos y generar una puntuación de identidad individual que representa la calidad de la coincidencia para cada propiedad fisicoquímica y generar un puntaje de identidad como una suma ponderada de los puntajes de identidad individuales.
- 25
2. El método de acuerdo con la reivindicación 1, en donde el paso d) comprende hacer rotar y trasladar la una o más nubes de puntos de consulta aleatoriamente.
- 30

Fig. 1

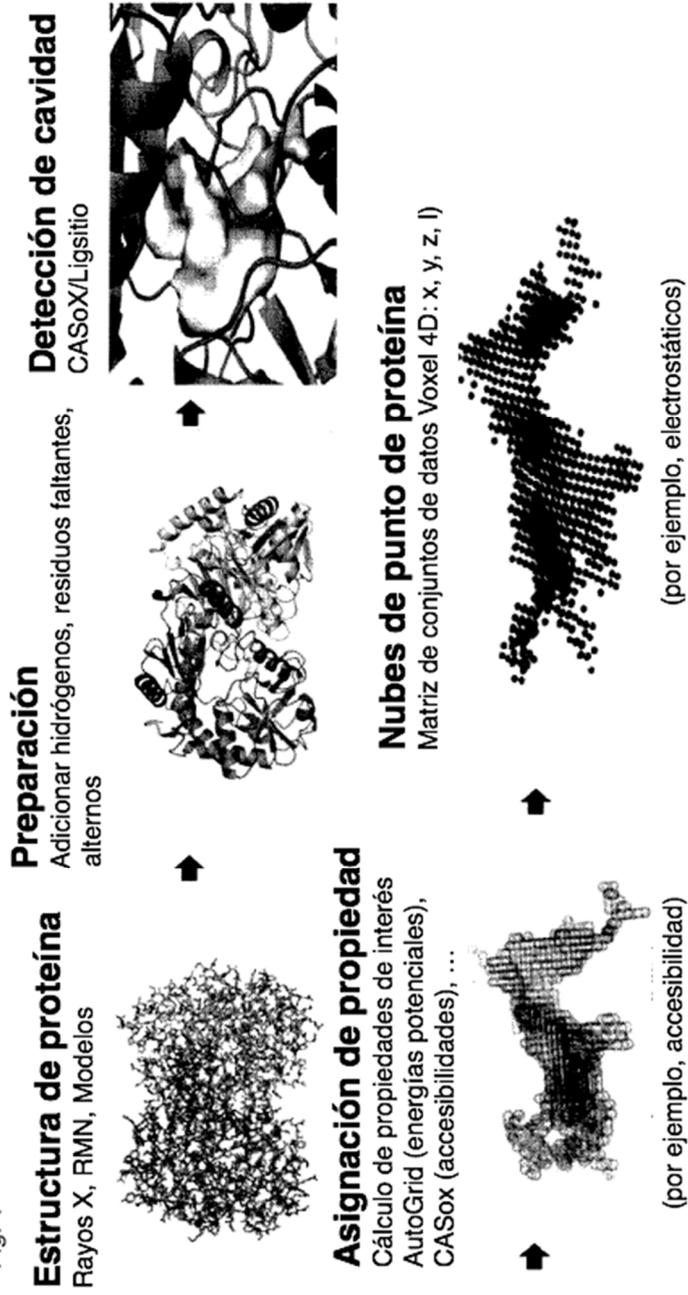


Fig. 2

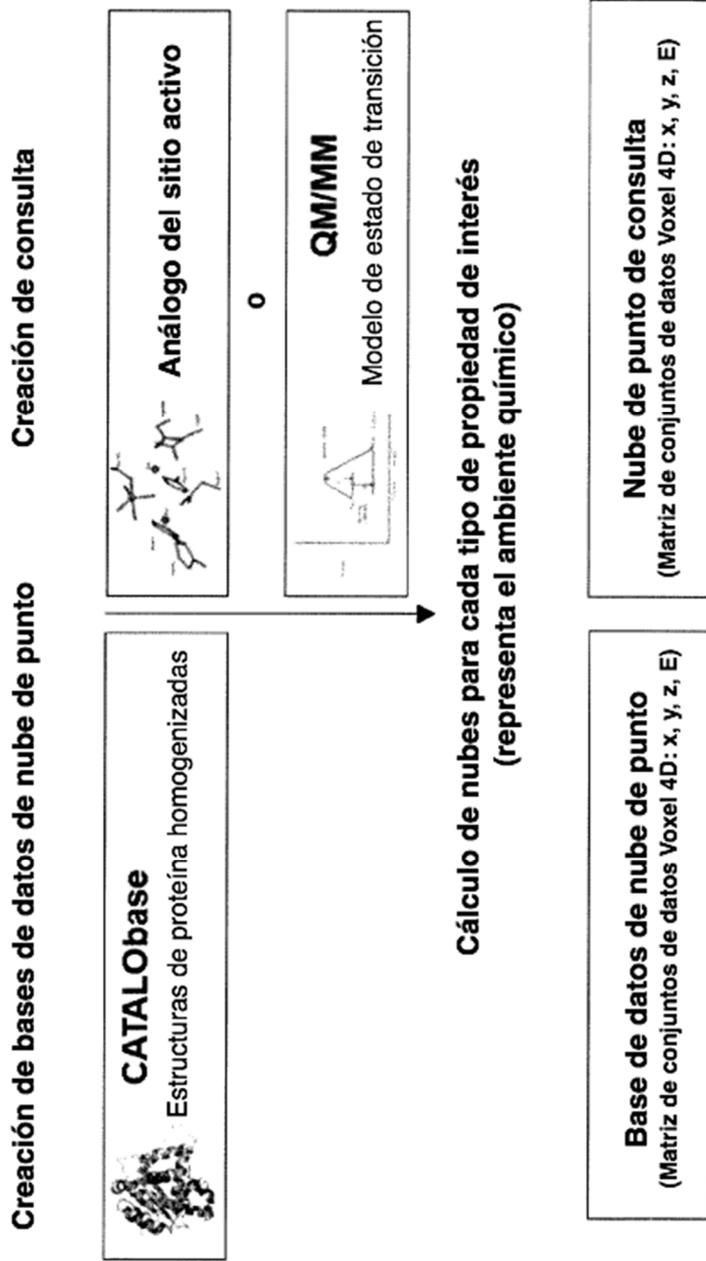


Fig. 3

