

Métodos de simulación QM/MM y de mecánica estadística: una aproximación teórica a la biología molecular

Albert Ardèvol

Resumen: Los avances técnicos y metodológicos en el campo de la química teórica y la enzimología computacional han ampliado considerablemente el espectro de aplicaciones y la precisión en las predicciones de los métodos *in silico*. En este sentido, las técnicas híbridas de mecánica cuántica/mecánica molecular (QM/MM) juntamente con los métodos de mecánica estadística han contribuido decisivamente al estudio de diversos mecanismos de catálisis enzimática y de reconocimiento enzima-ligando. En este artículo se revisa el desarrollo de dichas técnicas y su impacto en la biología molecular.

Palabras clave: Enzimología computacional, dinámica molecular, QM/MM, química teórica, mecanismo enzimático.

Abstract: Technical and methodological advances in the field of theoretical chemistry and computational enzymology have greatly expanded the range of applications and predictive accuracy of *in silico* methods. In this sense, hybrid techniques of quantum mechanics/molecular mechanics (QM/MM) together with statistical mechanics methods have contributed decisively to the study of enzymatic reaction mechanisms and enzyme-ligand recognition. This article reviews the development of these techniques and their impact on molecular biology.

Keywords: Computational enzymology, molecular dynamics, QM/MM, theoretical chemistry, enzyme mechanism.

Introducción

La aproximación a la enzimología y la biología molecular se hace desde múltiples técnicas experimentales y teóricas, con el objetivo de entender el mecanismo de acción de las proteínas y sus propiedades intrínsecas. A pesar de ello, todavía hay cuestiones fundamentales que permanecen elusivas y para las que no hay respuestas: ¿Cómo reconocen el sustrato y catalizan la reacción? ¿Por qué se pliegan de una forma determinada? y ¿Cómo interactúan entre ellas? son preguntas recurrentes en el campo de la biología molecular.

La primera hipótesis propuesta presume una relación directa entre la estructura y la acción de la enzima. Dentro de este marco, Emil Fisher formuló en 1894 el modelo de “llave-cerradura” según el cual el sustrato y la enzima son estructuras estáticas complementarias, y la enzima dispone los residuos catalíticos de forma conveniente para que se dé la reacción. Posteriormente, basándose en el complejo activado enzima-sustrato de Michaelis y Menten, Koshland postuló en 1958 el modelo de “encaje inducido” según el cual la unión enzima-sustrato provoca un cambio conformacional en el

centro activo que activa la reacción. Por lo tanto, la enzima y el sustrato no son conformaciones estáticas ni complementarias sino que tienen un cierto grado de flexibilidad.

Sin embargo el modelo de encaje inducido también tuvo que ser revisado con la constatación de que en algunas proteínas, los cambios conformacionales no son necesariamente provocados por la unión del sustrato (o el factor alostérico) sino que varios microestados de la proteína coexisten incluso en la forma libre (*i.e.*; sin ligando). En el nuevo modelo (Figura 1), la unión del ligando reconoce y estabiliza más una de las formas (la reactiva o activada), por lo que se denomina modelo de “selección conformacional”. El debate entre el modelo de encaje inducido y de selección conformacional es intenso actualmente,^{1,2} y no hay una teoría unificada en este sentido.

La inclusión de la flexibilidad conformacional en el modelo enzimático no se limita al alosterismo sino que se extiende al plegamiento de proteínas y a las proteínas intrínsecamente desordenadas (IDP). Por consiguiente, la relación estructura-actividad dejó de ser suficiente para explicar todas las propiedades enzimáticas, para dar paso a la relación dinámica-actividad. En este sentido, se desarrollaron técnicas experimentales y teóricas para estudiar el comportamiento dinámico de las proteínas, como las técnicas de NMR, de smFRET (*single molecule Förster resonance energy transfer*) y dinámica molecular o de cálculo de energías libres entre otras, de forma que el esfuerzo para estudiar la relación dinámica-actividad en proteínas es hoy en día un fenómeno transversal y multidisciplinar.

Enzimología computacional

La enzimología computacional integra el desarrollo y aplicación de los métodos teóricos para estudiar sistemas enzimáticos y su evolución en los últimos años ha permitido estudiar sistemas de gran complejidad. Según el nivel de teoría empleado, se pueden estudiar desde fenómenos de plega-



Albert Ardèvol

Department of Chemistry and Applied Biosciences,
ETH Zurich.
USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano,
Switzerland.
C-e: albert.ardevol@phys.chem.ethz.ch

Recibido: 18/07/2013. Aceptado: 11/11/2013.

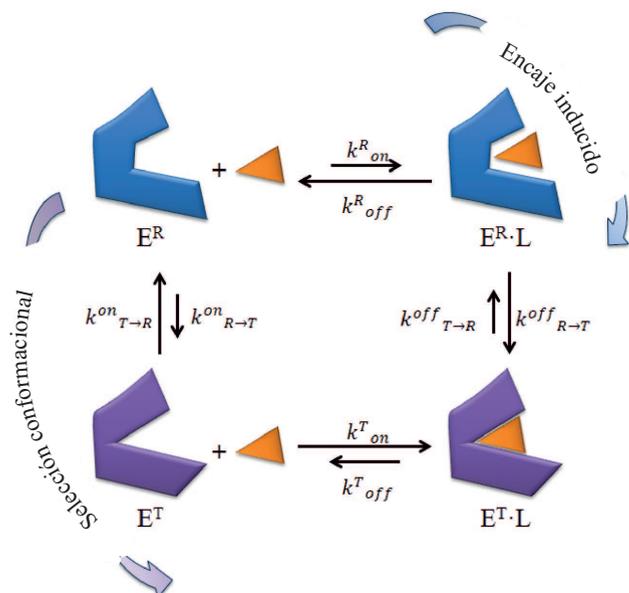


Figura 1. Modelo de encaje inducido y de selección conformacional para la formación del estado excitado $E^T \cdot L$ (el complejo de Michaelis). En el modelo de selección conformacional, la estructura tensionada de la proteína (E^T) coexiste en equilibrio con la forma relajada (E^R), y E^T une con mayor afinidad el sustrato, lo que desplaza el equilibrio hacia el complejo $E^T \cdot L$. En el modelo de encaje inducido es la unión del sustrato lo que provoca el cambio conformacional a la forma tensionada.

miento de proteínas, fenómenos de alosterismo, interacciones enzima-ligando, etc (usando métodos clásicos o de *coarse grain*) hasta reacciones enzimáticas o de transferencia de carga (usando métodos basados en la mecánica cuántica).³⁻⁵

Esta rápida evolución ha extendido la aplicación de dichos métodos teóricos desde cuestiones fundamentales a la solución de problemas técnicos reales, como el diseño de principios activos de aplicación farmacéutica, de enzimas con una actividad o especificidad deseada o de biomateriales con unas propiedades particulares. Hacer una descripción de todos los métodos de enzimología computacional queda fuera de las pretensiones de este artículo, el cual se quiere centrar en los métodos de dinámica molecular híbridos de mecánica cuántica/mecánica molecular (QM/MM por sus siglas en inglés).

Dinámica molecular

En la dinámica molecular, se simula la evolución temporal del sistema mediante la integración de las ecuaciones Newtonianas del movimiento. El nivel de teoría empleado es determinado por la función de energía potencial que rige el sistema, y se elige dependiendo del fenómeno que se quiera estudiar. Los niveles de teoría se clasifican en: métodos de mecánica cuántica (QM), de mecánica clásica o molecular atómica (MM) o de *coarse grain* (CG). En los métodos QM, el sistema se define a nivel electrónico, y la energía potencial se calculan mediante la resolución (aproximada) de la ecuación de Schrödinger o de las ecuaciones de Kohn-Sham; en los métodos de MM el sistema se define mediante átomos indivisibles mientras que en los métodos de CG, el sistema está formado por grupos de átomos. Tanto en los

métodos MM como en los CG, la función de energía potencial se calcula a partir de un campo de fuerzas que debe ser previamente parametrizado empíricamente o mediante cálculos realizados a un nivel teórico superior. Otros métodos incluyen el *empirical valence bond* (EVB) o los métodos multi-escala.⁵ La precisión de los resultados (la estructura electrónica/atómica, las energías de interacción, los estados de transición,...) así como el nivel de información sobre el sistema depende del nivel de teoría empleado. Pero al mismo tiempo, el coste computacional para resolver la función de energía potencial depende del nivel de teoría, de manera que, el tamaño del sistema y el número de pasos de integración que se puede calcular con un método QM es mucho más pequeño que el que se puede calcular con un método MM o CG.

Para simular una reacción química, se requiere un método de primeros principios, capaz de describir cambios importantes en la estructura electrónica del sistema, como la rotura y la formación de enlaces covalentes. Sin embargo, si la reacción ocurre en el seno de una enzima, no es posible simular todo el entorno, incluyendo la proteína, el disolvente y cofactores, mediante métodos QM. Por lo tanto, para simular reacciones químicas en el seno de una enzima, se debe usar un método multi-escala que combine una descripción cuántica del centro activo de la enzima con una descripción clásica del resto del sistema. La ventaja del método híbrido QM/MM es clara: se puede simular la ordenación electrónica que se da en el centro activo de la enzima cuando ocurre la reacción, teniendo en cuenta explícitamente el efecto del entorno enzimático y del disolvente, así como las variaciones temporales de dicho entorno, que naturalmente se adaptan a las variaciones electrónicas provocadas por la reacción enzimática.

Las simulaciones de dinámica molecular QM (o QM/MM) se pueden realizar mediante dinámica molecular de Born-Oppenheimer (BOMD) o dinámica molecular de Car-Parrinello (CPMD). La diferencia radica en que mientras en la BOMD se calcula la función de onda a cada paso de integración, en la CPMD sólo se calcula al inicio de la dinámica. En el paso de integración de Car-Parrinello⁶ los grados de libertad electrónicos se incluyen en el Lagrangiano como partículas ficticias que evolucionan simultáneamente con los núcleos. Esto representa una gran ventaja respecto la BOMD, ya que el cálculo de la densidad electrónica es la operación más costosa en recursos computacionales. Sin embargo, el tiempo de paso debe ser más pequeño que en la BOMD (aproximadamente 0.1 fs). Ambos métodos se han usado satisfactoriamente en aplicaciones biológicas.^{7,8}

QM/MM

El método QM/MM se caracteriza por la partición del sistema en i) una "región QM" que corresponde a la parte químicamente relevante (el centro activo) y que se trata mediante un método basado en la mecánica cuántica y ii) una "región MM" que incluye el resto del sistema y que se describe mediante un método de mecánica molecular (Figura 2). El cálculo QM/MM se define por los métodos QM y MM empleados, por cómo se combinan los Hamiltonianos (según si es por acoplamiento mecánico o electrónico) y por la forma de tratar la interfase entre las regiones QM y MM.

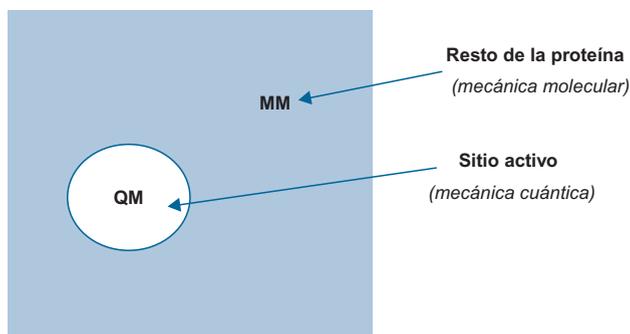


Figura 2. Partición del sistema en el método híbrido QM/MM. El sitio activo de la proteína se describe mediante mecánica cuántica, mientras que el entorno proteico y el solvente se describe con un campo de fuerzas.

El método QM

La elección del método QM depende de la precisión que se desea y el coste computacional del método. Los métodos QM más usados para el cálculo de reacciones enzimáticas son los basados en la teoría del funcional de la densidad (DFT), los semiempíricos o los post-Hartree Fock (post-HF).

- Los métodos semiempíricos (basados en la aproximación NDDO, *neglect the diatomic differential overlap*) son los menos costosos computacionalmente, lo que permite hacer simulaciones más largas y/o con una región QM mayor. El principal inconveniente es que son muy dependientes de los parámetros que los definen. Los métodos semiempíricos más usados incluyen el MNDO, AM1, PM3 y SCC-DFTB. Actualmente se usan en combinación con métodos más precisos como un método rápido para explorar la reacción para luego calcular la energía potencial con un método DFT o post-Hartree Fock.⁹
- Los métodos basados en DFT requieren más recursos computacionales pero son más precisos y versátiles que los semiempíricos (ya que se requieren mucho menos parámetros y éstos son independientes del sistema). Dentro de los métodos DFT encontramos una gran variedad de funcionales, incluyendo los LDA (*local density approximation*), GGA (*gradient corrected approximation*), meta-GGA, y los híbridos o meta-híbridos.¹⁰ Los más populares para cálculos QM/MM son los GGA (BP86, BLYP, PBE, ...) y los híbridos (B3LYP, PBE0, M06, ...). Los GGA son menos costosos en recursos computacionales y dan una buena descripción de la geometría y las interacciones del sistema. Los funcionales híbridos que incluyen un cierto porcentaje de intercambio de Hartree-Fock presentan un error menor al calcular estados de transición (barreras energéticas), y seguramente por ello el B3LYP es con diferencia, el funcional más usado hoy en día. La fuerte dependencia de los resultados a nivel cuantitativo en función del funcional aplicado es un inconveniente importante del método. Además, las limitaciones de los métodos DFT en interacciones de van der Waals son conocidas.¹¹
- Los métodos post-HF incluyen el cálculo de la energía de correlación mediante alguna de las siguientes

aproximaciones: la interacción de configuraciones¹² (CI, CISD, QCISD(T), ...), la cual representa la función de onda como combinación lineal de varios estados ocupados; añadiendo una corrección al resultado HF basada en la teoría de perturbación de Møller-Plesset¹³ (MP2, MP3, ...); o el método de *coupled-cluster*^{14,15} (CCD, CCSD, ...) entre otros. Estos métodos son mucho más precisos que los anteriores. Sin embargo, su elevado coste computacional impide que puedan usarse en modelos realistas para resolver problemas de enzimología computacional.

El método MM

El método MM se define por el campo de fuerzas usado. Actualmente hay una gran variedad de campos de fuerzas para proteínas (CHARMM22*,¹⁶ AMBERFF99SB-ildn,¹⁷ OPLS-AA/L,¹⁸ etc), carbohidratos (GLYCAM06h¹⁹) o ácidos nucleicos (CHARM27,²⁰ parmbsc0,²¹ etc). La evolución de los campos de fuerza para simulaciones QM/MM se centra en considerar el efecto de la región QM sobre las cargas parciales de la región MM mediante campos de fuerza polarizables.

Acoplamiento mecánico vs electrónico

La energía total del sistema (E_{total}) se puede calcular de forma aditiva o sustractiva.

- Según la forma aditiva (o de integración electrónica) se define:

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (1)$$

donde E_{QM} es la energía de la región QM calculada con mecánica cuántica, E_{MM} es la energía de la región MM calculada mediante mecánica clásica y $E_{QM/MM}$ es la energía de interacción QM/MM:

$$E_{QM/MM} = E_{QM/MM} (electr.) + E_{QM/MM} (vdW) + E_{QM/MM} (enlace) \quad (2)$$

donde los términos de van der Waals y de enlace se calculan con mecánica molecular y el electrostático se calcula con mecánica cuántica (de forma que se considera la polarización de la región QM por el efecto de la región MM).

- Según la forma sustractiva (o de integración mecánica), se define:

$$E_{total} = E_{QM} + E_{MM} (total) - E_{MM} (QM) \quad (3)$$

donde E_{QM} es la energía de la región QM, $E_{MM} (total)$ es la energía clásica de todo el sistema y $E_{MM} (QM)$ es la energía clásica de la región QM. Por consiguiente, en este caso las interacciones electrostáticas QM/MM se calculan a partir de cargas puntuales y sin considerar la polarización de la región QM. El método multicapas ONIOM²² es un ejemplo de esta categoría.

La partición QM/MM e interfase

Uno de los factores cruciales en una simulación QM/MM es la partición del sistema entre las regiones QM y MM. No hay ninguna regla para decidir cuáles son los átomos que deben

incluirse en la región QM y cuáles no. El procedimiento más robusto es considerar primero un sistema mínimo QM basado en un criterio químico, y valorar la variación de los resultados cuando se añaden más residuos (por ejemplo, la variación en una diferencia de entalpías o variaciones en la geometría del centro activo). Otros procedimientos incluyen el uso de regiones QM y MM que son variables a lo largo de la simulación.²³

Si el truncamiento entre las regiones QM/MM se da en un enlace covalente (por ejemplo, entre la cadena lateral y el esqueleto de un residuo), se debe asegurar que la estructura electrónica no se ve afectada. La forma más común es la de saturar la región QM con un átomo ficticio (generalmente un hidrógeno), con un pseudoenlace que sustituye el enlace QM-MM o con un orbital híbrido sobre el átomo MM.²⁴

Técnicas de simulación avanzadas

La principal aplicación de las simulaciones de dinámica molecular QM/MM en la enzimología computacional es el estudio del mecanismo de reacción en enzimas. La principal ventaja de las técnicas *in silico* es que se puede identificar y caracterizar todas las especies químicas que intervienen en el camino de reacción, aunque sean tan efímeras como un estado de transición. Sin embargo, conocer el mecanismo de reacción en un sistema enzimático no es una tarea trivial, ya que suelen ser sistemas multidimensionales definidos por perfiles de energía libre rugosos y con varios estados metaestables (Figura 3).

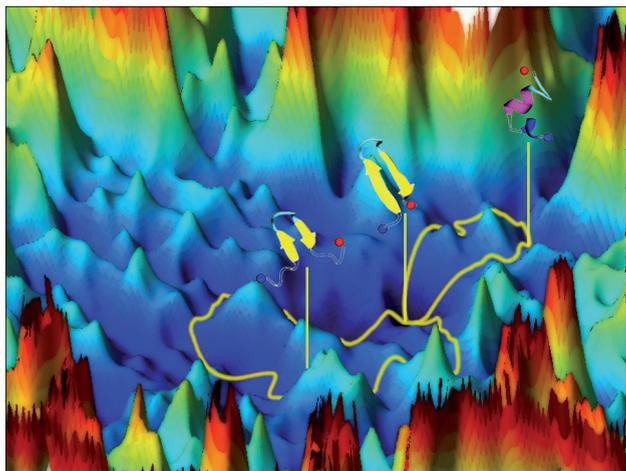


Figura 3. Ilustración de la superficie de energía libre conformacional de un péptido en disolución en función de 2 variables colectivas arbitrarias.

En consecuencia, comúnmente se obtiene el camino de reacción como el camino de mínima energía (potencial o libre) definido en un espacio reducido de coordenadas de reacción o variables colectivas (CV; por ejemplo, la distancia entre un nucleófilo y un electrófilo). A partir del perfil de energía libre y asumiendo la teoría del estado de transición,²⁵ se puede estimar la constante de velocidad:

$$k(T) = \gamma(T)(k_B T/h)(C^0)^{n-1} \exp[-\Delta G^{TS,0}(T)/RT] \quad (4)$$

donde C^0 es la concentración en condiciones estándar, n es el orden de la reacción y $\Delta G^{TS,0}$ es la diferencia de energía libre

entre el estado de transición y los reactivos. $\gamma(T)$ es el coeficiente de transmisión el cual incluye las contribuciones de recruzamiento dinámico, de efecto túnel y de las desviaciones de la distribución de equilibrio de Boltzmann. Es importante notar que esta reducción de la dimensionalidad es en sí una simplificación que puede acarrear problemas derivados de la superposición de conformaciones que son realmente distintas en el mismo valor de las CV. Por lo tanto, es necesario definir las CV de forma que las conformaciones superpuestas están separadas por barreras bajas o irrelevantes. El perfil de energía potencial se obtiene mediante cálculos puntuales de minimización de la energía a lo largo del camino de reacción para una de las múltiples conformaciones del sistema. El perfil de energía libre en cambio, requiere un muestreo extenso de todos los estados de la reacción. La dificultad principal para ello reside en que es prácticamente imposible explorar los estados de alta energía mediante “simple” dinámica molecular, dado que la probabilidad de explorar una configuración determinada depende exponencialmente de la energía F^* de dicha configuración conforme a $\sim \exp[-F^*/k_B T]$. Esto significa que, en la práctica, el sistema no puede escapar de los mínimos locales de energía, por lo que el cálculo de energías libres es uno de los problemas más difíciles de la mecánica estadística.

Para solventar este inconveniente, se han desarrollado una gran variedad de métodos que aceleran la exploración del espacio conformacional y la transición de barreras de energía. Estos métodos se pueden dividir en dos grupos generales:

- En el primer grupo, el conjunto de conformaciones exploradas no sigue una distribución de Boltzmann sino que se favorecen las conformaciones de alta energía. Este grupo incluye el *transition path sampling* (TPS),²⁶ *nudged elastic bands* (NEB),²⁷ *locally enhanced sampling* (LES),²⁸ o el *replica exchange molecular dynamics* (REMD),²⁹ entre otros.

En el TPS, se busca simular una gran cantidad de trayectorias reactivas (cortas) que crucen el estado de transición y las zonas de alta energía, escogiendo los puntos del camino mediante un método de Monte Carlo. El NEB es un método para optimizar el camino de reacción mediante la minimización de la energía de un número de puntos que unen reactivos y productos sujetos a un potencial externo que los une (simulando una banda elástica). El REMD no requiere definir ninguna variable colectiva ni camino de reacción. En cambio, varias réplicas del mismo sistema se simulan a distintas temperaturas, y periódicamente se intenta intercambiar las configuraciones entre réplicas adyacentes con una probabilidad $p = \min(1, e^{\Delta_{i,j}})$ en la que $\Delta_{i,j} = (\beta_i - \beta_j)(U(\mathbf{R}_i) - U(\mathbf{R}_j))$ y $U(\mathbf{R}_i)$ es la energía potencial de la réplica i . De esta forma, el sistema se calienta y se enfría varias veces a lo largo de la simulación, manteniendo la colectividad canónica. La idea detrás del REMD es que es más fácil cruzar las barreras a temperaturas mayores.

- El segundo grupo se compone por los métodos que aplican un potencial externo sobre la energía potencial del sistema que sesga la probabilidad de explorar las distintas regiones del perfil de energía libre. Este grupo incluye métodos como el *umbrella sampling* (US),³⁰ *conformatio-*

nal flooding,³¹ accelerated molecular dynamics (aMD),³² hyper-MD,³³ o metadynamics (MTD),³⁴ entre otros.

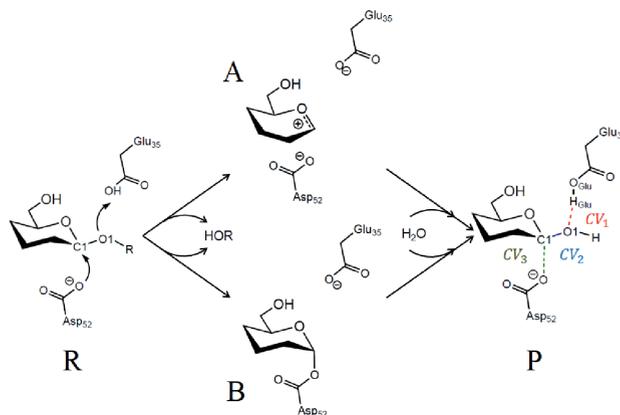
El método de US realiza un número de dinámicas moleculares a lo largo del camino de reacción y/o el espacio de las variables colectivas de forma que exploren regiones distintas del perfil de energía libre (limitadas mediante un potencial armónico), pero con un cierto solapamiento entre ellas. La energía libre se obtiene combinando las distintas distribuciones de probabilidad mediante el *weighted histogram analysis method* (WHAM)³⁵ o similar. En el *conformational flooding* se aplica un potencial externo al Hamiltoniano que desestabiliza el estado de reactivos y por lo tanto, disminuye la barrera de energía. Similarmente, la MTD también aplica un potencial externo sobre el Hamiltoniano que empuja el sistema a explorar regiones de alta energía. Sin embargo, en el caso de la MTD el potencial se obtiene como la suma de pequeños potenciales de tipo Gaussiana que se añaden periódicamente en las regiones del perfil de energía libre que se hayan explorado. Por lo tanto, se trata de un “potencial histórico” que evoluciona a medida que avanza la simulación. Ello presenta tres ventajas fundamentales: a) No es necesario saber ni estimar el perfil de energía libre *a priori* ya que el potencial se construye según la simulación avanza. b) El sesgo actúa sobre todo el espacio de energía libre, con lo que se puede escapar de múltiples mínimos y que favorece el recruzamiento de los estados de transición. c) Al final de la simulación, el potencial efectivo (la suma del potencial real y el sesgo) es aproximadamente plano. Por lo tanto, el sistema difunde libremente por todo el espacio de las variables colectivas y además, el potencial real se puede obtener como el inverso del potencial externo.

El mecanismo de reacción de la lisozima

La lisozima es una glicosil hidrolasa que actúa con retención de la configuración en el carbono anomérico y ha sido un modelo enzimático tanto experimental (fue la primera enzima para la que se resolvió la estructura de rayos X³⁶) como computacionalmente. En el campo de la química teórica, la primera simulación de mecánica cuántica/mecánica molecular (QM/MM) fue realizada por Warshel y Levitt³⁷ en 1976 sobre la formación del complejo de Michaelis y el mecanismo de reacción en la lisozima. En este trabajo, se estudió el efecto del encaje inducido y el efecto electrostático del residuo Asp52 en el mecanismo de reacción y la energía de activación en la lisozima (Esquema 1).

Los autores calcularon la superficie de energía potencial mediante una serie de optimizaciones de la geometría y cálculos puntuales, y encontraron una estabilización importante por efecto electrostático del Asp52 sobre el sustrato en la formación del complejo de Michaelis que favorece la reacción. Sin embargo, no contemplaron la posible formación de un intermedio covalente glicosil-enzima en el mecanismo de reacción (rígido) que estudiaron.

Posteriormente,⁹ la formación del intermedio (covalente o de par-iónico) en la misma enzima fue estudiada por métodos de dinámica molecular QM/MM y *umbrella sampling* (incluyendo explícitamente la flexibilidad del sistema). La simu-



Esquema 1. Mecanismo de reacción de la lisozima. A) El mecanismo propuesto por Phillips³⁸ predice un intermedio de reacción oxocarbenio. B) El mecanismo propuesto por Koshland³⁹ predice un mecanismo de doble desplazamiento con un intermedio covalente glicosil-enzima.

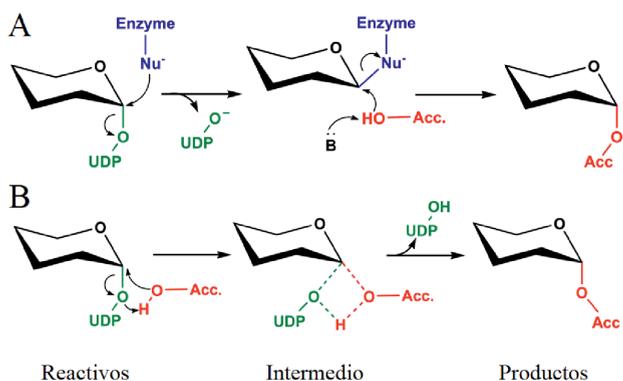
lación de US se inició desde el estado de productos (P en el esquema 1) y se usaron dos variables colectivas: la diferencia de distancias del protón del Glu52 (H_{Glu}) con los oxígenos O1 y O_{Glu} para describir la transferencia de protón y la distancia C1-O1 para simular la formación del intermedio. Se usó el método semiempírico PM3/CHARMM22 para explorar >1ns de MD, y se promedió la energía libre con el método WHAM. Debido a la poca precisión del método semiempírico se aplicó unas correcciones a la energía mediante cálculos B3LYP/6-311+G(2d) y MP2/6-311+G(2d) sobre un modelo del sistema en fase gas.

Con esta metodología, las barreras de energía para la transferencia electrónica y la posterior rotura del enlace C1-O1 se estimaron en 16 y 18 kcal·mol⁻¹ respectivamente. Es interesante notar que, antes de aplicar las correcciones a la energía, las mismas barreras energéticas obtenidas con el nivel de teoría semiempírico eran de 35 y 44 kcal·mol⁻¹ respectivamente. Significativamente, después de la transferencia de protón y al romper el enlace C1-O1, el sistema evolucionó espontáneamente a la formación del intermedio covalente glicosil-enzima con el Asp52. En una tercera simulación de US (usando CV_3 en el Esquema 1), se estimó la diferencia de energía libre entre el intermedio covalente y el par iónico en 30 kcal·mol⁻¹. Los autores concluyeron que el mecanismo de la lisozima pasa necesariamente por la formación de un intermedio covalente, lo que coincide con evidencias experimentales previas.^{40,41} Recientemente, se ha simulado la formación del intermedio covalente glicosil-enzima en otras glicosidasas⁴²⁻⁴⁶ también con métodos basados en la dinámica molecular, confirmando el mecanismo de doble desplazamiento propuesto por Koshland.

Mecanismo de reacción de la trehalosa-6-fosfato sintasa

Las glicosil transferasas (GTs) catalizan la transferencia de un monosacárido con inversión o retención de la configuración del carbono anomérico.⁴⁷ El mecanismo de las GTs que actúan con inversión de la configuración es bien conoci-

do:⁴⁸ una única reacción de tipo S_N2 en la que un residuo de tipo base general aumenta la nucleofilia de la molécula aceptora, análogo al mecanismo usado por las glicosil hidrolasas que actúan con inversión de la configuración (Esquema 2).



Esquema 2. Mecanismos de reacción propuestos para las GTs que actúan con retención de la configuración. A) Mecanismo de doble desplazamiento. B) Mecanismo *front-face* propuesto para las GTs que no tienen nucleófilo en el centro activo.

Por el contrario, el mecanismo de las GTs que actúan con retención de la configuración es desconocido. Por analogía con las glicosil hidrolasas que actúan con retención de la configuración (esquema 1B), se ha propuesto un mecanismo de doble desplazamiento, que implica la presencia de un intermediario covalente glicosil enzima (esquema 2A).⁴⁹ Sin embargo, hay otras GTs para las que no se ha encontrado ningún residuo cerca del centro activo que pueda actuar de base general, como la glicógeno fosforilasa, la α -galactosiltransferasa (LgtC) o la misma OtsA (Figura 4). El mecanismo de doble desplazamiento no es viable para estas enzimas sin contar con un cambio conformacional que cambie completamente la disposición de los residuos que conforman el sitio activo. Por consiguiente, se propuso un mecanismo (esquema 2B) similar a una S_N1 o una S_Ni.⁵⁰ A diferencia del mecanismo de doble desplazamiento, el intermediario de reacción no es covalente glicosil-enzima sino que sería de tipo oxocarbenio, o incluso podría no haber ningún intermediario (habría un simple estado

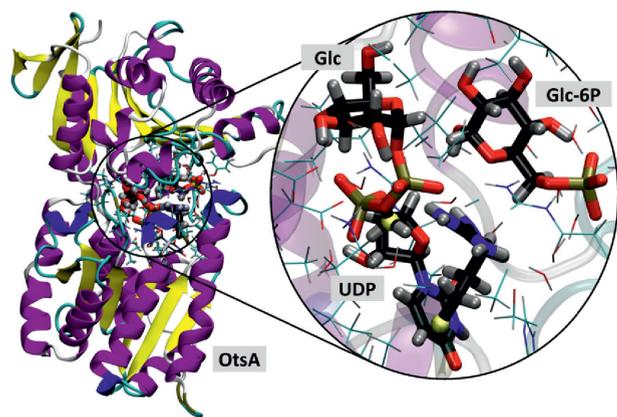
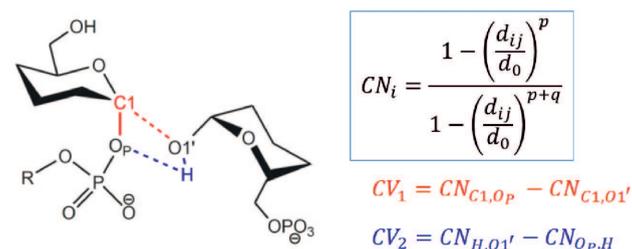


Figura 4. Complejo de Michaelis de la OtsA con los sustratos UDP-Glu y Glc-6P. La región QM se muestra en la ampliación.

de transición). Entre las evidencias en favor del mecanismo de doble desplazamiento destacan: la observación por espectrometría de masas de la formación de un intermediario covalente glicosil-enzima en el mutante E303C de la α -1,3-N-acetilgalactosaminil transferasa y de la α -1,3-galactosil transferasa (las enzimas GTA y GTB respectivamente, responsables de la síntesis de los antígenos que definen los grupos sanguíneos A y B);⁵¹ el rescate de la actividad enzimática del mutante inactivo E317A de la α 3-galactosil transferasa (α 3GalT) mediante un nucleófilo externo (azida),⁵² o los estudios teóricos del mecanismo de reacción sobre la α 3GalT.^{53,54}

En favor del mecanismo de tipo S_Ni se publicó recientemente un estudio estructural por difracción de rayos X del complejo enzima-ligando usando la OtsA y un inhibidor análogo al estado de transición [validoxilamina-6-fosfato (VA6P)].⁵⁵ Así mismo, un estudio de efecto isotópico cinético y de relación lineal de energía libre (LFER) con la misma enzima OtsA apoya también un mecanismo de S_Ni.⁵⁶ Desde el punto de vista teórico, el mecanismo de S_Ni se estudió primeramente en un modelo en fase gas de la LgtC, concluyendo que la reacción ocurre en un solo paso.⁵⁷ En el ejemplo que se presenta,⁵⁸ utilizamos un método de cálculo de energías libres (metadinámica) para estudiar el mecanismo de reacción de la OtsA. El sistema se describe mediante un método híbrido QM/MM.

La simulación de metadinámica QM/MM de la reacción de transferencia de glucosa se empezó desde el complejo de Michaelis (Figura 4) modelado a partir de la estructura cristalográfica del complejo enzima-UDP-inhibidor (OtsA·UDP·VA6P, código PDB: 2WTX).⁵⁵ El potencial histórico se aplicó en el espacio de dos variables colectivas que describen la rotura y formación de los enlaces covalentes durante la reacción (Esquema 3), definidas como diferencias de índices de coordinación.



Esquema 3. Variables colectivas usadas en la simulación de metadinámica. CV1 describe la rotura del enlace Glc-UDP y la formación del enlace glicosídico; CV2 describe la transferencia de protón.

En la superficie de energía libre obtenida de la simulación de metadinámica (Figura 5) se aprecian dos mínimos estables que corresponden a los estados reactivos (R) y productos (P), separados por una barrera de 23 kcal·mol⁻¹, parecido a los valores de entre 16 y 21 kcal·mol⁻¹ estimados teórica y experimentalmente para otras glicosil transferasas.^{51,52,56,59}

La reacción empieza con la elongación del enlace C1-OP (Figura 6). Simultáneamente, el hidroxilo del C1 del aceptor pasa de hacer puente de hidrógeno con uno de los oxígenos libres del fosfato para hacerlo con el oxígeno Op. Este nuevo puente de hidrógeno estabiliza la carga negativa que se está desarrollando en Op y ayuda a posicionar la molécula acep-

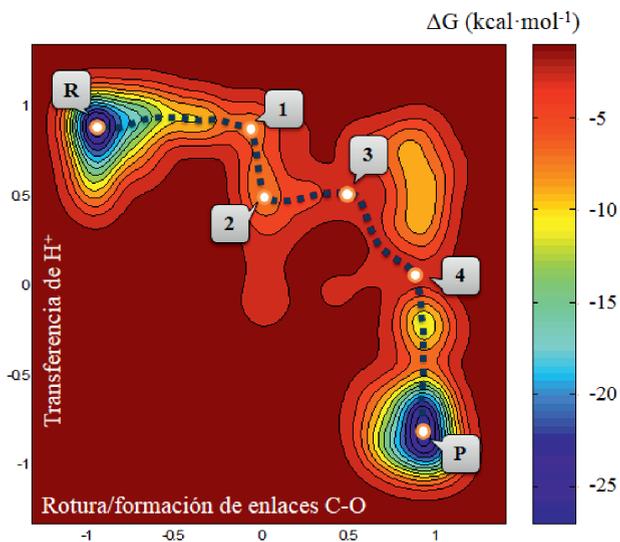


Figura 5. Superficie de energía libre de la reacción de transferencia de glucosa (en función de las variables colectivas descritas en el texto). Cada línea de contorno representa 2 kcal·mol⁻¹.

tora adecuadamente para reaccionar con el glicosilo que se está formando. El resultado de la ruptura del enlace C1-O_P es la formación de un par iónico UPD/oxocarboación, en el que el C1 está lejos de ambos oxígenos O_P y O1'. En este punto de la reacción, el carbono C1 pasa de tener una hibridación sp³ a una sp², lo que conlleva a una disposición plana de los átomos adyacentes. La deficiencia electrónica

en C1 es estabilizada por una transferencia de carga por parte del oxígeno piránico que se hace evidente por el acortamiento de la distancia entre ambos átomos. Estos cambios conformacionales producen un cambio de la conformación del anillo de glucosa que pasa de una conformación de silla (⁴C₁) a una más tensionada de media-silla (⁴H₃) en la que los átomos C2-C1-O5-C5 están en el mismo plano en los estados de la reacción 1 a 3. Durante el estado 2, la distancia O1'-H...O_P se hace más corta y el glucosil se desplaza cerca del átomo O1' con el cual forma un enlace covalente en el estado 3. Seguidamente y sin ninguna barrera se transfiere el protón del enlace glicosídico al grupo fosfato (estado 4) y se llega a productos.

La reacción puede dividirse en dos pasos, siendo el primero la ruptura de la molécula dadora (de R a 2) para formar la especie reactiva oxocarboación, y el segundo la formación del enlace glicosídico concomitantemente con la transferencia de protón (de 3 a P). Este mecanismo puede describirse como una S_N1 o como una S_Ni con un intermedio con un tiempo de vida muy corto, y concuerda con los resultados de efecto isotópico cinético⁵⁶ y estructurales⁵⁵ sobre la OtsA. Posiblemente este mecanismo de tipo S_Ni/S_N1 es común en aquellas glicosil transferasas que no tienen un residuo nucleófilo en el centro activo. En un estudio reciente realizado en el grupo sobre la α3galT se ha observado la presencia de un intermedio covalente glicosil-enzima y por lo tanto un mecanismo de doble desplazamiento. Sin embargo, podría ser que dicho mecanismo esté compitiendo con el mecanismo de tipo S_Ni.

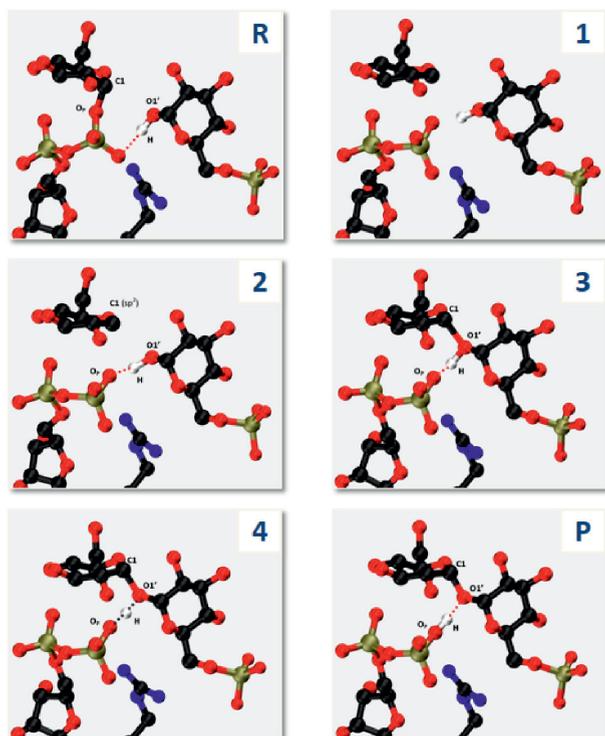


Figura 6. Instantáneas de las diferentes especies encontradas en el camino de reacción. Los átomos de hidrógeno no se muestran (excepto el que interviene en la reacción).

Conclusiones

En el presente artículo se ha querido mostrar la evolución de los métodos teóricos aplicados a la enzimología computacional, haciendo énfasis en los métodos QM/MM y de cálculo de energías libres. La combinación de estas dos metodologías permite la consideración explícita del entorno enzimático y los efectos de segunda capa, y al mismo tiempo posibilita un muestreo mucho más eficiente del perfil de energía libre, incluyendo los factores entrópicos y la flexibilidad de la proteína o el ligando. Hoy es comúnmente aceptada la necesidad de incluir la dinámica en cualquier cálculo de mecanismo enzimático, y el debate se centra en decidir hasta qué punto la flexibilidad de la enzima participa en la catálisis. A pesar de que los resultados a nivel cuantitativo son todavía dispares y dependen fuertemente del método usado, a nivel cualitativo son suficientes para contribuir al desarrollo y prueba de hipótesis de mecanismos enzimáticos, de transferencia electrónica o al análisis de interacciones enzima-ligando. En este sentido, los métodos *in silico* han demostrado ser especialmente útiles en combinación con técnicas experimentales, y no es de extrañar que las colaboraciones entre grupos experimentales y teóricos sea cada vez más extendida.

Agradecimientos

El autor agradece a la EMBO por la beca post-doctoral y a la Prof. C. Rovira y a Dra. E. Formoso por las correcciones al manuscrito.

Bibliografía

1. Q. F. Yu, W. Ye, W. Wang, H. F. Chen, *PLoS One* **2013**, *8*.
2. D. D. Boehr, P. E. Wright, *Science* **2008**, *320*, 1429.
3. R. O. Dror, R. M. Dirks, J. P. Grossman, H. F. Xu, D. E. Shaw, *Ann. Rev. Biophys.* **2012**, *41*, 429.
4. M. L. Klein, W. Shinoda, *Science* **2008**, *321*, 798.
5. K. Meier, A. Choutko, J. Dolenc, A. P. Eichenberger, S. Riniker, W. F. van Gunsteren, *Angew. Chem. Int. Ed.* **2013**, *52*, 2820.
6. R. Car, M. Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
7. M. W. van der Kamp, A. J. Mulholland, *Biochemistry* **2013**, *52*, 2708.
8. C. Rovira, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, n/a.
9. A. L. Bowman, I. M. Grant, A. J. Mulholland, *Chem. Commun.* **2008**, 4425.
10. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
11. A. J. Cohen, P. Mori-Sanchez, W. T. Yang, *Chem. Rev.* **2012**, *112*, 289.
12. J. A. Pople, R. Seeger, R. Krishnan, *Int. J. Quant. Chem.* **1977**, 149.
13. C. Moller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 0618.
14. R. J. Bartlett, G. D. Purvis, *Int. J. Quant. Chem.* **1978**, *14*, 561.
15. J. A. Pople, R. Krishnan, H. B. Schlegel, J. S. Binkley, *Int. J. Quant. Chem.* **1978**, *14*, 545.
16. S. Piana, *Biophys. J.* **2011**, *101*, 1015.
17. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins* **2010**, *78*, 1950.
18. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474.
19. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622.
20. A. Perez, F. Lankas, F. J. Luque, M. Orozco, *Nucleic Acids Res.* **2008**, *36*, 2379.
21. A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817.
22. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
23. A. Heyden, H. Lin, D. G. Truhlar, *J. Phys. Chem. B* **2007**, *111*, 2231.
24. Y. K. Zhang, *Theor. Chem. Acc.* **2006**, *116*, 43.
25. M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186.
26. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
27. G. Henkelman, H. Jonsson, *J. Chem. Phys.* **2000**, *113*, 9978.
28. A. Roitberg, R. Elber, *J. Chem. Phys.* **1991**, *95*, 9277.
29. Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141.
30. G. M. Torrie, J. P. Valleau, J. P. *J. Comp. Phys.* **1977**, *23*, 187.
31. H. Grubmuller, *Phys. Rev. E* **1995**, *52*, 2893.
32. D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, *120*, 11919.
33. A. F. Voter, *Phys. Rev. Lett.* **1997**, *78*, 3908.
34. A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
35. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J. Comput. Chem.* **1992**, *13*, 1011.
36. R. Diamond, *J. Mol. Biol.* **1974**, *82*, 371.
37. A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.
38. D. C. Phillips, *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 484.
39. D. E. Koshland, *Biol. Rev. Camb. Philos. Soc.* **1953**, *28*, 416.
40. A. J. Kirby, *Nature Struct. Biol.* **2001**, *8*, 737.
41. D. J. Vocadlo, G. J. Davies, R. Laine, S. G. Withers, *Nature* **2001**, *412*, 835.
42. J. Jitonnorn, V. S. Lee, P. Nimmanpipug, H. A. Rowlands, A. J. Mulholland, *Biochemistry* **2011**, *50*, 4697.
43. X. Biarnes, A. Ardevol, J. Iglesias-Fernandez, A. Planas, C. Rovira, *J. Am. Chem. Soc.* **2011**, *133*, 20301.
44. L. Petersen, A. Ardevol, C. Rovira, P. J. Reilly, *J. Phys. Chem. B* **2009**, *113*, 7331.
45. I. J. Barker, L. Petersen, P. J. Reilly, *J. Phys. Chem. B* **2010**, *114*, 15389.
46. L. Petersen, A. Ardevol, C. Rovira, P. J. Reilly, *J. Am. Chem. Soc.* **2010**, *132*, 8291.
47. D. J. Vocadlo, G. J. Davies, *Curr. Opin. Chem. Biol.* **2008**, *12*, 539.
48. L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, *Annu. Rev. Biochem.* **2008**, *77*, 521.
49. L. N. Gastinel, C. Bignon, A. K. Misra, O. Hindsgaul, J. H. Shaper, D. H. Joziassse, *EMBO J.* **2001**, *20*, 638.
50. K. Persson, H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers, N. C. Strynadka, *Nature Struct. Biol.* **2001**, *8*, 166.
51. N. Soya, Y. Fang, M. M. Palcic, J. S. Klassen, *Glycobiology* **2011**, *21*, 547.
52. A. Monegal, A. Planas, *J. Am. Chem. Soc.* **2006**, *128*, 16030.
53. H. Gomez, J. M. Lluch, L. Masgrau, *Carbohydr. Res.* **2012**, *356*, 204.
54. H. Gomez, J. M. Lluch, L. Masgrau, *J. Am. Chem. Soc.* **2013**, *135*, 7053.
55. J. C. Errey, S. S. Lee, R. P. Gibson, C. M. Fleites, C. S. Barry, P. M. J. Jung, A. C. O'Sullivan, B. G. Davis, G. J. Davies, *Angew. Chem. Int. Ed.* **2010**, *49*, 1234.
56. S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies, B. G. Davis, *Nature Chem. Biol.* **2011**, *7*, 631.
57. I. Tvaroska, *Carbohydrate Res.* **2004**, *339*, 1007.
58. A. Ardevol, C. Rovira, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10897.
59. H. S. Seo, Y. J. Koo, J. Y. Lim, J. T. Song, C. H. Kim, J. K. Kim, J. S. Lee, Y. D. Choi, *Appl. Environ. Microbiol.* **2000**, *66*, 2484.