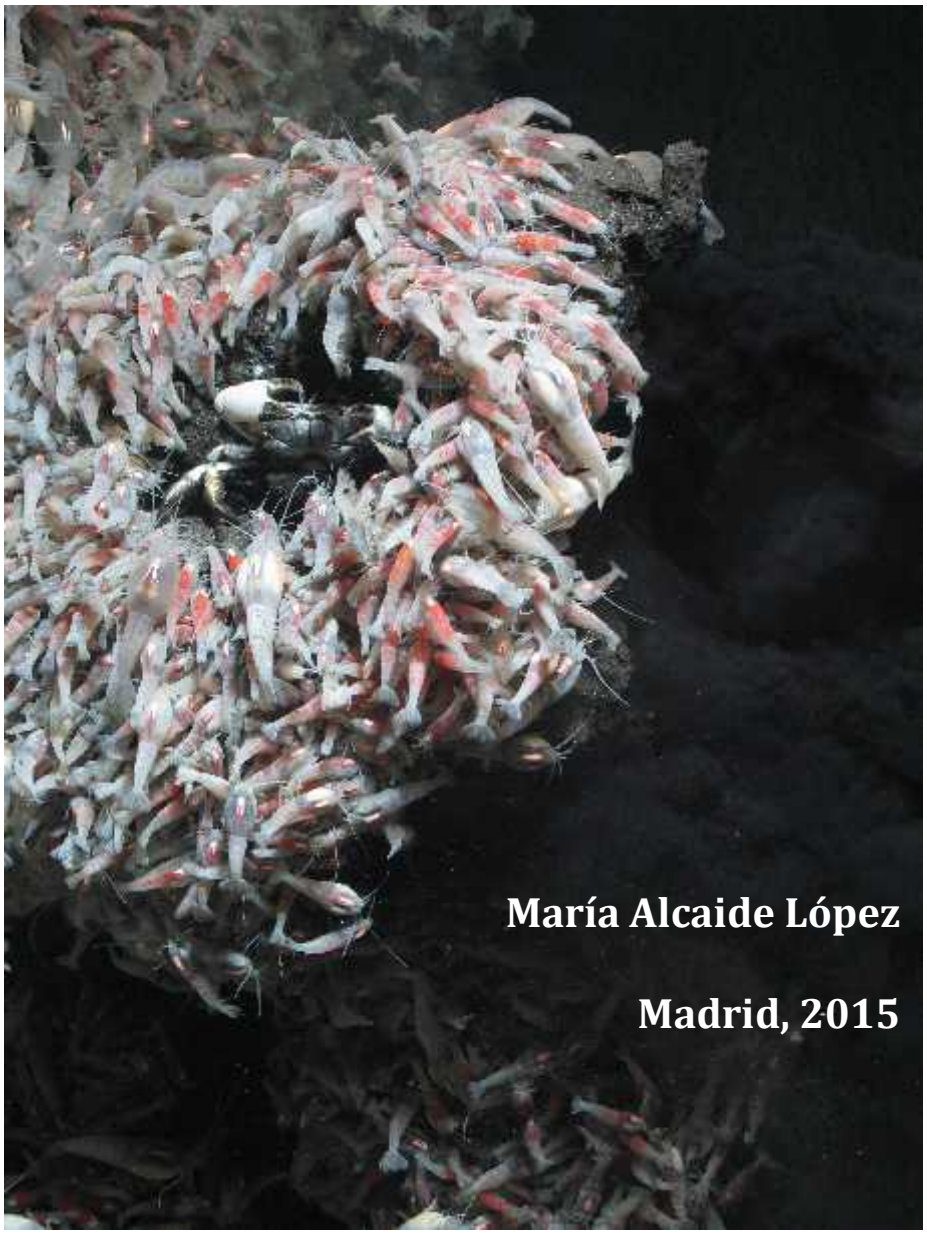
PROGRAMA DE DOCTORADO EN BIOQUÍMICA, BIOLOGÍA MOLECULAR, BIOMEDICINA Y BIOTECNOLOGÍA
DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID

Meta-genómica aplicada a estudios estructura-función en comunidades microbianas de diferentes sistemas acuáticos



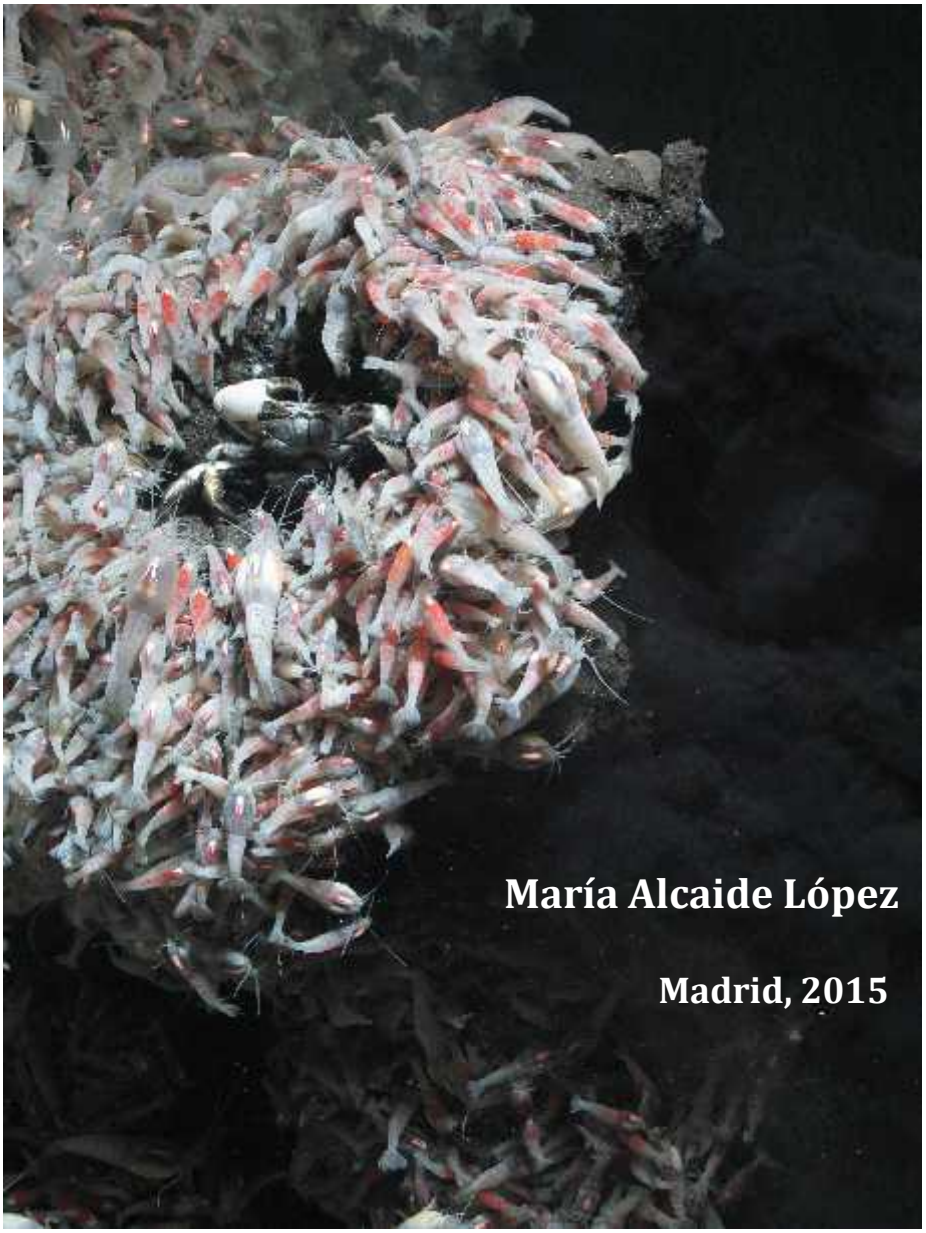
María Alcaide López

Madrid, 2015

Imagen de portada cortesía de Ifremer-Victor/BICOSE 2014. Especímenes de *Rimicaris exoculata* a lo largo de una chimenea profunda en un gradiente de fluidos hidrotermales fríos (8.7ºC), oxigenados, y ligeramente salinos (23.03 g kg^{-1}) en la dorsal Atlántica (*Rainbow Ridge*). La gamba *R. exoculata* viene en poblaciones de gran densidad (aproximadamente 2,055 individuos por m^2) a lo largo de las chimeneas. Las gambas tienen dos complejas comunidades microbianas de simbiontes independientemente de las condiciones geoquímicas. Una localizada en las agallas y está implicada en la nutrición de la gamba a través de quimio-síntesis. Una segunda está localizada en su intestino. En la foto se muestran especímenes adultos y juveniles alrededor de un cangrejo, *Segonzacia mesatlantica*.

PHD PROGRAM OF EN BIOCHEMISTRY, MOLECULAR BIOLOGY, BIOMEDICINE AND BIOTECHNOLOGY
DEPARTMENT OF MOLECULAR BIOLOGY
FACULTY OF SCIENCE
AUTONOMOUS UNIVERSITY OF MADRID

Meta-genomic approaches applied to structural-functional studies in microbial communities from different aquatic habitats



María Alcaide López

Madrid, 2015

Cover photograph (Courtesy of Ifremer-Victor/BICOSE 2014; reprinted with permission.): *Rimicaris exoculata* specimens along deep chimney walls in the gradient between hydrothermal fluids and cold (8.7°C), oxygenated, and slightly saline (23.03 g kg^{-1}) waters in the Mid-Atlantic Rainbow Ridge. The symbiotic shrimp *R. exoculata* lives in large aggregates (about 2,055 individuals per m^2) covering active chimneys. This shrimp harbors two complex symbioses regardless of the geochemical conditions. One is located in its enlarged gill chamber and is at least implied in the shrimp's nutrition through chemosynthesis. A second one is located in its gut. In this photo, adults and recently recruited juveniles are surrounding a *Segonzacia mesatlantica* crab.

Bajo la supervisión de:

1. Dr. Manuel Ferrer Martínez
Laboratorio de Biotecnología de Sistemas
Departamento de Biocatálisis Aplicada
Consejo Superior de Investigaciones Científicas
Instituto de Catálisis y Petroleoquímica

MANUEL FERRER MARTÍNEZ, DR. EN CIENCIAS QUÍMICAS, INVESTIGADOR CIENTÍFICO DEL CSIC

CERTIFICA: Que el presente trabajo "Meta-genómica aplicada a estudios estructura-función en comunidades microbianas de diferentes sistemas acuáticos", que constituye la Memoria que presenta la Licenciada en Químicas por la Universidad Complutense de Madrid, María Alcaide López, para optar al grado de Doctora, ha sido realizado bajo su dirección en el Departamento de Biocatálisis Aplicada del Instituto de de Catálisis y Petroleoquímica del CSIC, Campus de Excelencia Internacional UAM+CSIC, Madrid.

Y para que conste, firma el presente certificado en Madrid, a 10 de febrero de 2015.

Dr. Manuel Ferrer Martínez

AGRADECIMIENTOS

Después de todo este tiempo, y todas las cosas que han pasado, no me creo que esté escribiendo los agradecimientos a todos los que habéis estado conmigo de una manera u otra y que habéis colaborado en que este sueño se haga realidad. A todos muchas, muchas, muchas gracias; sin vosotros no hubiera sido capaz de llegar a donde he llegado, GRACIAS.

A mi director de tesis, Dr. Manuel Ferrer Martínez, por ser mi maestro, mi gran ayuda en los buenos y malos momentos, por creer en mí, por darme una oportunidad, por estar siempre dispuesto a escucharme, por no dejar de confiar en mí en todo este tiempo y por enseñarme e inspirarme en todo lo que he aprendido, gracias de corazón.

A mi tutora académica, la Prof. María Fernández Lobato, del Centro de Biología Molecular Severo Ochoa, por toda la ayuda que siempre ha estado dispuesta a darme.

A los Drs. Alexei Savchenko y Alexander F. Yakunin, del Departamento de Ingeniería Química y Química Aplicada de la Universidad de Toronto (Canadá), por darme la oportunidad de haber trabajado y aprendido en sus laboratorios durante mi estancia de dos meses en Toronto. Fue una experiencia maravillosa! Y a todas las personas que forman su grupo, en especial a Hong, Xiaohui, Peter, Anna, Anatoli, Alexander, Greg, etc. Por hacer de mi estancia unos días inolvidables y por todo lo que me enseñaron, nunca les olvidaré, muchas gracias.

Al Dr. Michail M. Yakimov, del Instituto de Costas del Medio Marino (CNR, Messina, Italia), por permitirme la experiencia de aprender en su laboratorio y conocer a su equipo, Enzo, Francesco, Violetta….. Por lo bien que me trataron y todo lo que aprendí de ellos.

Al Dr. Peter N. Golyshin y a la Dra. Olga V. Golyshina, de la Escuela de Ciencias Biológicas en la Universidad de Bangor (UK), por toda su ayuda y apoyo en este trabajo.

A la Dra. Ruth Matesanz, al Dr. José M. Andreu y a Dra. María A. Olivia (CIB-CSIC, Madrid), por enseñarme todo lo que sé de los espectros de dicroísmo circular, y por estar siempre dispuestos a ayudarme.

A las instituciones financiadoras y colaboradoras que han hecho posible la realización de esta Tesis Doctoral, al Ministerio de Economía y Competitividad, a la Unión Europea, al CSIC y a la UAM. En particular el presente trabajo se ha desarrollado dentro de los proyectos MAMBA y BIO2011-25012.

A los directores del Instituto de Catálisis y Petroleoquímica, Prof. Joaquín Pérez Pariente y Prof. José Carlos Conesa Cegarra.

Al gerente del Instituto de Catálisis y Petroleoquímica, José María Martínez Pacheco.

A los miembros del Tribunal de Tesis por haber aceptado formar parte de él.

Al Instituto de Catálisis y Petroleoquímica (ICP-CSIC) y a su personal, por su ayuda con los papeleos y/o ayuda técnica e informática (Nuria, Alberto, Enrique, Rosa, Patricio, Chema, Puerto, Paloma, Jesús, José, Ramón…)…, a Pili y, especialmente, a Carmen.

Al Dr. Antonio López de Lacey, porque fue el primero en querer darme una oportunidad junto con el Dr. Marcos Pita Martínez. A los Grupos del Prof. Antonio Ballesteros, el Dr. Francisco José Plou y el Dr. Miguel Alcalde. A todos y cada uno de mis compañeros del Instituto, Bárbara, Paloma, David, Patri, Berni, Nieves, Oscar, Eva, Diana…., y en especial darle las gracias a Lucía, por todos sus sabios consejos nada más llegar y por toda su ayuda.

Quiero dar las gracias de una manera muy emotiva a mis compañeros del Laboratorio 207, por cada día a día, por nuestras risas, nuestros bailes, nuestras peleas y sobre todo por hacerme todo este tiempo tan feliz, muchas gracias. Manolo gracias por todo, por poner humor en el día a día, Jesús por todas tus risas, Álvaro por todos los buenos ratos que hemos pasado juntos, por toda tu gran ayuda en los malos momentos y porque nunca dejaste de confiar en mí, Mónica por todos tus consejos y toda tu ayuda, Rafa por tu gran ayuda informática, a Ester por los ratos vividos, y sobre todo a mi amiga Mercedes, porque empezamos y terminamos juntas y sé que me llevo a una gran amiga. Muchas gracias a todos, os llevaré siempre conmigo.

A todas mis amigas de la "Uni", por hacerme feliz, Sandra, Mónica, Elena, Elenita, Patri, Gema y Milena. Y a las amigas de mi madre, "mis titas", porque siempre las tengo para lo que necesite.

Pero yo, mi vida y mi trabajo no tendrían sentido sin mi familia.

Gracias a mis abuelos. Mi abuelo que sigue creyendo que soy la mejor y mi abuela que allí donde esté, sé que estará muy orgullosa de mí porque siempre confiaba en que iba a llegar muy lejos en el colegio. A mi tía, Merche. A mi padre, porque aunque sea un poco cascarrabias, sé que siempre ha creído en mí y sabía que iba a conseguir todos mis sueños.

A mi marido, mi compañero en este viaje de la vida, por estar conmigo cada día y por querer seguir conmigo cada día formando nuestra familia, te quiero. A nuestro bebé, Carlos, que está en camino, porque espero que algún día pueda leer este trabajo y pueda sentirse muy orgulloso.

Pero sobre todo, me gustaría dedicarle todo, a mi madre. Mamá no tengo palabras, solo decirte que esto es de las dos. Gracias mamá.

Muchas gracias a todos, gracias por formar parte de mi vida y de mis sueños.

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS

Figura 10 | Mapa global de la localización de las principales fosas marinas profundas en el Mar Mediterráneo identificadas y estudiadas en esta Tesis Doctoral.

Figura 11 | A. Mapa global de la localización de la fosa marina *Medee*. Los paneles B y C muestran la superficie y los parámetros geoquímicos principales que caracterizan cada una de las interfases presentes en la fosa. Las interfaces se forman principalmente debido a las diferencias en la concentración de sal.

Figura 12 | Árbol filogenético de los grupos taxonómicos principales encontrados en la fosa marina *Medee* (Yakimov *et al.,* 2013).

Figura 13 | A. Mapa global de la localización de la dorsal Atlántica donde existen fumarolas hidrotermales donde habita *R. exoculata*.

Figura 14 | Mapa global de la localización de zonas marinas donde se han constatado vertidos de crudo.

Figura 15 | Modelos tridimensionales de la cuenca del lago de Arreo. Las líneas marrón y amarilla de trazo discontinuo representa el límite de la cuenca. La línea azul continua muestra el trazado del arroyo del Lago y la discontinua la de un pequeño cauce temporal. Las áreas verdes se corresponden con las formaciones palustres presentes en la cuenca y en azul con la lámina de agua del lago. Los números que aparecen en el mapa y en las fotos se corresponde con: Prao Segundo (1), Prao Tercero (2), manantial salino (3), lago de Arreo (4) y fuente del Hambre (5).

Figura 16 | Mecanismo de actuación de las enzimas esterasas-lipasas (adaptado de Stock *et al*., 2004).

Figura 17 | Mapa de color indicando los parámetros óptimos de actividad de las enzimas caracterizadas en la presente Tesis Doctoral. Los parámetros, en orden de aparición de izquierda a derecha, incluyen, pH, temperatura y concentración de sal (NaCl).

Figura 18 | Agrupación gráfica de la similitud y/o diferencias en la reactividad (A) y actividad específica (units mg^{-1}) (B) frente a un set de 137 ésteres diferentes de las diferentes esterasas caracterizadas en la presente Tesis Doctoral, así como otras enzimas similares comerciales (CalA, CalB, 100L, RM, RO, y PLE) y otras descritas de ambientes similares en la bibliografía (ABO 1197, ABO 1251, MGS01010, MGS 0105, y MGS0109). El panel A muestra la agrupación jerárquica en base a una distribución binomial de la presencia o ausencia de actividad para un stock de 137 sustratos, para cada una de las enzimas analizadas. Abreviaturas: CalA, lipasa A de *Candida antarctica* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); CalB, lipasa B de *Candida antarctica* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); 100L, lipasa de *Thermomyces lanuginosa* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); RM, lipasa de *Rhizomocur miehei* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); RO, lipasa de *Alcaligenes* sp. (suministrada por Meito Sangyo Co. (Japón); PLE, esterasa de hígado de cerdo.

Figura 19 | Relación filogenética de las esterasas identificadas en la presente Tesis Doctoral (negrita y sombra). Ver detalles en Alcaide *et al*. (2015). La figura muestra las sub-familias de la familia de esterasas/lipasas a las que pertenecen las diferentes enzimas (Arpigny y Jaeger, 1999).

Figura 20 | Actividad específica máxima (para el mejor de los sustratos) de las diferentes esterasas/lipasas caracterizadas en la presente Tesis Doctoral, así como otras enzimas de ambientes similares descritas en la bibliografía (ABO 1197, ABO 1251, MGS01010, MGS 0105, y MGS0109).

Figura 21 | Comparación de las estructuras de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Las enzimas se representan en viñetas donde se aprecia la localización de la serina catalítica (para estersas en la parte superior, y la enzima CCSP0084 en la parte inferior) y de los sitios de unión a NADPH/NADH y sustrato (para las enzimas MGS-M4 y MGS-M5 en la parte inferior).

Figura 22 | Comparación de los centros activos putativos de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Para cada enzima

se muestran los sitios de unión a sustrato y aquellos que participan en la catálisis. En el caso de enzimas con actividad esterasa (parte superior y enzima CCSP084 en la parte inferior) se representa la triada catalítica. El canal de unión a sustrato de la enzima MGS-M2 y su homólogo estructural HsaD se representan como representación de superficie accesible a solvente; el potencial electroestático se representa con colores, donde se aprecian las diferencias en residuos cargados. Se muestran también los residuos de unión a NADH/NADPH y de sustrato para la aldo-ceto reductasa MGS-M4 y la lactato dehidrogenasa MGS-M5, y sus homólogos estructurales.

Figura 23 | Comparación de la superfie de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Para cada enzima se muestran los sitios de unión a sustrato y aquellos que participan en la catálisis. La estructura aparece como representación de superficie accesible a solvente; el potencial electroestático (en unidades kb/T = constante de Boltzmann frente a la temperatura) se representa con colores, donde se aprecian las diferencias en residuos cargados. Las flechas indican la localización de la serina catalítica (para estersas en la parte superior, y la enzima CCSP0084 en la parte inferior) y de los sitios de unión a NADPH/NADH y sustrato (para las enzimas MGS-M4 y MGS-M5 en la parte inferior).

Figura 24 | A. Vista detallada de las estructuras de MGS-M1, MGS-M2 y CCSP084 superpuestas. Como se aprecia las tres enzimas presentan un plegamiento α/β hidrolasas, si bien la esterasa MGS-M1 no presenta un sub-domino α-hélice, que si presentan MGS-M2 y CCSP0084. Estrella negra: localización del centro activo. El panel (B) representa la estructura detallada de MGS-M1 donde se aprecian los elementos estructurales ausentes en comparación con las otras dos proteínas.

Figura 25 | Vista detallada de las estructuras de MGS-M2 y CCSP084 donde se detallan los residuos cercanos al centro activo. Las estrellas negras representan los residuos potenciales que participan en la catálisis o unión a sustrato. B. Detalle del canal de la estructura de MGS-M2 donde se aprecia la presencia de dos ligandos desconocidos, uno de ellos unido a la Ser97.

ÍNDICE DE TABLAS

Summary

This PhD Thesis employs metagenomic tools to identify new versatile enzymes, including esterases, lipases, glycosidases, aldo-ketoreductases and lactate dehydrogenases. With this Thesis we access to a wide enzyme diversity by using approaches based on the screening of activities of interest in clone libraries. The libraries were created from DNA originated from microbial communities of different origin, that are clearly different from those previously reported in the specialized literature. The main objective is to achieve through an intensive screen programm to a wide collection of enzymes and to provide an in deep understanding of the enzyme characteristics on the basis of environmental constraints, as well as of the mechanisms reponsible for enzyme adaptation and promiscuity. These general objectives have been achieved through: (1) functional screens in metagenomic libraries in samples with a high and new microbial diversity with model substrates to access to a high number of positive clones for activities of interest; (2) in silico screen, in the obtained sequences from positive clones, of those encoding enzymes of interest; and (3) creating a database containing multiple information that includes sequences, biochemical data and structural information. The selected enzymes have been biochemically characterized and those most promising were subjected to high density fermentation and crystallization.

In particular, in the PhD Thesis, we have identified and characterized 25 new enzymes (22 esterases/lipases, 1 beta-glucosidase, 1 aldo-keto reductase, and 1 (L)-lactate dehydrogenase) from metagenomic libraries created from DNA of microbial communities from 8 different habitats and the genome of one marine bacterium: *i*) 4 deep sea basins (*Medee*, *Kryos*, *Bannock* and *Matapan*) of the Eastern Mediterranean Sea (3 of which being hypersaline); *ii*) a karstic lake (Lake Arreo); *iii*) the microbiome of the epiotic bacteria from the gill chamber of the deep-sea shrimp *Rimicaris exoculata*, that lives close to deep-sea hydrothermal vents (2,320 m depth); *iv*) superficial seawater contaminated with crude oil in the Barents Sea (close to Kolguev Island); *v*) superficial seawater from a hydrothermal vent in Saint Paul Island (Alaska); and *vi*) the hydrocarbonoclastic marine bacterium *Cycloclasticus* sp. ME7. The temperature, salinity and depth of the investigated habitats range from 4-16.5ºC, 1.1-348 g/kg and 0-4,908 m, respectively. The low accessibility to some the investigated habitats together the unique geochemical constraints of some of them, make them examples of habitats not previously subjected to enzyme screen programmes. In addition to that, the differences in temperature, salinity and pressure within them and as compared to other previously investigated environments make them suitable model habitats to perform structural-functional studies and to investigate protein adaptation to poly-extreme conditions and analysis of enzyme properties such as enzyme promiscuity.

The average insert size of the libraries that were subjected to activity screen ranged from 120 to 816 Mbp. The incidence rate of positive clones containing activities of interest varied from 1:667 to 1:15,000. The molecular mass and isoelectric point of the investigated enzymes varied from 24,190 to 84,278 Da, and from 4.66 to 10.04, respectively. At the sequence level, the investigated enzymes presented a sequence homology as compared to known homologous ranging from 25% and 99%. The optimal temperature for activity range from 12 and 75ºC, and the optimal concentration of salt for activity range from 0 to 4.0 M. The majority of the enzymes were from bacterial origin of the phylum Proteobacteria (18), followed from those of Tenericutes (5) and Firmicutes (1); in one case, no unambiguous identification was possible.

The comparative study presented in this PhD Thesis has not only allowed us to provide the widest collection (25) and crystal structures (6) of enzymes from aquatic environments (including deep sea), reported to date, but also to provide a wide understanding of the enzyme reactivities and adaptations in a number of aquatic habitats. In particular, among the most important finding we should mention the identification and characterization of hydrolases with dual esterase:*meta*-cleavage product (MCP) hydrolase, as well as the presence of thermo-active and thermo-stable enzymes in deep-sea environments where the seawater temperature is never higher than 16.5ºC. In relation to this last point we should highlight that we have provided first experimental evidences that link the resistance of enzymes to high temperature and pressure in deep-sea hypersaline habitats. Moreover, the extensive analysis of substrate profiles tested with at least 210 substrates provided a deep understanding of promiscuity of the different enzymes investigated within them and as compared to other reported ones in the specialized literature. The results points to rare enzyme substrates profiles of some of the enzymes herein reported, some of which may have biotechnological potential. The data also suggests that enzymes from the same habitat may have similar reactivities as compared to those from other habitats.

Resumen

La presente tesis Doctoral emplea técnicas metagenómicas para la identificación de enzimas novedosas y versátiles, que incluyen esterasas/lipasas, glicosidasas, aldo-ceto reductasas y lactato dehidrogenasas. A través de esta Tesis se ha accedido a una amplia diversidad enzimática mediante el uso de estrategias de rastreo de actividad enzimática en librerias de clones. Dichas librerías se han creado a partir de DNA extraído de comunidades microbianas de diferente procedencia, que se distinguen claramente de otros trabajos realizados con anterioridad en la literatura científica. El objetivo principal es proporcionar a través de rastreos enzimáticos una amplia colección de enzimas, y un entendimiento detallado de las carácterísticas de las mismas en el marco de las condiciones geoquímicas que caracterizan los hábitats de procedencia, y los mecanismos subyacentes a la actividad de las enzimas. Estos objetivos ambiciosos se han conseguido mediante el empleo de: (1) rastreos funcionales de librerias metagenómicas empleando muestras de alta y muy diferente biodiversidad con sustratos modelo para acceder a un mayor número de clones positivos; (2) rastreos en las secuencias obtenidas de los clones positivos de secuencias correspondientes a las enzimas objeto de estudio; y (3) una amplia base de datos de secuencias, y datos bioquímicos y estructurales de las enzimas estudiadas. Las enzimas selecciondas han sido caracterizadas bioquímicamente y aquellas más prometedoras se sometieron a fermentaciones con alta densidad celular y técnicas de modelado y cristalización.

En particular, en la presente Tesis Doctoral se han identificado y caracterizado 25 nuevas enzimas (22 esterasas/lipasas, 1 beta-glucosidasa, 1 aldo-ceto reductasa, y 1 (L)-lactato dehidrogenasa) de metagenomas creados a partir de DNA de comunidades microbianas procedentes de 8 hábitats diferentes y un genoma de una bacteria cultivable. En particular de: *i*) cuatro fosas marinas (*Medee*, *Kryos*, *Bannock* y *Matapan*) del Este del Mar Mediterráneo (3 de ellas hipersalinas); *ii*) un lago cárstico (Lago Arreo); *iii*) el microbioma de agallas de una gamba (*Rimicaris exoculata*) que viene a 2,320 m de profundidad en la zona cercana a un fuente hidrotermal en la Dorsal Mesoatlántica; *iv*) agua marina superficial contaminada con crudo cercana a la Isla de Kolguev en el Mar de Barents; *v*) una fuente hidrotermal no profunda en la Isla de San Pablo (Alaska); y *vi*) de la bacteria marina hidrocarbonoclástica, *Cycloclasticus* sp. ME7. La temperatura, salinidad y profundidad de los hábitats estudiados oscila entre 4-16.5ºC, entre 1.1 y 348 g/kg y entre 0 y 4,908 m, respectivamente. Las características geoquímicas de estos hábitats los convierten en ejemplos de ambientes poco explorados a nivel enzimático y la amplia diversidad de factores ambientales, en particular, salinidad, temperatura y presión, los conviertes en hábitats adecuados para estudios estructura-función, adaptaciones a medios extremos y análisis de promiscuidad catalítica.

El tamaño medio de las librerías sometidas a rastreos funcionales oscila entre 120 y 816 Mbp, y el número de clones positivos para las actividades de interés por genoteca oscila entre 1:667 a 1:15,000. La masa molecular y el punto isoeléctrico de las enzimas estudiadas oscila entre 24,190 y 84,278 Da, y 4.66 y 10.04, respectivamente. A nivel de secuencia las enzimas identificadas y analizadas presentan homología a nivel de identidad entre el 25% y el 99%. La temperatura óptima de actividad oscila entre los 12 y los 75ºC y las concentraciones de sal (NaCl) para actividad óptima varían entre 0 y 4.0 M. La mayoría de las enzimas proceden de bacterias del filo Proteobacteria (18), seguido de Tenericutes (5) y Firmicutes (1); en uno de los casos, no fue posible sugerir la posible bacteria de origen.

El estudio comparativo de las enzimas identificadas no solo ha proporcionado la mayor colección (25) y mayor número de estructuras (6) de enzimas de ambientes acuáticos (incluídos ambientes marinos profundos), sino también un amplio conocimiento de nuevas reactividades y adaptaciones de las mismas. En particular conviene destacar la identificación y caracterización de enzimas con actividad dual esterasa: C-C hidrolasa, así como la presencia de enzimas termo-activas y termo-resistentes en ambientes marinos profundos donde la temperatura no es superior a 16.5ºC. Los datos presentados en esta Tesis Doctoral demuestran una relación directa entre la resistencia a la presión y un aumento en la temperatura óptima y de desnaturalización en enzimas de microorganismos aislados de ambientes marinos profundos hiper-salinos. Así mismo cabe mencionar que el análisis exhaustivo de las reactividades con un set de más de 210 sustratos ha proporcionado un amplio conocimiento de los niveles de promiscuidad de las enzimas de diferentes hábitats entre sí y con enzimas comerciales y descritas en la literatura científica. Los resultados presentados en esta Tesis Doctoral no solo apuntan a la existencia de especificidades inusuales de interés biotecnológico, sino también indican que las enzimas de un mismo hábitat presentan, por lo general, reactividades parecidas.

ABREVIATURAS/ACRÓNIMOS

ACR	Aldo-ceto reductasa
ADN	Ácido desoxiribonucleico
AZCL	Azo-carboximetil
AZCL-HE	Azo-carboximetil hidroxietil
CD	Dicroismo circular
CSIC	Consejo Superior de Investigaciones Científicas
EC	*Enzyme Commision*
EST	Esterasa
GH	Glicósido hidrolasa
GLY	Glicosidasa
HOHD	2-Hidroxi-6-oxohepta-2,4-dienoato
HOPHD	2-Hidroxi-6-oxo-6-fenilhexa-2,4-dienoato
ICP	Instituto de Catálisis y Petroleoquímica
KEGG	*Kyoto encyclopedia of genes and genomes*
LA	Lago Arreo
LAE	Esterasa del Lago Arreo
LDH	Lactato dehIdrogenasa
MCP	*C-C meta-cleavage product*
MeUmbG2	4-Metilumberil celobiosa
MGS	Metagenome sequence / Secuencia metagenómica
MTT	3-(4,5-Dimetil-tiazolil)-2,5-difenil bromuro de tetrazolio
α-NA	Acetato de α-naftilo
NADPH	Nicotinamida adenina dinucleótido fosfato
ORF	*Open Reading Frame* / Marco de Lectura
PET	Tereftalato de polietileno
*p*NP	*p*-nitrofenil
P valor	Valor estadístico de probabilidad
PET	Tereftalato de polietileno
Td	Temperatura de desnaturalización
X-Caprilato	Caprilato de 5-bromo-4-chloro-3-indolilo
X-Gal	5-Bromo-4-cloro-3-indolil-β-D-galactopiranósido

CAPÍTULO 1
INTRODUCCIÓN GENERAL

RESUMEN DE CONTENIDO

INTRODUCCIÓN GENERAL

1.1 Enzimas: interés creciente

"*La naturaleza no hace nada en vano*"….decía Aristóteles.

La vida en los ecosistemas evoluciona constantemente dando origen a una gran biodiversidad. Se estima que la biodiversidad microbiana asciende a unos 5-30 millones de especies y que la mayor parte de ella se encuentra en los mares y océanos (Mora *et al*., 2011). Si bien se desconoce el tamaño real de la misma, estudios recientes han revelado que solo se encuentran clasificadas hasta la fecha aproximadamente 11,000 especies de Bacterias y Arqueas, y cada año se describen al menos 600 nuevas especies (Yarza *et al*., 2014). Estos números proporcionan no solo una referencia del potencial ampliamente inexplorado de nuestro Planeta y de las comunidades microbianas que lo habitan, sino también del potencial de las mismas en lo que a nuevos genes y actividades se refiere (Kyrpides *et al*, 2014). Dichas actividades son ampliamente desconocidas, si bien su potencial para futuros desarrollos económicos se ha reconocido ampliamente (Jemli *et al*., 2014; Timmis *et al*., 2014; Velasco-Bucheli, *et al*., 2015).

Actualmente existe un desfase entre el número de enzimas microbianas teóricas que dicha biodiversidad alberga, y el número de enzimas caracterizadas experimentalmente: la proporción relativa se acerca a cero (Anton *et al*., 2013). Hay, por tanto, un creciente reconocimiento de que se debe abordar este problema (Bastard *et al*., 2014). En esta línea hay algunos ejemplos de iniciativas de investigación internacionales, como el proyecto COMBREX financiado por el *National Institute of Health* (EEUU) desde 2001, que pretenden la caracterización sistemática de enzimas de una docena de microorganismos de referencia, que incluyen bacterias modelo como *Escherichia coli* y *Helicobacter pylori*. El resultado de dichas iniciativas se ha traducido en un considerable incremento de las proteínas caracterizadas experimentalmente. Sin embargo, sólo el 0.33% de las proteínas de las bacterias *E. coli* y *H. pylori* han sido caracterizadas experimentalmente, lo que sugiere que solo una ínfima parte de las enzimas en el Planeta lo han sido también (Anton *et al*., 2013).

El conocimiento de la biodiversidad microbiana y enzimática a través de iniciativas de investigación públicas y privadas es de gran interés desde un punto de vista de investigación básica, pero también puede constituir la fuente de futuros recursos económicos. En este sentido, actualmente hay una enorme necesidad de biocatalizadores sostenibles (enzimas o microorganismos) con una alta eficiencia como una alternativa "verde" a la síntesis química (Adrio *et al*., 2003; Fernández-Arrojo *et al*., 2010; Blomberg *et al*., 2013; Turner *et al*., 2013; Vergne-Vaxelaire *et al*., 2013). Se estima que el 40% de los procesos de síntesis química se sustituirán por procesos de catálisis enzimática en 2020 (Adrio *et al*., 2010), ya que éstos eliminan la necesidad de disolventes orgánicos dañinos para el medioambiente y reducen el gasto energético. Por otro lado, nuestra civilización ha sobrepasado el máximo grado de extracción de crudo ("*peak oil*"), lo que implica la necesidad de buscar alternativas a los derivados fósiles (Timmis *et al*., 2014) mediante bio-procesos. Actualmente, se generan casi 5 billones de dólares por la aplicación directa de enzimas en diferentes mercados y se estima que la demanda mundial de enzimas se incrementará casi 6.4-6.7% en 2017 (www.rnrmarketresearch.com/world-enzymes-to-2017-market-report.html). Además, la evolución del precio que se paga por las enzimas, lo que se denomina "*share price evolution*", ha aumentado casi dos veces en los últimos 3 años (**Figura 1**). Estos datos demuestran que las enzimas y las soluciones biotecnológicas y renovables que de ellas se obtengan tendrán un mayor impacto en los próximos años.

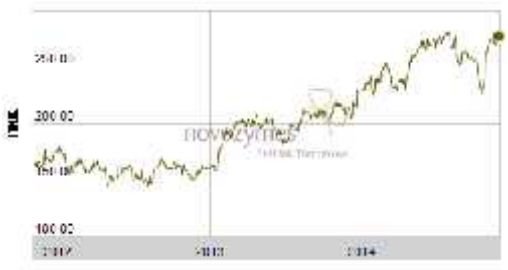


Figura 1 | Evolución del precio (en coronas danesas - DKK) que se paga por kilo de enzima. Estimaciones de la empresa líder en producción de enzimas Novozymes A/S (Bagsværd, Denmark). Figura cedida por Martin Simon Borchert (Novozymes A/S).

1.2 Biodiversidad enzimática: meta-genómica como herramienta de acceso

La primer pregunta que surge a la vista de lo expuesto anteriormente es ¿Cómo y de dónde pueden aislarse nuevas colecciones de enzimas? Inicialmente, éstas se extraían, aislaban o purificaban directamente de fuentes biológicas, principalmente microorganismos. Posteriormente, avances en biología molecular, bioquímica y microbiología permitieron optar por otras opciones más eficaces.

Resultados previos en la bibliografía, tanto a nivel teórico como experimental, han demostrado que es posible diseñar enzimas que no existen en la naturaleza o que muestran propiedades muy diferentes a las existentes. Así es posible diseñar enzimas con altos incrementos en la actividad y en la especificidad frente a un amplio número de moléculas (Reetz *et al*., 2010; Brustad y Arnold, 2011; Nobili *et al*., 2013; Liu *et al*., 2014; Sheng *et al*., 2014; Alcalde, 2015). Estas enzimas tienen secuencias, o modificaciones en las mismas, generadas por técnicas recombinantes, que causan cambios estructurales que producen una nueva enzima que contiene un centro activo específico para la reacción y el sustrato para el cual el método de rastreo se ha diseñado. En este proceso, conocido como ingeniería de proteínas, se pueden generar nuevas y /o mejores enzimas (Brustad y Arnold, 2011; Mate *et al*., 2010, 2013). En base a los resultados, el proceso se repite de forma iterativa bajo condiciones deseadas, para producir las variantes que cumplan los requisitos deseados. Los avances técnicos relacionados por los denominados sistemas *high-throughput screening* (sistemas de alto rendimiento) proporcionan una ventaja competitiva. Una revisión bibliográfica relacionada con la aplicación de estos métodos en estudios de ingeniería de proteína pone de manifiesto que la producción de librerías de mutantes es relativamente sencilla y posee un riesgo limitado, con un índice de éxito, es decir, un índice de encontrar un mutante mejorado respecto a las enzimas nativas en una librería de mutantes, que oscila entre 1:140 y 1:125,000 (**Figura 2**). Sin embargo, los estudios publicados indican que es posible generar nuevas enzimas, pero que la introducción de tales enzimas a nivel industrial es muy limitado (Davids *et al*., 2013; Koudelakova *et al*., 2013; Woodley *et al*., 2013).

El potential existente y reconocido de la microbiología ambiental para identificar nuevas enzimas se ha potenciado con la introducción de técnicas metagenómicas. Esta tecnología, a diferencia de la ingeniería de proteínas, tiene la capacidad de permitir el acceso a proteínas verdaderamente nuevas de microorganismos o comunidades microbianas sin la necesidad del cultivo individual de los mismos, que es técnicamente muy difícil (Fernández-Arrojo *et al*., 2010; Puspita *et al*., 2012).

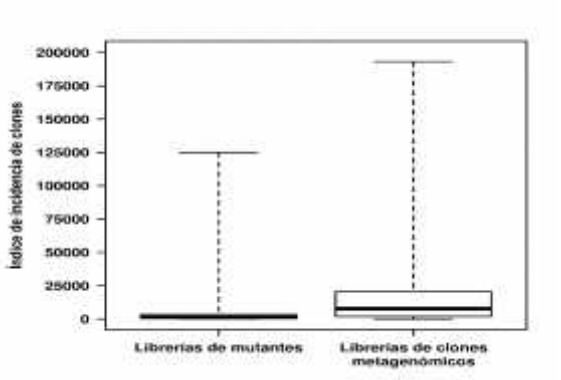


Figura 2 | Índice de incidencia de i) clones con actividad mejorada frente a clones con actividad nativa en experimentos de ingeniería de proteínas y ii) clones con actividad deseada en librería de clones metagenómicos. Los índices se basan en experimentos que emplean sistemas de rastreo de alto rendimiento y en los resultados de una amplia búsqueda de los trabajos publicados con ambas técnicas independientemente de la actividad objeto de estudio, en los últimos 25 años en la base de datos NCBI.

Más concretamente, es a partir de 1989 cuando la metagenómica entra en juego. La metagenómica es el estudio del conjunto de genomas de un determinado entorno, directamente a partir de muestras de ese ambiente, sin necesidad de aislar y cultivar esas especies. El auge de la metagenómica ha sido posible con la aparición de las tecnologías de alto rendimiento para secuenciar el DNA a bajo precio (Akondi y Lakshmi, 2013). En este sentido, el inicio del Proyecto Genoma Humano en 1990 marcó una revolución tecnológica. El desarrollo de éstas técnicas han permitido secuenciar una gran cantidad de genomas. Por ejemplo, a septiembre de 2014 en la base de datos GOLD (*The Genomes On Line Database"*; http://www.genomesonline.org/), aparecen 58,285 proyectos de secuenciación en marcha de los que se han completado 6,620 proyectos. En la **Figura 3** se aprecia una representación de la evolución de los proyectos de secuenciación masiva realizados en los últimos 7 años.

Posteriormente a la metagenómica se han desarrollado otras técnicas que nos permitían

completar el estudio de comunidades microbianas a nivel de microorganismos, genes y enzimas que éstas contienen. Estas técnicas se las denominan técnicas "ómicas" (Guazzaroni y Ferrer, 2011; Guazzaroni *et al*., 2012).

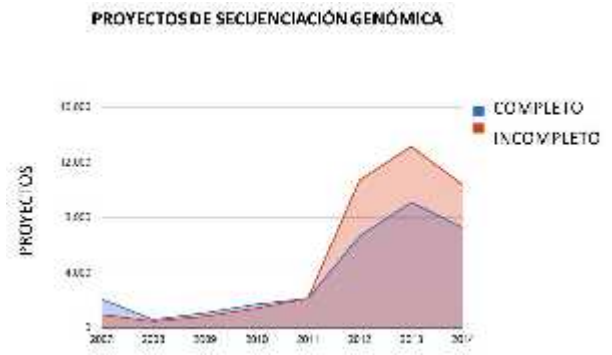


Figura 3 | Representación temporal de los proyectos de secuenciación masiva realizados o en vías de realización desde 2007.

La aparición de la metagenómica, como la genómica de los microorganismos incultivables, es una herramienta que ha supuesto una revolución en el estudio de comunidades de microorganismos directamente de su medio natural, ya que no hay la necesidad de aislar y crecer las especies individuales (Chistoserdova, 2014). Mientras que el genoma representa el material genético (DNA) de un organismo individual, el metagenoma es una técnica independiente del cultivo, lo que hace que se pueda aplicar teóricamente al estudio de cualquier tipo de muestra y nos ofrece la posibilidad de estudiar el DNA de una comunidad entera de organismos. El problema deriva de la alta complejidad de las muestras, es decir, mientras que un genoma estudia un individuo aislado, el metagenoma estudia una comunidad microbiana con miles de organismos diferentes y muchos de ellos desconocidos y además cada uno de ellos en diferente proporción aleatoria (Vieites *et al*., 2008).

En la **Figura 4** se puede observar un esquema de cuáles son los pasos a seguir para poder analizar el DNA procedente de una muestra ambiental. Se observan dos alternativas, en lo que a su uso para la identificación de nuevas enzimas se refiere: *i*) una donde se realiza un rastreo utilizando la secuencia obtenida por secuenciación masiva; y *ii*) otra donde se realiza un rastreo funcional del material genómico.

En la secuenciación directa (primera alternativa anteriormente mencionada), se obtiene como resultado un conjunto de tripletes de nucleótidos, llamados marcos de lectura abierta (ORF - *Only Read Frame*), cada uno de los cuales codifica un aminoácido que se encuentra entre un codón de iniciación que consta de tres nucleótidos (ATG) y un codón de terminación, que suele tener alguna de estas tres combinaciones de nucleótidos TAA, TAG o TGA. Del conjunto de secuencias obtenidas en un cualquier estudio se calcula que alrededor del 2-5% de los marcos de lectura abierta (ORF) en un genoma completo codifican enzimas de interés (Guazzaroni *et al*. 2014). Esto supone una proporción muy baja de enzimas con actividades de interés frente al número total de genes que codifican enzimas en los genomas microbianos (cerca del 40%) (Vieites *et al*., 2009). Una de las mayores limitaciones de los métodos de búsqueda de secuencias que codifican enzimas, es que las predicciones de funciones *in sílico* se basan en homología. Actualmente se están desarrollando herramientas bioinformáticas y de modelado a nivel de secuencia que permiten identificar secuencias que codifican enzimas de interés en los datos obtenidos por secuenciación masiva (Anton *et al*., 2013).

Dado que esta Tesis Doctoral no versa sobre este tipo de desarrollos bioinformáticos, nos centraremos a continuación en la descripción de las características y estado del arte de la denominada metagenómica funcional y los avances y desarrollos en la misma.

1.3 Problemas y posibles soluciones asociados a la metagenómica funcional

El rastreo de nuevas activades enzimáticas mediante la aplicación de técnicas metagenómicas ha supuesto avances significativos, principalmente en lo que se refiere a la búsqueda de actividades en ambientes extremos (Rothschild y Mancinelli, 2001; Feller, 2013; Alcaide *et al*., 2014). El establecimiento de colecciones de enzimas de genomas ambientales ya no es un problema. Como media es posible acceder sin grandes recursos técnicos a una media de 44 clones activos por cada librería de clones con una media de aproximadante 53,000 clones. Estos datos se han obtenido después de una extensa revisión bibliográfica de los trabajos publicados en los últimos 25 años, en esta área.

Sin embargo, existen una serie de problemas que limitan el uso de la metagenómica para la identificación de nuevas actividades enzimáticas y por ello se están realizando nuevos desarrollos que acorten el tiempo requirido que va desde la identificación de una enzima a su aplicación en

desarrollos tecnológicos (Jemli *et al*., 2014). Los mayores problemas tecnológicos, incluyen: (1) la baja proporción de genes presentes en el DNA metagenómico que codifica enzimas con actividades de interés (Guazzaroni *et al*., 2014); (2) la baja proporción de enzimas seleccionadas en condiciones requeridas en los procesos industriales (Martínez-Martínez *et al*., 2013); (3) la falta de sustratos industrialmente relevantes para los rastreos funcionales (Fernández-Arrojo *et al*., 2010); (4) la baja eficacia de los métodos de rastreo para actividades raras (Fraaije y Scrutton, 2013); (5) el reducido rendimiento de las enzimas aisladas mediante la aplicación de éstas técnicas, bajo condiciones no naturales (Fernández-Arrojo *et al*., 2010]; (6) el alto grado de enzimas identificadas que son inactivas después de la expresión en *Escherichia coli* (Loeschcke *et al*., 2013); (7) la falta de fuentes bioinformáticas fiables para el análisis de datos de secuenciación masiva (Nyyssönen *et al*., 2013); y (8) la falta de sistemas fiables de predicción de actividades enzimáticas de secuencias que codifican proteínas hipotéticas (Anton *et al*., 2013; Bastard *et al*., 2014).

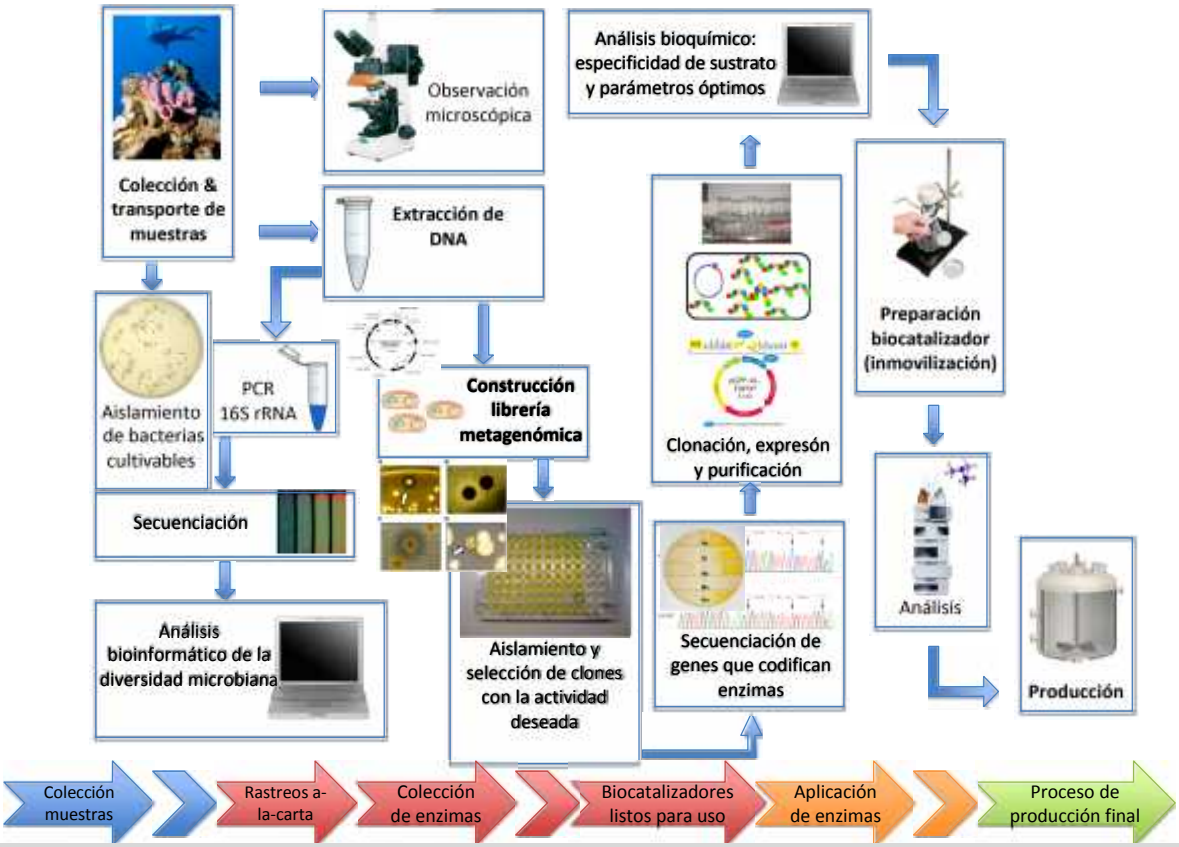


Figura 4 | Esquema con las aproximaciones y etapas más comunes para analizar el DNA procedente de una muestra ambiental. Las aproximaciones incluyen aquellas para el estudio por homología de secuencia y mediante rastreo funcional. Esta última aproximación ha sido empleada en la presente Tesis Doctoral.

Para solventar tales problemas se viene trabajando en los últimos años en una serie de soluciones, entre las que se incluyen: (1) un plan más selectivo de rastreo de actividades enzimáticas a partir de extensas librerías de clones y de selección de genes que codifican enzimas de interés (Yoon *et al*., 2014); (2) un pre-enriquecimiento bajo condiciones similares a las requeridas en los procesos o biotransformaciones a desarrollar (Kunze *et al*., 2013); (3) la selección de muestras ambientales donde la proporción de dichos genes sea más elevada, el aislamiento de cDNA y el análisis de la expresión de proteínas en enriquecimientos selectivos bajo condiciones y con sustratos de interés a fin de seleccionar las enzimas con mayores actividades específicas frente a sustratos de interés industrial (Akeroud *et al*., 2013; Chang *et al*., 2013); (4) priorizar la selección de enzimas con actividades múltiples, amplio rango de especificidad de sustrato y alta estabilidad en una amplio rango de condiciones (Alcaide *et al*., 2013), a fin de que la misma enzima puede ser empleada en múltiples procesos y bajo múltiples

condiciones; (5) la síntesis a la carta *de novo* de pequeñas moléculas o sustratos con funcionalidad similar a las requeridas a escala industrial con los que realizar los rastreos o enriquecimiento selectivos (Lim *et al*., 2013; Majdi *et al*., 2014); (6) el desarrollo de vectores a-la-carta y huéspedes para el rastreo y la expresión de genes que codifican actividades de interés (Valero, 2012; Loeschcke *et al*., 2013; Terrón-González *et al*., 2013; Furubayashi *et al*., 2014); (7) el diseño *in silico* o experimental mediate ingeniería de proteínas de las nuevas enzimas identificas por técnicas metagenómicas con el fin de producir las variantes biotecnológicas más favorables (Alcaide *et al*., 2013); (8) el desarrollo de un flujo de trabajo computacional para la identificación de genes en secuencias de DNA mediante la integración de métodos customizados bioinformáticos; y (9) el desarrollo de las denominadas técnicas "*Unknown BLAST*" que permitan el mapeado o búsqueda de ortólogos a enzimas con actividad desconocida (Anton *et al*., 2013).

1.4 El paradigma: racionalizando el proceso de identificación hacia la aplicación

Mientras que la producción de librerías de clones presenta un riesgo limitado, de acuerdo a los datos bibliográficos en los últimos 25 años, la baja incidencia de clones positivos si constituye un problema. Pese a ello, actualmente, mediante búsquedas por homología o por rastreo de clones es posible encontrar, en un par de meses, cientos de enzimas usando un solo sustrato. El cultivo o pre-enriquecimento bajo condiciones selectivas puede acelear significativamente la selección de enzimas de interés (Jacquiod *et al*., 2013; Verastegui *et al*., 2014). Por lo tanto, establecer una colección de enzimas por técnicas metagenómicas ya no constituye un problema. Además, es posible usar los genomas secuenciados de organismos cultivables o metagenomas secuenciados a un bajo precio para inspeccionar genes de enzimas que pueden ser clonados, caracterizados y expresados a escala piloto (Kube *et al*., 2013; Schallmey *et al*., 2014).

Este proceso de encontrar enzimas con marcado interés a nivel básico y aplicado es, sin embargo, un proceso largo y costoso que solo pocas enzimas pasan. Esto se ha demostrado para enzimas ampliamente usadas a nivel industrial, como esterasas y lipasas (Martínez-Martínez *et al*., 2013). Como se muestra en la **Figura 5**, una escala de tiempo de más de 7 años desde la identificación de un gen que codifica una enzima hasta el establecimiento de un proceso es la realidad (Fernández-Arrojo *et al*., 2010). Esto se debe a que las enzimas solo pueden aplicarse a escala industrial si cumplen una serie de criterios industriales: actividad frente a condiciones drásticas y muy diferentes a las naturales, alta selectividad y una alta velocidad de reacción (Singh, 2010).

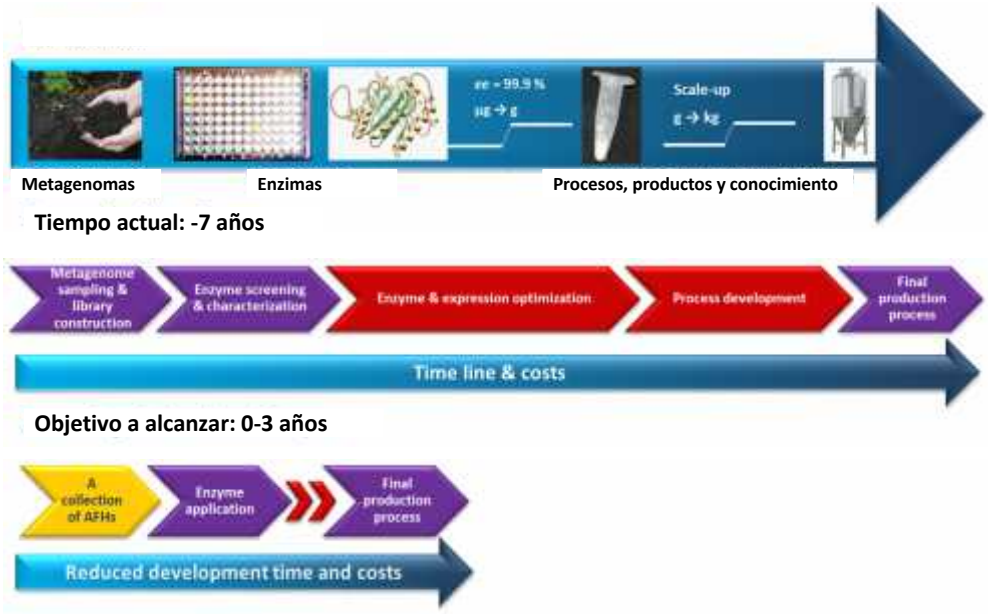


Figura 5 | Etapas requeridas para el desarrollo de procesos biotecnológicos basados en el uso de enzimas aisladas de comunidades microbianas. El panel superior indica la escala actual de tiempo requerido para realizar tales etapas, y la inferior, la predicción actual y futura, para que el desarrollo de procesos enzimáticos partiendo de una nueva enzima sea económicamente viable a escala industrial.

Pero ¿Cuáles son los pasos que producen tal retraso o escala de tiempo tan amplia? Desde mi punto de vista, los problemas mayores son los altos tiempo requeridos para la validación experimental de las actividades de las enzimas, que resulta en una tardía entrada en el mercado y altos costes (**Figura 5**). Solo muy pocas enzimas en las colecciones disponibles o enzimas descritas en la literatura (menos de un 5%) cumplen criterios requeridos a escala industrial (Martínez-Martínez *et al*., 2013). La baja promiscuidad (o bajo número de sustratos capaces de ser transformados eficazmente), la baja selectividad y la baja tolerancia a condiciones amplias de temperatura, etc., son problemas concretos y muy recurrentes. Otros problemas lo constituyen la baja actividad a bajos/altos pH y la baja tolerancia a altas concentraciones de sal y disolventes (Blomberg *et al*., 2013; Turner *et al*., 2013; Vergne-Vaxelaire *et al*., 2013). Pese a ello, la demanda de nuevas enzimas que cumplan criterios industriales es muy alta (Esteban-Torres *et al*., 2014; Pandya *et al*., 2014).

Las etapas de expresión y desarrollo del proceso constituyen los pasos más costosos en tiempo y en recursos económicos. De hecho, estas etapas constituyen en muchos casos más del 90–95 % de los costes de un desarrollo biotecnológico. Claramente, la presente Tesis Doctoral parte de la hipótesis de que "Para desarrollos de primera clase" con nuevas moléculas producidas por nuevas enzimas es crucial identificar e incorporar nuevas enzimas en el mercado". Para ello, solo el estudio detallado de nuevas y mejores enzimas puede reducir los tiempos y costes. Estimamos que un proceso de 3 años es actualmente el límite que se considera adecuado para tales desarollos y para que una nueva enzima pueda ser competitiva frente las existentes (**Figura 5**).

A continuación se resumen datos relacionados con el estado del arte en relación con los diferentes métodos empleados para la búsqueda de enzimas por técnicas metagenómicas y estimaciones de los avances que han supuesto estas técnicas en lo que al estudio de la biodiversidad en el Planeta se refiere.

1.5 Eficacia del rastreo: dependencia de la enzima a buscar, el sustrato, el método, el huésped y el vector a emplear

Una revisión bibliográfica de los estudios relacionados con la aplicación de métodos de metagenómica funcional pone de manifiesto que la producción de librerías de clones para la búsqueda de clones activos posee un riesgo limitado, usando cualquiera de los sustratos comerciales modelo y otros que se han diseñado a la carta a tal efecto. Como ejemplo, más de 200 sustratos diferentes se han empleado para la identificación de al menos 4,000 clones con actividad esterasa y lipasa en librerías de clones metagenómicos. La incidencia de los clones positivos, o sea, el número de clones positivos frente al número de clones totales rastreados, oscila entre 1:11 y 1:193,200 (**Figura 6, inserto**). Este número es similar a la incidencia de mutantes mejorados en experimentos de ingeniería de proteínas (**Figura 2**).

La incidencia (como medida de la frecuencia por la que un clon positivo aparece en una librería de clones) depende del tipo de enzima o actividad a seleccionar, además de la abundancia de los genes que codifican dicha actividad en los genomas microbianos presenten en la muestra analizada, y el nivel de actividad de los mismos. Así, clones que contienen enzimas poco activas pueden no ser dentificados en los tiempos comúnmente empleados en los ensayos colorimétricos, que oscilan en el margen de segundos a meses. Por ejemplo, para la búsqueda de actividad esterasa/lipasa con tributirina (1 volumen/ volumen (*v*/*v*)), los típicos tiempos de ensayo varían de 12 horas a 8 días (Jeong *et al*., 2011); para glicosidasas, los tiempos varían desde una escala de minutos (usando derivados de *p*-nitrofenol) a 30 días (para 5-bromo-4-cloro-3-indolil-β-D-galactopiranósido (X-Gal), derivados de caseina (p.e. AZCL-caseina), amilosa, curdlan, celulosa, xilano y galactomanano (Vester *et al*., 2014); para proteasas, los tiempos varían desde las 12 horas hasta las 4 semanas empleando leche desnatada (Niehaus *et al*., 2011); para fitasas, hasta 5 días usando fitato (Tan *et al*., 2014); y para las mono- y di-oxigenasas desde 2 horas (usando estireno/NADPH) a 3 días (para la producción de pigmentos como índigo) (Singh *et al*., 2010).

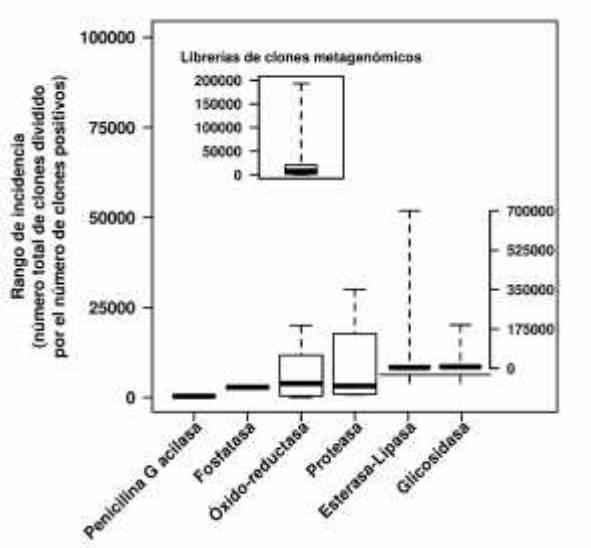


Figura 6 | Incidencia de clones positivos en librerías de clones para las seis actividades más estudiadas por técnicas metagenómicas. Los resultados se basan en valores medios para estudios metagenómicos, independientemente de los sustratos empleados. El inserto, representa la media de incidencia para todos los trabajos publicados.

A nivel de enzimas, considerando las 6 actividades más estudiadas a nivel metagenómico en la bibliografía, se aprecia que hay actividades más recurrentes que otras, es decir que pueden encontrarse en mayor proporción que otras en las librerías de clones. Se puede observar el siguiente orden en la incidencia: acilasas (1 clon positivo cada 333 clones [1:333]), fosfatasas (1:2,843), óxido-reductasas (1:6,670), proteasas (1:9,388), esterasas y lipasas (1:17,320) y glicosidasas (1:31,190) (**Figura 6**).

Otro de los factores que afectan a la eficacia de los rastreos es el sustrato a emplear. Así, en un reciente estudio, la incidencia variaba desde 1:188 a 1:3,937 y 1:15,625 cuando se empleada 1% (*v*/*v*) de tributirina, tricaprilina y trioleina, respectivamente, cuando se examinaba la misma librería de clones para actividad esterasa/lipasa (Glogauer *et al*., 2011) (**Figura 7A, inserto**). Esto sugiere que las enzimas con actividad lipasa son mucho menos abundantes (en este trabajo, 83-veces) que aquellas con actividad esterasa.

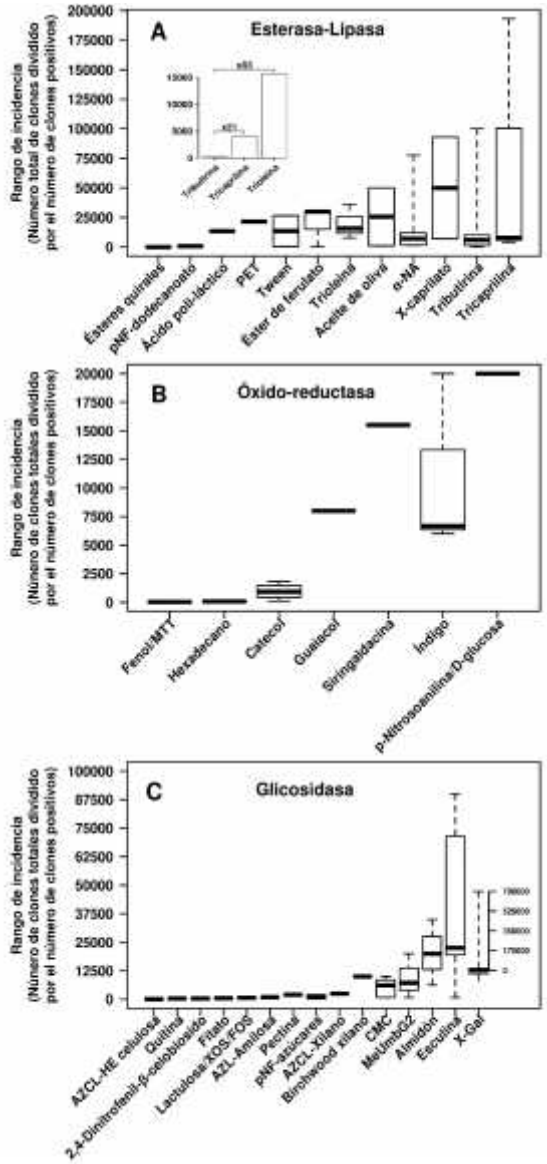


Figura 7 | Incidencia de clones positivos en librerías de clones para tres tipos de actividades empleando sustratos diferentes. A, actividad esterasa-lipasa; B, actividad óxido-reductasa; C, actividad glicosidasa. Abreviaturas: PET, tereftalato de polietileno; α-NA, acetato de α-naftilo; X-caprilato, caprilato de 5-bromo-4-chloro-3-indolilo; MTT, 3-(4,5-dimetil-tiazolil)-2,5-difenil bromuro de tetrazolio; AZCL, azo-carboximetil; AZCL-HE, azo-carboximetil hidroxietil; MeUmbG2, 4-metilumberil celobiosa; X-Gal, 5-bromo-4-cloro-3-indolil-β-D-galactopiranósido.

Entre los sustratos y métodos más empleados para el rastreo de actividad esterasa y lipasa, aquellos basados en el empleo de indicadores de pH resultan en una mayor incidencia de clones positivos (1:29), seguido de métodos basados en el uso de ácido poli-láctico (1:13,334), tributirina (1:15,478), acetato de α-naftilo (1:19,925), tereftalato de polietileno (1:21,400), trioleina/aceite de oliva y rhodamina B (1:22,061), *Tween*-20 (1:26,496), ferulato de metilo y etilo (1:26,496), caprilato de 5-bromo-4-

chloro-3-indolilo (1:50,000) y tricaprilina (1:68,279), en este orden (**Figura 7A**), por citar los sustratos más comunes y un tipo de actividad ampliamente estudiada.

Para las óxido-reductasas, entre los 7 sustratos más comunes empleados, el uso de fenol resulta en mayores niveles de incidencia (1:32), mientras que el empleo de *p*-nitrosoanilina complementada con D-glucosa para la búsqueda de actividad α-glucosa dehidrogenasa, se ha descrito como uno de los sustratos y/o actividades menos recurrentes (1:20,000) (**Figura 7B**). Finalmente, en el caso del rastreo de actividad glicosidasa, aproximadamente 15 sustratos diferentes han sido empleados con éxito, con derivados de azurina y hidroxietil celulosa (*azurine hydroxyethyl cellulose*), sustratos para la búsqueda selectiva de *endo*-celulasas, siendo los más eficaces (1:108); por el contrario, X-Gal, un sustrato común para la búsqueda de actividad β-galactosidasa, es uno de los peores sustratos (incidencia 1:142,900) (**Figura 7C**).

Los datos mencionados anteriormente sugieren que en los programas de búsqueda de actividades enzimáticas mediante la aplicación de técnicas de metagenómica funcional, tanto el sustrato a emplear como la actividad a rastrear pueden causar un sesgo importante en el número y diversidad de clones positivos y debe ser tenido en cuenta en cualquier experimento de este tipo. Independientemente de ello, una vez identificada la actividad a estudiar y el o los sustratos a emplear, la selección de método de búsqueda (*screening*) eficaz es de vital importancia. Así, los ensayos en placas de agar, donde la señal es visible al ojo humano, permite rastrear al menos 10,000 clones en cuestión de días, si bien presentan una sensibilidad limitada y se requiere una alta concentración del sustrato a emplear, que oscila entre 2.5 y 10,000 mg/l (Glogauer *et al*., 2001; Martínez-Martínez *et al*., 2014). Este problema se puede mitigar empleando ensayos colorimétricos o de fluorescencia que exhiben una sensibilidad mayor empleando menores concentraciones de sustrato (desde ng/l a pocos mg/l de sustrato) y que pueden emplearse en sistemas de alto rendimiento tales como micro-placa (Kourist *et al*., 2009; Martínez-Martínez *et al*., 2013). *Cell sorting* y las tecnologías de micro-fluidos permiten actualmente la selección de clones en librerías de 10^7–10^8 clones en un día de una manera cuantitativa (Kintses *et al*., 2012; Chang *et al*., 2013; Lim *et al*., 2013; Ma *et al*., 2013; Najah e*t al*., 12013; Hosokawa *et al*., 2014); en algunos casos, estas técnicas han permitido el rastreo de hasta 50,000 clones por segundo, o casi 1 billón de clones por día (Lim *et al*., 2013). En los últimos años se han desarrollado también las conocidas como trampas genéticas que permiten identificar actividad enzimática a unos niveles de sensibilidad muy altos ya que permiten identificar enzimas capaces de transformar solo 10^5 moléculas por célula (Qureshi, 2007; Phelan *et al*., 2012; van Rossum *et al*., 2013; Choi *et al*., 2014).

Los datos descritos anteriormente ponen de manifiesto las diferentes eficacias de los diferentes métodos comúnmente empleados. Dicho esto, los datos reportados en la bibliografía han demostrado que en muchos casos es mejor emplear sustratos modelo de amplio espectro, capaces de ser transformados por la mayoría de las enzimas descritas que se buscan, y luego someter los clones seleccionados a rastreos selectivos con la molécula o sustrato de interés. Esto se ha probado con éxito para la búsqueda de clones con actividad hidrolítica específica frente a ésteres de (*S*)-ketoprofeno (Yoon *et al*., 2007) o piretroides (Li *et al*., 2008), que se seleccionaron realizando primero un cribado con acetato de α-naftilo o X-caprilato, respectivamente.

El desarrollo de sistemas de rastreo se ha producido en paralelo al desarrollo de sistemas de expresión heteróloga, uno de los problemas principales de la metagenómica. Para solventar esto, se han descrito recientemente los denominados sistemas TREX para la expresión, transferencia y expresión de genes en múltiples huéspedes. El sistema se basa en que ambos extremos del vector donde se clona el DNA metagenómico, se marca en ambos extremos con las denominadas casetes L-TREX y R-TREX, que permiten transferencias por conjugación y la integración en el cromosoma de múltiples bacterias hospedadoras (Loeschcke *et al.*, 2013). También se han empleado componentes virales en los vectores de clonación de DNA ambiental para prevenir la terminación e incrementar la expresión de fragmentos de DNA de gran tamaño (Terrón-González *et al*., 2013). Dicho esto, los vectores más utilizados para la creación de librerías de clones a partir de DNA ambiental son cósmidos (p.e. pLAFR3), fósmidos (p.e. pCC1FOS) y Cromosomas Bacterianos Artificiales (BACs). También se trabaja actualmente en el diseño de nuevos vectores, como por ejemplo modificando genéticamente el vector comercial pCC1FOS dando lugar a otros vectores como pCC1FOS-CeuI, pMPO571 o pMPO579 (Terrón-González *et al.*, 2013). El

organismo hospedador más utilizado para la producción de estas genotecas es *Escherichia coli* aunque también pueden ser utilizados otros organismos heterólogos como *Streptomyces* (Gómez-Escribano *et al.*, 2012) o *Pseudomonas* (Martínez *et al.*, 2004), entre otros.

Actualmente existen en paralelo a los ensayos de selección y sistemas de clonación y expresión descritos antoriormente, herramientas bioinformáticas que permiten también hacer una preselección de enzimas que posteriormente se pueden estudiar en detalle (Nyyssönen *et al*., 2013). Así, las secuencias obtenidas por métodos de secuenciación masiva se pueden filtrar de acuerdo a su similitud con bancos de secuencias generalistas (UniProt, NCBI NR), bases de datos con dominios de secuencias conservadas (Pfam y Common Domains database (CDD)), o mediante bases de datos espeíficos que contienen secuencias e información bioquímica de determinados grupos de enzimas, tales como la base de datos *Carbohydrate-Active Enzyme* (Cantarel *et al*., 2009), que se usa para identificar glicosidasas así como esterasas y lipasas que actúan frente a ésteres de carbohidratos, y una bases de datos accesible desde 2011 para identificar lacasas (Sirim *et al.*, 2011). Métodos bioinformáticos de alto rendimiento se han empleado con éxito también para la búsqueda y rastreo en secuencias accesibles en las bases de datos o generadas mediante secuenciación masiva de secuencas que codifican epoxido hidrolasas y haloalcano dehalogenasas (Barth *et al*., 2004). Finalmente, herramientas de predicción tales como antiSMASH, se han generado recientemente para la búsqueda de genes que codifican enzimas para la síntesis de metabolitos secundarios o bio-activos tales como lactonas, bacteriocinas, sideroforos, ectonias, y policétidos (Blin *et al*., 2013, 2014). Más recientemente, Schallmey y colaboradores (Schallmey *et al*., 2014) han usado motivos conservados para identificar 37 nuevas halohidrin dehalogenasas, muy promiscuas, tras verificación experimental, examinando las bases de datos públicas. Mencionar que en una segunda etapa, para cada secuencia seleccionada es posible obtener datos generales tales como masa molecular y pK, así como la existencia de dominios conservados o estructuras que son específicos para la actividad deseada (Marsh *et al*., 2011).

1.6 Una mínima parte de nuestro Planeta ha sido explorada a nivel enzimático: ambientes marinos como ejemplo

Los datos e investigaciones recientes expuestas anteriormente demuestran que la metagenómica es una herramienta de gran potencial y en expansión. Se habla mucho de que con ella es posible llegar a entender el funcionamiento de comunidades microbianas y extraer nuevos microorganismos, genes y enzimas de cualquier hábitat de nuestro Planeta. Pero la realidad actual es muy diferente. A continuación resumiré que se conoce a nivel de hábitats explorados y enzimas obtenidas mediante la aplicación de técnicas metagenómicas.

En primer lugar, hay que recordar que las diferentes formas de vida que constituyen la diversidad microbiana descrita en la sección 1.1, han sabido adaptarse a las diferentes condiciones donde habitan. Pese a que los ecosistemas acuáticos constituyen las ¾ partes de la superficie terrestre (http://water.usgs.gov/edu/earthwhere water. html), éstos constituyen uno de los entornos naturales menos estudiados en lo que se refiere al aislamiento y caracterización de enzimas de las comunidades microbianas que lo habitan (De Long, 2005). Así, aunque el análisis de comunidades microbianas y el descubrimiento de nuevas enzimas ha progresado considerablemente (Martínez-Martínez *et al*., 2013), éste solo se ha realizado en un número muy limitado de hábitats. Una búsqueda exhaustiva en las bases de datos SCOPUS, PubMed, WOK y IMG/M (*US Department of Energy Joint Genome Institute*; http://www.jgi.doe.gov/) muestra que solo se ha estudiado el contenido genómico de las comunidades microbianas de aproximadamente 1,800 lugares del Planeta (**Figura 8**).

Hay que resaltar además que pese a que el 60% de ellos corresponden a ambientes marinos, el número de clones activos y/o enzimas analizas en los mismos es muy desfavorable a dichas muestras, en comparación con otros ambientes, como por ejemplo, ambientes terrestres. Así, mediante el uso de aproximaciones metagenómicas, aproximadamente solo 160 enzimas o clones con actividades de interés de microorganismos o comunidades microbianas marinas han sido descritas en la bibliografía, mientras que este número asciende a casi 4,000 en el caso de muestras de origen terrestre (**Figura 9**).

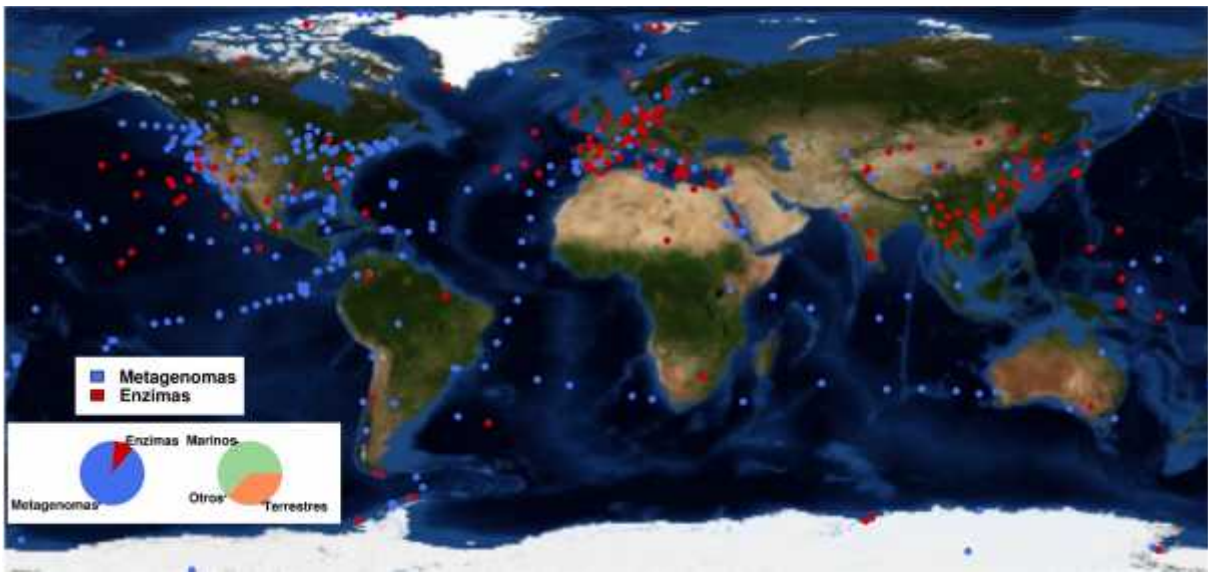


Figura 8 | Sondeo de muestras acuáticas (marinas y lagos) y terrestres distribuidas por el Planeta que han sido analizadas a nivel de secuenciación de DNA y taxonomía microbiana (en azul) y enzimático (en rojo) por técnicas metagenómicas. Como se observa en la figura, solo un número muy limitado de lugares ha sido analizado en lo que respecta al análisis de actividades enzimáticas. El mapa ha sido creado con programas de language R (2008) y el paquete OpenStreetMap (Eugster and Schlesinger, 2012) usando un mapa tipo "*mapquest-aerial*". La figura incluye estudios publicados (en los que constan las coordenadas GPS), usando las bases de datos SCOPUS, PubMed, WOK y IMG/M (http://www.jgi.doe.gov/). Como se muestra en la figura, se ha analizado el contenido genético y/o diversidad de comunidades microbianas de aproximadamente 1,800 lugares, y solo en 188 (o 11% del total) se han identificado y estudiado enzimas. Por lo tanto solo una pequeña fracción de los hábitats presentes en el Planeta ha sido sometida a intensos programas de búsqueda de actividades de interés.

Estos números reflejan el bajo nivel de estudio de las enzimas de microorganismos de ambientes acuáticos, incluidos hábitats marinos y lagos de baja salinidad. Esto nos lleva a una pregunta: ¿Es posible que las enzimas marinas tengan menos interés o menores niveles de actividad o potencial biotecnológico y por ello hayan sido sometidos a menos estudio? La respuesta es no, por múltiples razones. Primera, porque como quedará reflejado en la presente Tesis Doctoral las enzimas marinas poseen actividades y características únicas que las hacen de interés tanto a nivel básico como aplicado. Segunda, porque en la actualidad solo hay una empresa que comercializa enzimas marinas, *ArticZymes* (http://arcticzymes.com), y hay que conseguir aumentar el mercado de las mismas. Tercera, porque los hábitats marinos están caracterizados por condiciones geoquímicas muy diferentes y heterogéneas, no solo entre los diferentes hábitats que lo componen, sino también en relación a otro tipo de ambientes. En particular, dichas características incluyen diferentes niveles de profundidad, concentración de oxígeno, salinidad (http://water.usgs.gov/edu/earthwherewater.html), temperatura, presión, aporte de nutrientes y actividad antropogénica (por ejemplo, la presencia de vertidos de crudo) (Harrison *et al*., 2013). Dichas condiciones pueden influir en la distribución y diversidad de las poblaciones microbianas y también en las actividades y mecanismos de adaptación de las enzimas que éstas poseen.

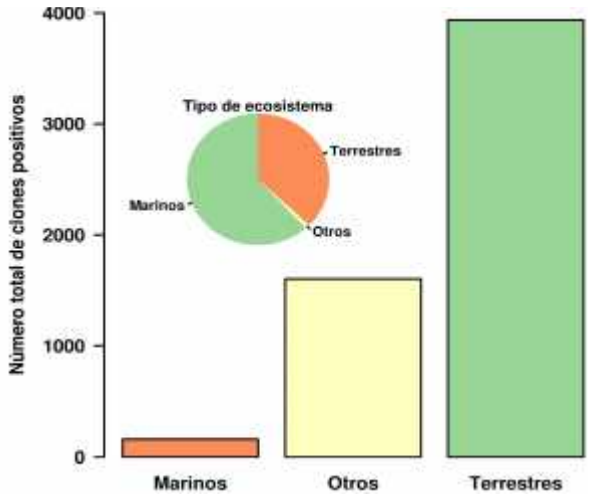


Figura 9 | Distribución del número de enzimas y/o clones activos identificados en estudios de metagenómica funcional empleando librerías de clones procedentes de comunidades microbianas de muestras de diferente procedencia. La figura insertada muestra el porcentaje relativo de lugares sometidos a rastreo de enzimas por técnicas metagenómicos.

A continuación se detallan algunos ambientes acuáticos caracterizados por propiedades geoquímicas muy diferentes que han sido explorados a nivel de rastreo de enzimas en la presente Tesis Doctoral y que pueden constituir la base de conocimiento básico y aplicado. Se trata de hábitats que han sido poco explorados hasta la fecha en lo que al aislamiento de enzimas de refiere y que han constituido la base de la presente Tesis Doctoral.

1.7 Fosas marinas profundas del Mar Mediterráneo: fuente inexplorada de poli-extremo-enzimas

El primer ejemplo, lo constituyen aquellos hábitats que han permitido el desarrollo de comunidades microbianas autóctonas que han sufrido pequeñas fluctuaciones ambientales: las fosas marinas creadas en el Mar Mediterráneo. La crisis salina del Messiniense (subdivisión comprendida entre los 7.2 y los 5.3 millones de años) fue un periodo en la historia geológica del Mediterráneo caracterizado por una acumulación masiva de sal en su fondo y que probablemente incluyó un periodo de desecación casi completa al quedar aislado el Mar Mediterráneo del Océano Atlántico. Como consecuencia de la ruptura del balance hídrico entre ambos se produjo la progresiva desecación del Mediterráneo (Hsü *et al*., 1973), apareciendo en su fondo enormes depósitos residuales de agua hipersalina. La estabilidad actual de estas formaciones viene marcada por un compromiso entre la topografía de la zona, su conveniente protección frente a las corrientes marinas y una permanente reposición de sal aportada por la disolución de los depósitos circundantes (Cita *et al*., 2006). Las primeras fosas se descubrieron en el año 1993, habiéndose descrito hasta la fecha al menos una docena de ellas. Algunos ejemplos de estos ambientes son las fosas *Hephaestus*, *Bannock*, *Tyro*, *Kryos*, *Discovery*, *L'Atalante*, *Urania* y *Medee* (MEDRIFF Consortium, 1995; Wallman *et al*., 1997; Westbrook *et al*., 2002; Cita *et al*., 2006; Daffonchio *et al*., 2006; Alexander *et al*., 2009; Borin *et al*., 2009; La Cono *et al*., 2011; Ferrer *et al*., 2012; Smedile *et al*., 2013; Yakimov *et al*., 2013). Las cuatro últimas se ubican en la llamada "región de las fosas hipersalinas anóxicas", área vecina a la dorsal mediterránea en la cual todavía se espera encontrar más de estas formaciones. La profundidad media de las mismas oscila aproximadamente entre los 2,150 y los 5,000 m.

La **Figura 10** muestra la localización de las principales fosas marinas identificadas en el Mar Mediterráneo, con un detalle de aquellas que se han estudiado en la presente Tesis Doctoral.

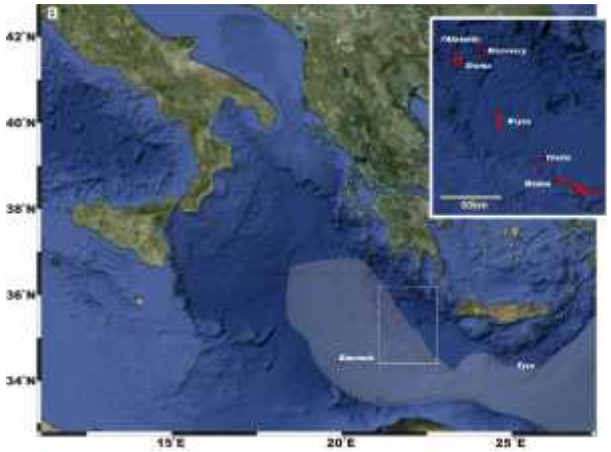


Figura 10 | Mapa global de la localización de las principales fosas marinas profundas en el Mar Mediterráneo identificadas y estudiadas en esta Tesis Doctoral.

Una de las primeras preguntas que puede surgir en relación a estas fosas marinas, es ¿Cuál es su tamaño? Sirva como ejemplo, la reciente fosa marina *Medee* (Yakimov *et al*., 2013) que presenta una superficie de 110 km^2 (la sexta parte de la superficie de la ciudad de Madrid) y un volumen de 9 km^3 (**Figura 11**).

Una segunda pregunta es ¿Cual es el nivel de salinidad en dichas fosas? La concentración de sal de las fosas marinas hipersalinas varía entre 283 (para *Urania*) y 487 (para *Discovery*) g/L, que es significativamente superior a la salinidad del agua de mar (38.5 g/L). Dicho esto cabe mencionar, que algunas de estas fosas presentan una salinidad similar a la del agua marina; es el caso de la fosa *Matapan* (Smedile *et al*., 2013), cuya salinidad es de 38.6 g/L. Como consecuencia de la mayor densidad de esta agua saturada estas fosas están separadas del resto de la masa salina por una fina pero muy estable capa que dificulta no solo los intercambios químicos, sino también, microbianos. Los factores principales que condicionan que la vida microbiana es posible en estos ambientes hipersalinos son la falta de oxígeno, la alta presión (>200 bares) y la temperatura estable de aproximadamente 14-16.5ºC (Filker *et al*., 2013). Hay que destacar, que la temperatura es una de las características más relevantes que caracterizan las profundidades del Mar Mediterráneo, ya que se trata del único lugar marino profundo con una temperatura de 14-

16.5ºC, mientras que en otros lugares está cercana a los 4ºC. Las características geoquímicos de las principales fosas se resumen en la **Tabla 1**.

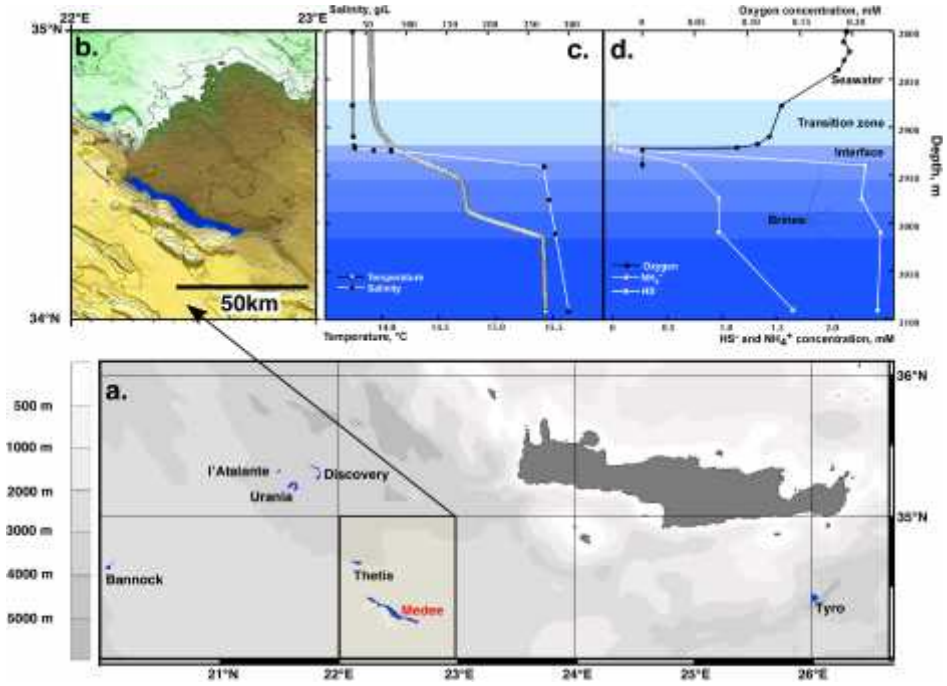


Figura 11 | A. Mapa global de la localización de la fosa marina *Medee*. Los paneles B y C muestran la superficie y los parámetros geoquímicos principales que caracterizan cada una de las interfases presentes en la fosa. Las interfaces se forman principalmente debido a las diferencias en la concentración de sal.

Fosa	Prof. (m)	Sal. (g/kg)	T (ºC)	pH	[O_2]
M	3010	347.80	15.1	6.87	0
K	3340	313.92	16.5	6.7	0
B	3342	263.20	14.5	6.55	0
TH	3258	348	14.5	-	0
U	3493	283	16.0	-	0
TY	3500	411	15.5	-	0.5
D	3581	480	15.0	-	0.5
MT	4908	38.6	14.3	8.12	5..0

Tabla 1 | Parámetros geoquímicos de las fosas marinas más representativas en el Mar Mediterráneo. Abreviaturas: Medee (M), Kryos (K), Bannock (B), Thetis (TH), Urania (U), Tyro (TY), Discovery (D) Matapan (MT), Prof. (Profundidad), Sal. (Salinidad), T (Temperatura), [O_2] (Concentración de oxígeno en mL/L).

Dada las condiciones de poli-extremofília (salinidad, temperatura, presión y concentración de O_2) que caracterizan las fosas marinas, solo unas pocas especies de anaerobios son capaces de soportar estas condiciones (Wallman *et al*., 1997; Westbrook *et al*., 2002; Cita *et al*., 2006; Daffonchio *et al*., 2006; Borin *et al*., 2009; Timmis *et al*., 2010; La Cono *et al*., 2011; Oren, 2011; Andrei *et al*., 2012; Ferrer *et al*., 2012; Smedile *et al*., 2013; Yakimov *et al*., 2013). Según sus características, las fosas no solo se diferencian a nivel geoquímico, sino también por gran variedad de diferentes microorganismos que las habitan, si bien las δ- y ε-Proteobacterias son las más abundantes (Borin *et al*., 2009). Pese a ello, existen Procariotas que se encuentran muy representados en algunas de ellas, en particular, las arqueas del grupo MSBL y bacterias del grupo KB1, muy abundantes en las fosas *Medee* (**Figura 12**) (Yakimov *et al*., 2013). En el caso de la fosa *Matapan* (Smedile *et al*., 2013), la más profunda (4,908 m) de las descubiertas hasta la fecha, podemos encontrar como bacteria mayoritaria *Alteromonas macleodii*, que participa en reacciones de fijación de CO_2 autótrofo y reacciones anapleróticas (Smedile *et al*., 2013). Se han encontrado también Eucariotas en la fosa *Thetis* (3,258 m de profundidad) (Stock *et al*., 2012).

Pese a que se ha empezado a conocer ampliamente la diversidad microbiana de estas fosas (Ferrer *et al*., 2012; Yakimov *et al*., 2013), pocas son las enzimas que se han estudiado en las mismas; de hecho solo se han caracterizado 5 esterasas de la fosa *Urania* (Ferrer *et al*., 2005). Esto se debe principalmente a su difícil acceso y a la falta de métodos que permitan extraer DNA de suficiente calidad. Dicho esto, una pregunta que puede surgir es: ¿Cuántas enzimas microbianas de ambientes profundos, incluidas las fosas marinas del Mar Mediterráneo, se conocen? ¿Qué se conoce de ellas a nivel estructural y a nivel bioquímico? Por otro lado, ¿Dado que las fosas marinas presentan características de poli-extremofilia (alta salinidad, alta presión y baja-media temperatura),

muestran sus enzimas propiedades y estructuras adaptadas a tales condiciones de forma diferente a las que presentan otras enzimas de otros ambientes extremos?

Obtener este tipo de información es de gran interés dada la importancia de definir el efecto combinado de valores extremos de salinidad, presión y temperatura en la diversidad y adaptación microbiana y la actividad y estructura de las proteínas. Sin embargo poco se conoce a nivel enzimático; una búsqueda exhaustiva en las bases de datos revela que la mayoría de las enzimas del fondo marino publicadas hasta la fecha han sido de ambientes fríos (4ºC) o ambientes hidrotermales ligeramente salinos (ver detalles en la Tabla S1 en Alcaide *et al*., 2014), junto con el previamente mencionado trabajo de la fosa Urania (Ferrer *et al*., 2005). Por otro lado, solo se han caracterizado estructuralmente 6 proteínas de organismos cultivables del fondo marino (De Vos *et al*., 2006; Shirai *et al*., 2008; Xu *et al*., 2008; Shin *et al*., 2009; Sineva and Davydov, 2010; Pietra, 2012); hay que mencionar que ninguna de estas enzimas ha sido aislada de hábitats que reflejen valores extremos de salinidad, presión y temperatura conjuntamente. Por lo tanto, la identificación y caracterización bioquímica y estructural de nuevas enzimas obtenidas de tales ambientes puede aportar nuevas informaciones en lo que al potencial básico y aplicado de las mismas se refiere.

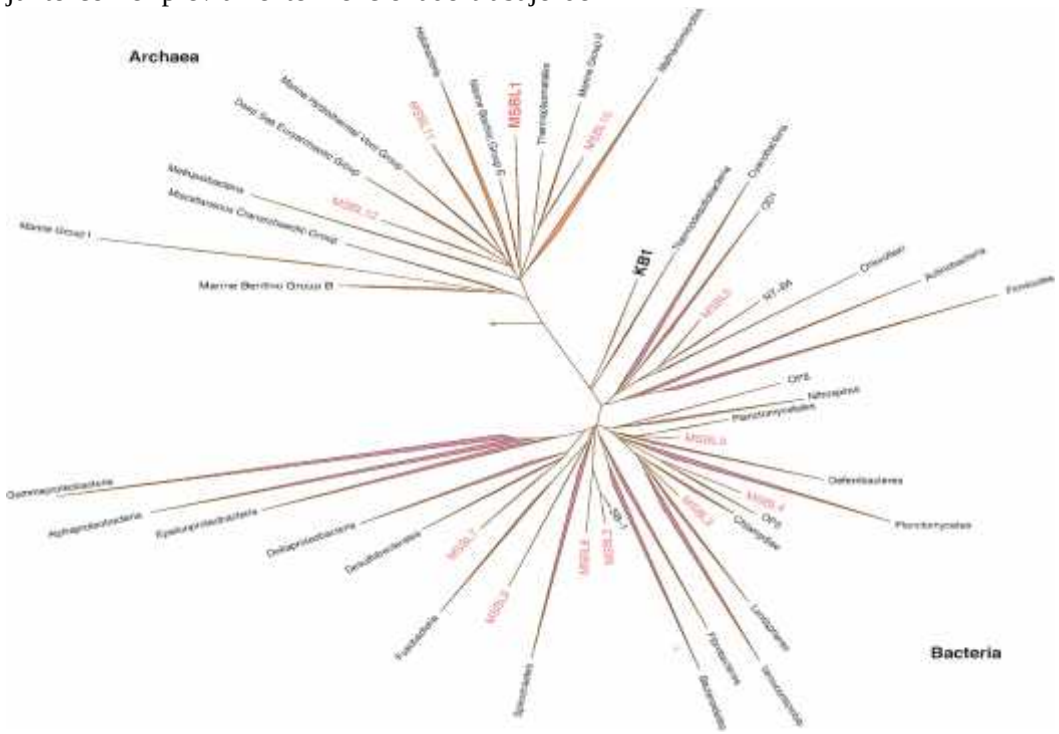


Figura 12 | Árbol filogenético de los grupos taxonómicos principales encontrados en la fosa marina *Medee* (Yakimov *et al.,* 2013).

1.8 Fumarolas hidrotermales: fuente inexplorada de extremo-enzimas de macro-organismos que la habitan

Un ejemplo de ambiente marino en el que las comunidades microbianas y los macro-organismos que la habitan han sido poco explorados lo constituyen las fumarolas hidrotermales. La primera fue descubierta en 1977 (Prieur *et al*., 1995) y se caracterizan por una temperatura por encima de los 350ºC en el origen de la zona de emisión de gases.

Una de las características más peculiares de estos ambientes es la presencia de una alta densidad de macro-organismos que son portadores de bacterias simbiontes y epibiontes, cuya diversidad ha comenzado a conocerse recientemente. Las bacterias simbiontes y epibiontes constituyen la base de la cadena alimentaria en los sistemas eco-quimio-sintéticos marinos y constituyen la base para la extracción de fuentes de energía. Las primeras bacterias simbiontes, identificadas en 1981, fueron las bacterias quicio-litótrofas del gusano *Riftia pachyptila* en el Pacífico; éstas bacterias son capaces usar compuestos de azufre como fuente energética. Posteriormente, en 1985 se descubrieron los primeros simbiontes que utilizan metano como fuente de energía; éstos se

descubrieron en los mejillones procedentes de filtrados de hidrocarburos en el Golfo de México (Petersen *et al*., 2011; Anderson *et al*., 2014).

La gamba *Rimicaris exoculata* (Williams y Rona, 1986) domina la fauna en muchas fumarolas hidrotermales de la dorsal Atlántica (<2,320 m de profundidad). *R. exoculata* vive en zonas cercanas a las chimeneas en el gradiente entre los fluidos hidrotermales y el agua marina, que tiene una temperatura entre 3 y 25ºC (media de 8.7ºC). La salinidad media, debido a las características especiales de estos ambientes, es inferior a la del agua de mar y en el caso particular de la dorsal Atlántica (**Figura 13**) se sitúa alrededor de 23 g/l de sal. Estudios recientes han demostrado que *R. exoculata* posee una densa comunidad bacteriana en sus agallas (Schmidt *et al*., 2008). Éstas contribuyen a la utilización de compuestos reducidos de azufre, hidrogeno y metano (Schmidt *et al*., 2008; Petersen *et al*., 2011), lo que proporciona energía y nutrientes al huésped, incapaz de metabolizar tales compuestos. Así mismo, dichos micro-organismos son capaces de metabolizar compuestos inorgánicos y orgánicos procedentes de la columna de agua o del fluido hidrotermal (nitratos, fosfatos, sulfatos, bicarbonato, vitaminas, aminoácidos, policarbonatos, etc. (Charlaou *et al*., 2002; Zbinden *et al*., 2008; Hügler *et al*., 2011; Konn *et al*., 2011; Guri *et al*., 2012; Wahl *et al*., 2012; Ponsard *et al*., 2013).

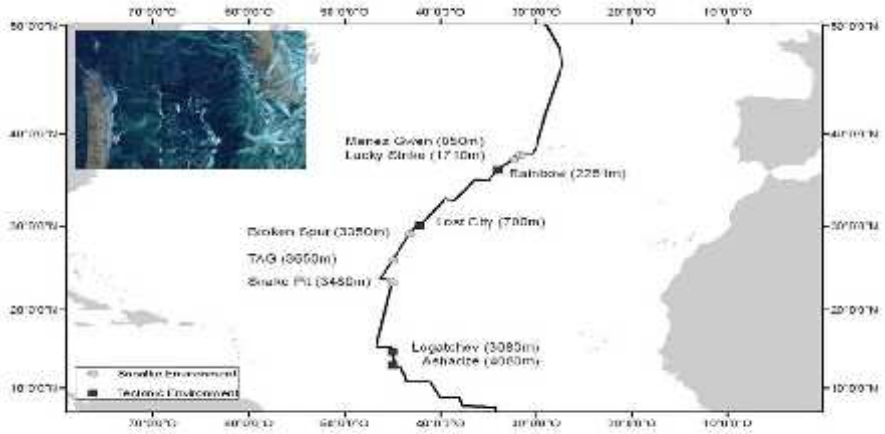


Figura 13 | A. Mapa global de la localización de la dorsal Atlántica donde existen fumarolas hidroterma les donde habita *R. exoculata*.

Aunque el fondo marino es un hábitat para macro-organismos y las comunidades microbianas asociadas a los mismos, hasta la fecha en que esta Tesis Doctoral fue escrita, no se habían descrito enzimas de microorganismos (simbiontes o epibiontes) asociadas a macro-organismos de ambientes profundos, incluidos aquellos que habitan fumarolas marinas. De forma minoritaria solo se ha descrito un pequeño número de enzimas bacterianas asociadas a esponjas y moluscos marinos (Karpushova *et al*., 2005; Okamura *et al*., 2010; Selvin *et al*., 2012; Kim *et al*., 2011; Sim *et al*., 2012), ninguno de los cuales procede de ambientes profundos (profundidad inferior a 200 m). Entre las enzimas caracterizadas se encuentran 3 esterasas de esponjas marinas (Karpushova *et al*., 2005; Okamura *et al*., 2010; Selvin *et al*., 2012), y una celulasa y una alginato liasa (Kim *et al*., 2011; Sim *et al*., 2012) de moluscos marinos. La mayoría de estas enzimas presentan una homología a nivel de secuencia de aminoácidos del 54-100% a secuencias de enzimas homólogas descritas en las bases de datos, son halo-tolerantes y son más activas a pH cercano a la neutralidad y a 40ºC. Mientras que una esterasa asociada a la esponja *Hyrtios erecta* retiene más del 60% de su actividad en el rango 30-50ºC (Okamura *et al*., 2010), una lipasa de la esponja *Haliclona simulans* es capaz de mantener tales niveles de actividad a 10-80ºC (Selvin *et al*., 2012). La baja información bioquímica de éstas y otras enzimas asociadas a macro-organismos de ambientes profundos, limita nuestro conocimiento sobre las características y adaptaciones ambientales y el potencial biotecnológico de las mismas

1.9 Contaminación antropogénica en ambientes marinos: ¿fuente de promiscuidad enzimática?

La influencia antropogénica ha marcado claramente los ambientes acuáticos, en particular, los marinos. Así las aguas pueden verse sometidas a constantes vertidos que hacen que los microorganismos que habitan dichos

ambientes cambien sus fuentes de carbono y se enfrenten a especies químicas a las que antes no se habían enfrentado (Dafoncchio *et al*., 2013). Hasta qué punto las enzimas de dichos organismos evolucionan para poder asimilar dichos nutrientes o moléculas y cuáles son las bases moleculares y estructurales que lo favorecen, está aún por esclarecer a nivel enzimático.

Los hidrocarburos en los sistemas marinos pueden proceder de filtraciones naturales de petróleo y de depósitos de gas natural, además de de accidentes de transporte de petróleo y deliberadas descargas o accidentes como los ocurridos en el Golfo de Mexico (Frodie *et al.*, 2011; Gonzalez-Lozano *et al.*, 2010), y en la costa gallega (Laffon *et al.*, 2014). La **Figura 14** muestra la localización de las zonas marinas donde se han producido vertidos. Como se puede apreciar en la figura, existen miles de lugares donde se han reportado vertidos accidentales de crudo. Algunas de las zonas donde más vertidos se han producido se localizan en el Mar Mediterráneo, Mar del Norte y zonas contiguas.

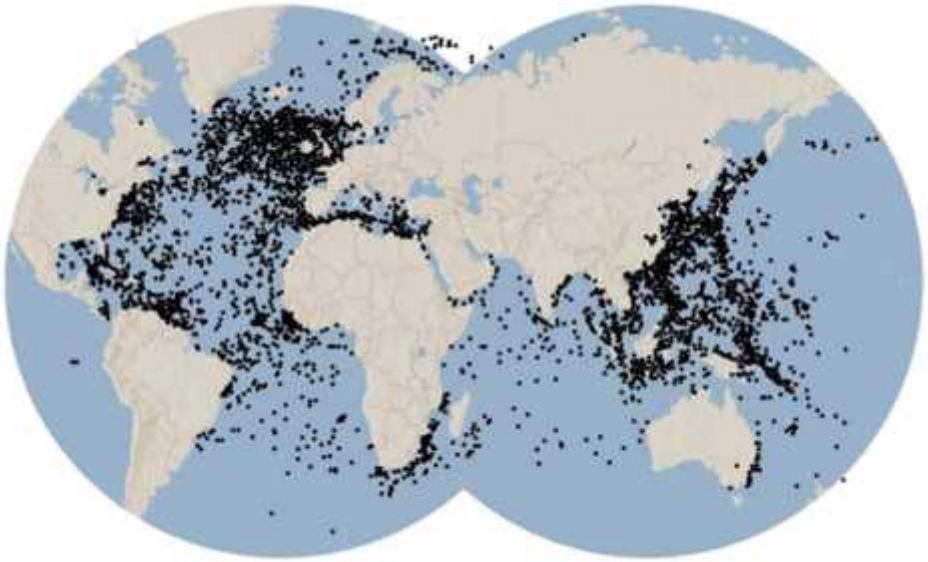


Figura 14 | Mapa global de la localización de zonas marinas donde se han constatado vertidos de crudo.

Hasta hace poco tiempo no se conocían verdaderos microorganismos marinos capaces de degradar hidrocarburos. A diferencia de sus homólogos terrestres que son muy versátiles, los microorganismos marinos degradadores han demostrado ser muy especializados. Hasta la fecha, la mayoría de estos microorganismos pertenecen a sólo 19 géneros de bacterias, entre las que cabe mencionar *Alcanivorax*, *Marinobacter*, *Thallassolituus*, *Cycloclasticus*, y *Oleispira* (Yakimov *et al.*, 2007), entre los más conocidos. Estudios recientes han revelado que, pese a su importancia en la biodegradación natural de crudo, poco se sabe sobre el arsenal enzimático de éstos u otros microorganismos que pueblan ambientes contaminados de forma crónica. Esto contrasta con el hecho de que investigaciones recientes han demostrado que las enzimas (p.e. esterasas) de estas bacterias presentan actividades específicas y una especificidad de sustrato superiores a las de enzimas modelo comerciales (Kube *et al*., 2013; Tchigvintsev *et al*., 2014). Así mismo, pese a que la mayoría de estos organismos están adaptados a una baja temperatura (0-25ºC), presentan un arsenal de enzimas muy versátil ya que pueden actuar eficazmente en un rango que oscila entre los 4 y los 65ºC (Kube *et al*., 2013; Tchigvintsev *et al*., 2014).

Hasta qué punto la introducción en el medio ambiente de especies químicas diferentes a las presentes de forma natural (por ejemplo, moléculas presentes en el crudo), pueden alterar o modular las actividades enzimáticas presentes en comunidades microbianas autóctonas, es un aspecto del que poco se conoce. Es éste un aspecto que se ha tratado de investigar con éxito en la presente Tesis Doctoral.

1.10 Lagos salinos poco profundos: hábitats accesibles pero poco explorados

A diferencia de las fosas marinas del Mar Mediterráneo que han permanecido inalteradas durante millones de años, hay ambientes que han sufrido modificaciones en una escala geológica

menor, en lo que a variaciones en la salinidad se refiere. Un ejemplo lo constituyen los lagos cársticos, que son lagos que se alojan en una depresión cárstica (dolina, uvala, poljé), que se ha producido por la disolución de las calizas. En la mayoría de estos lagos tanto la alimentación de los ríos afluentes como el río emisario, se hace gracias a ríos subterráneos. Un ejemplo es el Lago Arreo (**Figura 15**), con una temperatura que fluctúa entre 4.7 y 19.8 ºC según los meses del año y con una profundidad máxima de 24.8 m. Este Lago, localizado al noroeste de España, contiene propiedades de halofília moderada (<1 g/L) (González-Mozo *et al.*, 2000; Corella *et al.*, 2011) y constituye el lago cárstico más profundo en Europa. Otro ejemplo lo constituye la Laguna de Carrizo en Madrid (Ferrer *et al.*, 2011).

¿Qué se sabe de las enzimas de este tipo de ambientes? La respuesta es que si bien se conoce la diversidad microbiana y eucariota en los mismos (González-Mozo *et al.*, 2000; Corella *et al.*, 2011; Ferrer *et al.*, 2011), no se han descrito apenas enzimas de estos ambientes (Martínez-Martínez *et al*., 2013). Hasta qué punto la presencia de bajas concentraciones de sal y la baja-moderada temperatura del agua son factores que han permitido la evolución de determinadas características enzimáticas, está por investigar. Este es también un aspecto que se ha tratado de investigar con éxito en la presente Tesis Doctoral.

1.11 Esterasas y lipasas: un caso de estudio práctico

Lo expuesto en las secciones anteriores resume la limitada información a nivel de enzimas en un amplio número de hábitats de nuestro Planeta, y en particular, en ambientes acuáticos. Dado que un objetivo principal en los grupos de investigación es encontrar enzimas con características singulares que sean de interés en estudios básicos de estructura-función y en biotecnología es necesario profundizar en el análisis de la bioquímica y las estructuras de nuevas enzimas en hábitats únicos, incluidos aquellos sometidos a condiciones extremas, ya sea por sus condiciones físico/químicas ambientales o por su situación geográfica. Claramente, estos ambientes pueden albergar una biodiversidad enzimática que en algún caso puede ser muy diferente a la conocida hasta la fecha.

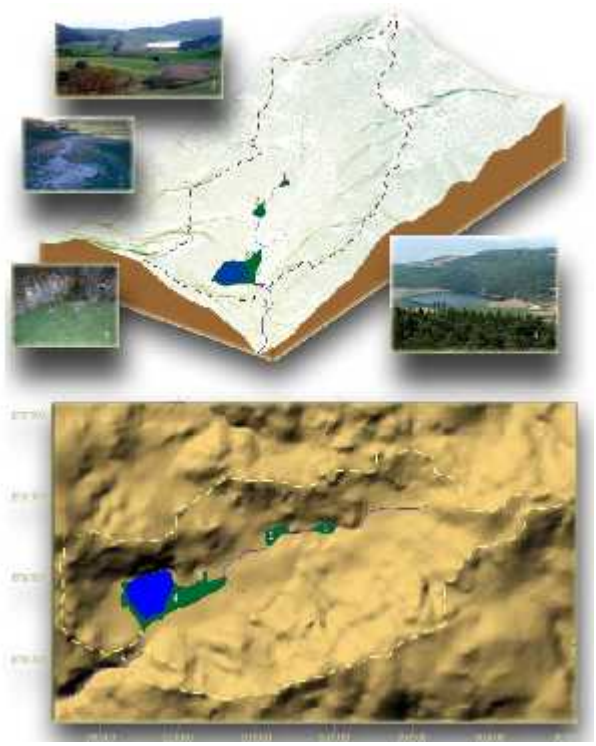


Figura 15 | Modelos tridimensionales de la cuenca del lago de Arreo. Las líneas marrón y amarilla de trazo discontinuo representa el límite de la cuenca. La línea azul continua muestra el trazado del arroyo del Lago y la discontinua la de un pequeño cauce temporal. Las áreas verdes se corresponden con las formaciones palustres presentes en la cuenca y en azul con la lámina de agua del lago. Los números que aparecen en el mapa y en las fotos se corresponde con: Prao Segundo (1), Prao Tercero (2), manantial salino (3), lago de Arreo (4) y fuente del Hambre (5).

Una vez que se ha concretado el primer objetivo, concretamente identificar nuevas enzimas de hábitats poco explorados, el siguiente objetivo es concretar el tipo de enzimas a estudiar. Un grupo de enzimas de amplio interés lo constituyen las enzimas esterasa-lipasa, un grupo de enzimas que se encuentran ampliamente distribuidas en la naturaleza, tanto en animales y plantas como en microorganismos. Independientemente de su naturaleza, todas ellas tienen una actividad común: catalizan la hidrólisis de los enlaces ésteres formados entre un ácido y un alcohol (Ro *et al*., 2004). Las lipasas son específicas para acilgliceroles con ácidos grasos de cadena larga (>10 átomos de carbono), siendo la trioleína su sustrato de referencia, mientras que las esterasas actúan específicamente sobre acilgliceroles de cadena corta (<10 átomos de carbono), y otros ésteres simples, siendo la tributirina su sustrato estándar. Las lipasas muestran preferencia por sustratos altamente hidrófobos, insolubles y agregados,

presentando sitios de unión a sustrato largos para poder acomodar al ácido graso (Reyes-Duarte *et al*., 2005). Las esterasas, por el contrario, actúan sobre sustratos más solubles y con un grado de hidrofobicidad más variable. Además las lipasas muestran un número mayor de aminoácidos no polares localizados en las zonas expuestas al solvente y en la región del centro activo en comparación con las esterasas, así como un rango de sustrato más amplio y estéreo-especificidad y regio-selectividad mayor (Fojan *et al.*, 2000). Pese a estas diferencias, ambas no requieren cofactores, suelen ser estables y son activas en disolventes orgánicos, lo que las convierte en biocatalizadores de amplio espectro y de interés industrial.

Arpigny y Jaeger (1999) realizaron una clasificación, dividiendo las esterasas-lipasas descritas en 8 familias de lipasas bacterianas en función de la similitud de secuencias de aminoácidos y algunas propiedades bioquímicas. Esta clasificación ha venido actualizándose debido a la identificación de nuevas enzimas y secuencias (Hausmann y Jaeger 2010; Kourist *et al*., 2010).

Estructuralmente las esterasas-lipasas suelen presentar un plegamiento típico de las α/β hidrolasas, y contienen mayoritariamente una triada catalítica constituida por Ser-Asp-His (Arping y Jaeger 1999; Hausmann y Jaeger 2010; Kourist *et al*., 2010. Algunas lipasas presentan cambios (Arping y Jaeger 1999), tales como: i) la sustitución de Glu en lugar de Asp; ii) la sustitución del motivo Gly-x-Ser-x-Gly cerca de la Serina del centro activo, por la secuencia consenso GDSL (Gly-Asp-Ser-Leu); y iii) y que la triada Ser-Asp-His en algunos casos es sustituida por una diada catalítica Ser-His. Pese a estos cambios, el mecanismo para la formación o hidrólisis de los enlaces éster es esencialmente el mismo para lipasas y esterasas y consiste en cuatro pasos (Navarro-González y Pariego, 2012) que se resumen en la **Figura 16**: (1) en primer lugar se une el sustrato a la enzima, después se produce el ataque nucleofílico por parte del grupo hidroxilo de la serina catalítica sobre el enlace éster del sustrato. Este ataque nucleofílico inicial forma un tetraedro intermedio que es estabilizado por la presencia de dos glicinas cercanas al centro activo, formando el centro oxianiónico; (2) este tetraedro es rápidamente descompuesto, con la ayuda de la histidina protonada, lo que provoca la liberación del alcohol del enlace ester y la formación del complejo acil-enzima; (3) a continuación, el complejo acil-enzima es atacado por una molécula de agua, con la ayuda de la histidina, formándose un segundo tetraedro intermedio; y (4) finalmente después de la formación del tetraedro intermedio, se produce la liberación del ácido graso y como consecuencia de ello la regeneración del centro activo (generando una nueva enzima libre).

Las esterasas y lipasas son el tercer gran grupo de enzimas comerciales, después de las proteasas y de las carbohidrasas, y son consideradas entre uno de los principales grupos de enzimas que producen más beneficios. Entre las aplicaciones tecnológicas más importantes destacan las siguientes (Plou *et al*., 2003; Navarro-González *et al.*, 2012):

1. Industria farmacéutica y química (Muñoz-Solano *et al*., 2012): ejemplos de su utilización en procesos químicos son la hidrólisis enzimática de ésteres para la obtención de intermedios en la síntesis de paroxentina, que es un potente antidepresivo (Guisan *et al.*, 2002), la síntesis de antibióticos β-lactámicos (Sánchez-Ferrer *et al.*, 2007), y la síntesis de ésteres de azúcares con actividad antitumoral (Ferrer *et al*., 1999).
2. Industria alimentaria: ejemplos de su utilización en la industria panadera son la feruloil esterasa de *Streptomyces thermocarboxydus* empleada como coadyuvante tecnológico en los procesos de panificación de masas de harina (Copa *et al.*, 2006); ejemplos en la industria de los aditivos alimentarios lo constituyen la reacción para la liberación de los ésteres de carotenoides se emplean esterasas y lipasas procedentes de *Pseudomonas fluorescens* y *Pseudomonas alcalígenes* (Ralf *et al.*, 2007; Harrison *et al.*, 2000); aplicaciones en la industria láctea son aquellas derivadas de su capacidad de hidrólisis ya que las lipasas juegan un papel importante en la preparación de saborizantes con sabor a queso (Silvert *et al.*, 2007; Esteban-Torres *et al*., 2014) y en la industria de grasas y aceites, ya que, por ejemplo las grasas con menor valor comercial pueden ser convertidas en grasas con un mayor valor comercial (Sharma *et al.*, 2001). Recientemente las lipasas se han empleado con éxito en la producción de antioxidantes fenólicos (Torres *et al*.,2010, 2012; Medina *et al*., 2010).
3. Industrial de detergentes y agentes limpiadores: las esterasas y lipasas pueden ser empleadas como aditivos ya que son activas a altas temperaturas y pH alcalino;

además son esenciales en la producción de jabón, lavavajillas, disolventes de limpieza en seco y limpieza de lentes de contacto.

La presente Tesis Doctoral se ha centrado en el estudio de esterasas y lipasas. Como se ha comentado anteriormente, éstas han sido seleccionadas por varias razones. Primero, porque con enzimas ampliamente estudiadas en la bibliografía y existe un amplio número de datos bioquímicos y estructurales que nos permitirán estudios comparativos. Segundo, porque son pocas las enzimas de este tipo que han sido estudiadas en hábitats acuáticos, y la identificación y caracterización bioquímica y estructural, puede ayudar a entender el potencial básico y aplicado de las mismas en el contexto de la información para enzimas de otros ambientes. En este sentido, hasta la fecha en la que la presente Tesis Doctoral fue escrita no se había descrito hasta la fecha la estructura de ninguna enzima con actividad esterasas y/o lipasa en ambientes marinos profundos. Tercero, porque en el marco de estudios anteriores y los presentados en esta Tesis Doctoral son enzimas que pueden ser usadas como marcadores para el estudio de mecanismos de adaptación de las proteínas a condiciones de poli-extremofília. Cuarto, porque constituyen uno de los biocatalizadores de mayor proyección biotecnológica (Jaeger y Eggert, 2002; Turner y Truppo, 2013).

Pese a que esterasas y lipasas constituyen la base principal objeto de estudio en la presente Tesis Doctoral, a modo comparativo se han estudiado otras enzimas, en particular, glicosidasas, aldo-ceto reductasas y lactato dehidrogenasas, que igualmente han sido poco estudiadas en los ambientes investigados en la presente Tesis Doctoral. Aunque conocemos los avances que se están desarrollando en el diseño de mejores erramientas de clonación y expresión de DNA metagenómico y sistemas de rastreo (*screening*) que permitan aumentar la eficacia de los sitemas de búsqueda de enzimas en librerías de clones o secuencias obtenidas por secuenciación masiva, la presente Tesis Doctoral emplea herramientas y ensayos comúnmente usados, ya que entendemos que con ellos es posible abarcar una amplia diversidad enzimática.

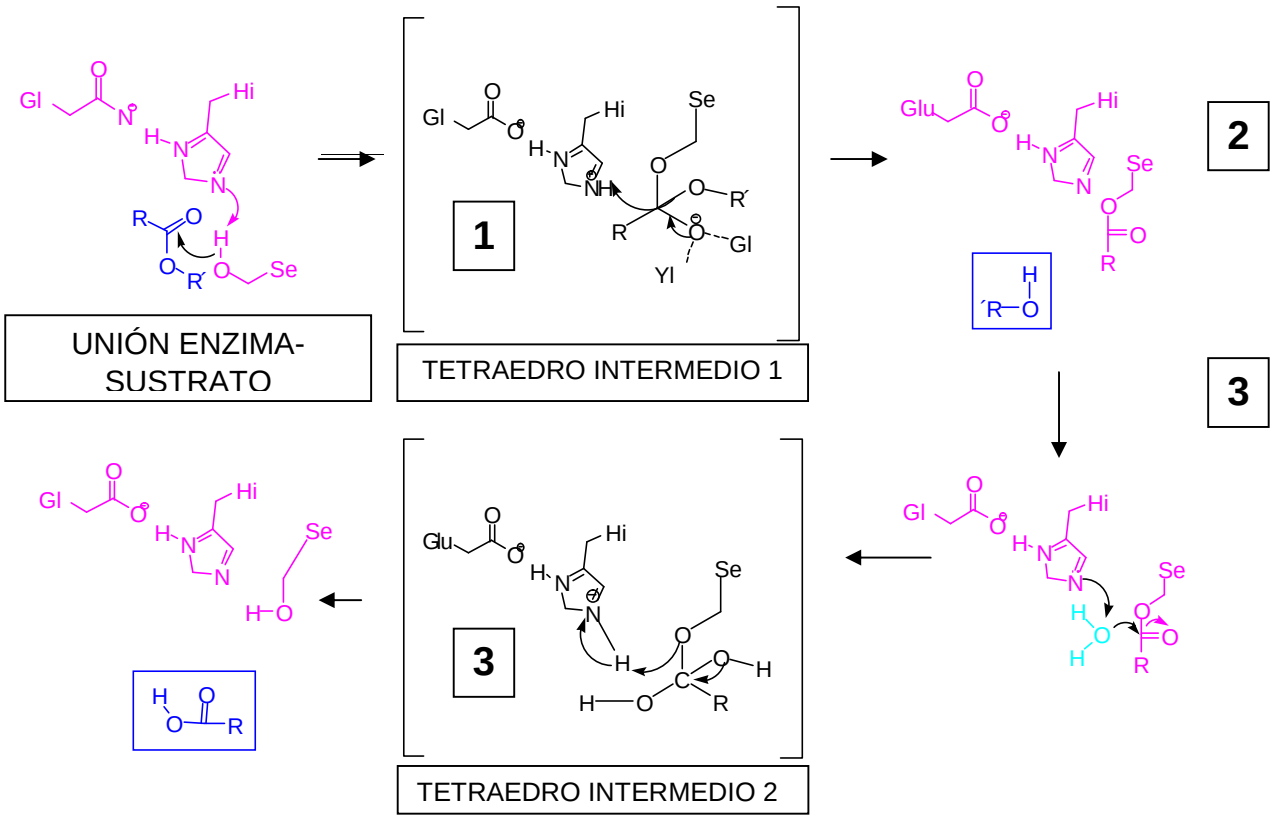


Figura 16 | Mecanismo de actuación de las enzimas esterasas-lipasas (adaptado de Stock *et al.*, 2004).

FINALIDAD Y OBJETIVOS

El Objetivo general de la presente Tesis Doctoral es acceder a una **amplia diversidad enzimática**, mediante el desarrollo y aplicación de métodos metagenómicos. Se pretende con ello proporcionar una **amplia colección de enzimas**, y un **entendimiento detallado de los mecanismos** subyacentes a las características bioquímicas y físico-químicas de las mismas en el marco de factores estructurales y ambientales. Estos objetivos ambiciosos se conseguirán mediante los siguientes sub-objetivos específicos:

• **Objetivo 1**. Obtener una amplia colección de enzimas de comunidades microbianas autóctonas de ecosistemas de diferente procedencia, en particular aquellos poco estudiados hasta la fecha, para garantizar el acceso a una amplia biodiversidad. Se emplearán **técnicas metagenómicas** para el rastreo funcional, en librerías de clones creadas a partir de DNA microbiano procedente de muestras de origen marino y cárstico, de enzimas con alto potencial biotecnológico, que incluyen **esterasas y lipasas,** como objetivo principal, así como otras enzimas de interés como **glicosidasas, aldo-ceto reductasas y lactato dehidrogenasas**.

• **Objetivo 2**. Generar una amplia base de datos de secuencias y datos bioquímicos de las enzimas aisladas. Para ello tras los rastreos funcionales se procederá a la secuenciación de los fragmentos de DNA que contienen las enzimas de interés y, tras identificar las secuencias que codifican dichas enzimas, se procederá a la **clonación** de los genes correspondientes, y a la **expresión, purificación y caracterización** de las enzimas objeto de estudio.

• **Objetivo 3**. Proporcionar un **entendimiento detallado** de las especificidades de sustrato y características físico-químicas de las enzimas objeto de estudio, y generar una amplia colección de **datos estructurales de las enzimas** con mayor relevancia a nivel de actividad y secuencia. Se pretende con ello sugerir los factores estructurales responsables de las correspondientes características enzimáticas y el posible efecto de las condiciones geoquímicas de los hábitats de procedencia sobre la promiscuidad y mecanismos de adaptación. Para ello, las enzimas más prometedoras se someterán a fermentaciones con alta densidad celular y técnicas de cristalización de proteínas, junto con técnicas de mutagénesis.

CAPÍTULO 2
ESTERASAS DEL LAGO ARREO

RESUMEN DEL CONTENIDO

En este primer capítulo presentamos la identificación y el análisis bioquímico de siete enzimas con actividad esterasa/lipasa obtenidas mediante rastreos funcionales en una libreria metagenómica de comunidades microbianas presentes en una muestra de sedimentos del Lago de Arreo. Se trata de un lago de origen cárstico localizado en el norte de España (42º46'N, 2º59'W) y a 655 m de altitud, que presenta una temperatura media que fluctúa entre los 4.7 y 19.8 ºC, según los meses del año, y una profundidad máxima de 24.8 m. El Lago de Arreo es uno de los pocos lagos cársticos en Europa de mayor profundidad que se han desarrollado en formaciones de yeso.

Para ello se han utilizado técnicas metagenómicas, incluidas rastreos de actividad en librerías de fósmidos (pCCFOS), secuenciación masiva del DNA de los clones positivos y técnicas bioinformáticas para la identificación de marcos de lectura abiertas que codifican las enzimas de interés. Posteriormente se han empleado técnicas de biología molecular para la expresión en la bacteria *Escherichia coli* de los genes que codifican las enzimas de interés, así como la posterior producción a pequeña escala y purificación con la ayuda de colas de 6-histidinas. El empleo de técnicas bioquímicas, incluidos ensayos espectrofotométricos para la determinación de parámetros óptimos, especificidad de sustrato y cálculo de actividades específicas y constantes cinéticas, permitió posteriormente la caracterización de las mismas. Finalmente se han realizado estudios comparativos a nivel de secuencia de aminoácidos y de frecuencia de tetra-nucleótidos para determinar no solo la similitud y novedad de las secuencias obtenidas sino también el posible origen microbiano de las mismas.

Tras rastrear con un sustrato modelo (acetato de naftilo) un total de 11,520 clones (342 Mpb) se obtuvieron 10 clones positivos. Esto supone una incidencia de 1 clon positivo por cada 1,152 clones. Tras la secuenciación del DNA contenido en los mismos se identificaron 378 genes que codificaban proteínas, de los que 7 eran genes que codificaban presuntas esterasas de la super-familia de las α/β-hidrolasas. La masa molecular y el punto isoeléctrico de las enzimas estudiadas oscila entre 24,329 y 34,003 Da, y 4.78 y 10.04, respectivamente. A nivel de secuencia las enzimas identificadas y analizadas presentan homologías a nivel de identidad entre el 25% y el 54%. El pH óptimo varía entre 8.0 y 8.5, la temperatura óptima de actividad oscila entre los 12 y los 40ºC y las concentraciones de sal (NaCl) para actividad óptima varían entre 0 y 1.2 M. Las 7 enzimas proceden de bacterias del filo Proteobacteria, en particular de los géneros *Cupriavidus*/*Ralstonia* (2), *Methylobacterium* (1), *Rhizobium* (1), *Azotobacter*/*Azospirillum* (1), *Sphingomonas* (1) y *Burkholderia* (1). Estos datos demuestran la presencia en el Lago de Arreo de proteobacterias que contienen enzimas moderadamente halófilas y que están adaptados a bajas temperaturas. El análisis de la especificidad de sustrato, testeada sobre un conjunto de 101 ésteres estructuralmente diversos, demostró posteriormente perfiles de reactividad muy diferentes entre sí y en comparación con otras 6 preparaciones enzimáticas comerciales con actividad esterasa y lipasa. En particular, una de las enzimas (LAE6) presenta reactividades y características superiores a las de la mejor enzima en su clase usada en procesos industriales (CalB) y es ampliamente promiscua, siendo capaz de transformar eficazmente triglicéridos, ésteres de ácidos grasos halogenados y no halogenados, ésteres aromáticos, ésteres de carbohidratos, lactonas y epóxidos quirales. Este estudio proporciona evidencias experimentales de que los e metagenomas bacterianos de ambientes cársticos pueden ser de un gran interés potencial para la identificación de enzimas de interés biotecnológico, ya que son tolerantes a la sal, son activas a bajas temperaturas y muestran una alta promiscuidad y enantioselectividad.

El CD adjunto a esta Memoria contiene Material Suplementario relacionado con este Capítulo.

AEM
Journals.ASM.org

Biochemical Diversity of Carboxyl Esterases and Lipases from Lake Arreo (Spain): a Metagenomic Approach

Mónica Martínez-Martínez,[a] María Alcaide,[a] Anatoli Tchigvintsev,[b] Oleg Reva,[c] Julio Polaina,[d] Rafael Bargiela,[a] María-Eugenia Guazzaroni,[a]* Álvaro Chicote,[e] Albert Canet,[f] Francisco Valero,[f] Eugenio Rico Eguizabal,[g] María del Carmen Guerrero,[g] Alexander F. Yakunin,[b] Manuel Ferrer[a]

Spanish National Research Council (CSIC), Department of Applied Biocatalysis, Institute of Catalysis, Madrid, Spain[a]; Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada[b]; Department of Biochemistry, University of Pretoria, Pretoria, South Africa[c]; CSIC, Instituto de Agroquímica y Tecnología de Alimentos, Paterna, Valencia, Spain[d]; Universidad de Castilla-La Mancha, Ciudad Real, Spain[e]; Departament d'Enginyeria Química, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain[f]; Department of Ecology, Faculty of Science, Autonomous University of Madrid, Madrid, Spain[g]

The esterases and lipases from the α/β hydrolase superfamily exhibit an enormous sequence diversity, fold plasticity, and activities. Here, we present the comprehensive sequence and biochemical analyses of seven distinct esterases and lipases from the metagenome of Lake Arreo, an evaporite karstic lake in Spain (42°46′N, 2°59′W; altitude, 655 m). Together with oligonucleotide usage patterns and BLASTP analysis, our study of esterases/lipases mined from Lake Arreo suggests that its sediment contains moderately halophilic and cold-adapted proteobacteria containing DNA fragments of distantly related plasmids or chromosomal genomic islands of plasmid and phage origins. This metagenome encodes esterases/lipases with broad substrate profiles (tested over a set of 101 structurally diverse esters) and habitat-specific characteristics, as they exhibit maximal activity at alkaline pH (8.0 to 8.5) and temperature of 16 to 40°C, and they are stimulated (1.5 to 2.2 times) by chloride ions (0.1 to 1.2 M), reflecting an adaptation to environmental conditions. Our work provides further insights into the potential significance of the Lake Arreo esterases/lipases for biotechnology processes (i.e., production of enantiomers and sugar esters), because these enzymes are salt tolerant and are active at low temperatures and against a broad range of substrates. As an example, the ability of a single protein to hydrolyze triacylglycerols, (non)halogenated alkyl and aryl esters, cinnamoyl and carbohydrate esters, lactones, and chiral epoxides to a similar extent was demonstrated.

Esterases and lipases from the α/β hydrolase family have received considerable attention, because they are widely distributed within the microbial communities operating in most of environments where they have important physiological functions (1) and because they are one of the most important groups of biocatalysts for biotechnological applications (2–4). Upon searching the list of genes using Pfam (the protein family database [5]) from the approximately 140 metagenomic projects in various stages of sequencing on the GOLD website (Genomes OnLine Database; http://www.genomesonline.org/) and the available sequences of esterases and lipases, more than 72,000 predicted esterases/lipases of the α/β hydrolase superfamily were retrieved, which revealed the richness of uncultured biodiversity (6), to provide wide collections of such biocatalysts. This is one of the largest protein families with available sequences. In relation to the cultivation-independent methods used to identify them, it should be highlighted that sequence-based metagenomics only provide the presumptive compositional and functional blueprint represented in the community genome (7, 8), but at the same time, this method causes serious problems regarding both sequencing errors (9) and the erroneous assignment of substrate specificity (10). In contrast, the activity-directed techniques have been shown to provide a direct view of known or new protein families and functionalities (for examples, see references 11–22, and 23). Whatever the platform used (i.e., either gene cloning from sequence resources or naïve activity screens), in many cases, enzyme properties have been shown to be modulated by ecosystem characteristics (24, 25).

From the whole set of sequences available, only a limited number of metagenomically derived esterases and lipases have been experimentally characterized (see Table S1 in the supplemental material), and only 27% of them (or approximately 60) have been shown to possess characteristics likely needed for industrial operations. In this respect, useful features, such as broad substrate specificity complemented with high activity levels under a wide range of thermal and pH conditions and salt concentrations, stability in organic solvents, and enantio- and stereoselectivity, are widely used for defining the potential application of esterases and lipases in biotechnology settings (4); however, not all characterized enzymes match these criteria (for example, see reference 26). Table S1 in the supplemental material provides an exhaustive list (with biochemical characteristics) of more than 200 different esterases/lipases from the α/β hydrolase superfamily which have been identified by metagenomic methods from various environments, including soils, compost piles, landfill leachate, bioreactors, activated sludges, marine water, and sediment samples (in-

Received 22 January 2013 Accepted 26 March 2013
Published ahead of print 29 March 2013
Address correspondence to M. Ferrer, mferrer@icp.csic.es.
* Present address: María-Eugenia Guazzaroni, Universidade de São Paulo, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Departamento de Química, Monte Alegre, Ribeirao Preto, São Paulo, Brazil.
M.M.-M. and M.A. contributed equally to this article.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00240-13.

doi:10.1128/AEM.00240-13
The authors have paid a fee to allow immediate free access to this article.

cluding tidal flat sediments, deep sea samples, and water columns) and freshwater samples (including drinking water, pond water, rivers, and hot springs).

In the present work, we identified and successfully cloned, expressed, purified, and characterized seven esterases/lipases from the α/β hydrolase superfamily via an activity-centered metagenome analysis from a microbial community of an evaporite karstic lake (Lake Arreo) in northwest Spain (42°46′N, 2°59′W; altitude, 655 m). Our findings point to the importance of evaporite karstic lakes as rich resources for novel low-temperature- and salt-adapted α/β hydrolases that may be useful for biotechnological applications because of their remarkable and broad substrate profiles. The substrate spectrum of newly isolated enzymes was compared to that of commercially available preparations, and structural models were also used to suggest molecular features that allow for the acceptance of different sets of substrates. Additionally, our work provided further insights into the placing of esterases/lipases at Lake Arreo by using genome linguistics, sequence similarity, and activity phenotyping.

MATERIALS AND METHODS

Materials, strains, and esterase/lipase preparations. Chemicals, biochemicals, and solvents were purchased from Sigma Chemical Co. (St. Louis, MO), Aldrich (Oakville, Canada), or Fluka (Oakville, Canada) and were of pro-analysis quality. The oligonucleotides used for DNA amplification and sequencing were synthesized by Sigma Genosys Ltd. (Pampisford, Cambs, United Kingdom) and are provided in Table S2 in the supplemental material. The nickel-nitrilotriacetic acid (Ni-NTA) His · Bind chromatographic media were from Sigma Chemical Co. (St. Louis, MO). The *Escherichia coli* EPI300-T1^R strain (Epicentre Biotechnologies; Madison, WI), used for the fosmid library construction and screening, and *E. coli* GigaSingles for the gene cloning and BL21(DE3) for the expression using the pET-46 Ek/LIC vector (Novagen, Darmstadt, Germany), were cultured and maintained according to the recommendations of the suppliers and standard protocols (27). Commercial esterase/lipase preparations Novozym 735 (lipase CalA from *Candida antarctica*), Novozym CALB L (lipase CalB from *C. antarctica*), Lipolase 100L (lipase from *Thermomyces lanuginosa*), and Lipozyme RM–Novozym 388L (lipase from *Rhizomocur miehei*) were provided by Novozymes A/S (Bagsvaerd, Denmark). Lipase from *Alcaligenes* sp. was kindly donated by Meito Sangyo Co. (Japan). Lipase from *Rhizopus oryzae* was obtained and prepared as described elsewhere (28).

Selection of hydrolases derived from the evaporite karstic lake (Lake Arreo) fosmid library. A large-insert pCCFOS1 fosmid library was created from the DNA of a microbial community inhabiting sediment samples from an evaporite karstic lake (Lake Arreo). Sediment sampling (10 cm deep) was carried out in February 2007 with a sterilized 50-ml polypropilene Falcon tube in the eastern shallow part of the lake (see Fig. S1 in the supplemental material). The sample was maintained cold and dark during transportation to the laboratory, where it was frozen at −20°C until processing. Total DNA (5.2 μg DNA/g sediment) was extracted using the G'NOME DNA isolation kit (Qbiogene, Heidelberg, Germany). Purified and size-fractioned DNA was ligated into the pCCFOS fosmid vector and further cloned in *E. coli* EPI300-T1^R according to the instructions of Epicentre Biotechnologies (WI) and a procedure described earlier (11). Fosmid clones (40,000; average insert size, 29.7 kbp) harboring approximately 1 Gbp of community genomes were arrayed and grown in 384-well microtiter plates containing Luria-Bertani (LB) medium with chloramphenicol (12.5 μg/ml) and 15% (vol/vol) glycerol and stored at −80°C.

Approximately 11,520 clones, plated onto small (12.5- by 12.5-cm) petri plates (each containing 96 clones) with LB agar containing chloramphenicol (12.5 μg/ml) and the induction solution (Epicentre Biotechnologies, WI), as recommended by the supplier to induce a high fosmid copy number, were screened with α-naphthyl acetate under previously described conditions (29). The positive clones were selected and their DNA inserts sequenced with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT) at Life Sequencing SL (Valencia, Spain). Upon completion of sequencing, the reads were assembled to generate nonredundant metasequences using Newbler GS De Novo Assembler v.2.3 (Roche). GeneMark software (30) was employed to predict potential protein-coding regions (open reading frames [ORFs] with ≥20 amino acids) from the sequences of each assembled contig, and deduced proteins were screened via BLASTP and PSI-BLAST (31). Sequences are available at NCBI under accession number SRA059294.

Cloning, expression, and purification of selected proteins. The cloning, expression, and purification of selected His_6-tagged proteins in the Ek/LIC 46 vector and *E. coli* BL21 were performed as described elsewhere (32), except that protein expression was performed using 1.0 mM isopropyl-β-D-galactopyranoside for 16 h at 16°C. After purification using a Ni-NTA His · Bind resin (Sigma Chemical Co., St. Louis, MO), protein solutions were extensively dialyzed with 20 mM HEPES buffer (pH 7.0) by ultrafiltration through low-adsorption, hydrophilic, 10,000-nominal-molecular-weight-limit-cutoff membranes (regenerated cellulose; Amicon, Madrid, Spain) and stored at −86°C until use. Purity was assessed as >95% using SDS-PAGE, performed on 12% (vol/vol) acrylamide gels as described by Laemmli (33), in a Bio-Rad Mini protein system. Protein concentrations were determined according to Bradford (34).

Biochemical assays. If not otherwise stated, hydrolase reactions using *p*-nitrophenyl (*p*NP) esters were performed in a microplate reader (Synergy HT multimode microplate reader; BioTek) as described previously (32), with minor modifications. Briefly, reaction mixtures contained 2 μg pure enzyme and 0.8 mM *p*NP ester (from a 20 mM stock solution in acetone) in 20 mM HEPES buffer, pH 7.0, at 30°C, in a total volume of 190 μl. Hydrolase-specific activity using other structurally diverse esters was also determined using *p*-nitrophenol as a pH indicator as described elsewhere (35), with small modifications. Briefly, specific activity was calculated by adding 2 μg pure enzyme (a smaller amount of enzyme [0.4 μg] was used for LAE6) and 2 mM substrate from a 100 mM ester stock solution (in acetonitrile) in 2 mM *N*,*N*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES) buffer (pH 7.2) containing 0.45 mM *p*-nitrophenol and 2.5% acetonitrile (vol/vol), at 30°C, in a total volume of 150 μl. In all assays, the reactions were monitored every 2 min by spectrophotometrically measuring the absorbance of *p*-nitrophenol at 410 nm (for *p*NP esters) or 404 nm (for substrates other than *p*NP esters) during 15 min (except for LAE6, for which a 10-min incubation time was used). Under our experimental conditions, the absorption coefficient for *p*-nitrophenol was measured as 15,200 $M^{-1} \cdot cm^{-1}$. One unit of enzyme activity was defined as the amount of enzyme hydrolyzing 1 μmol of substrate in 1 min under the assay conditions. In all cases, three independent experiments were performed, and graphs were plotted using mean values; the standard deviations were less than 5%. Kinetic parameters were calculated by using a conventional Lineweaver and Burk model, at 30°C, as outlined above, in 96-well microtiter plates where each well contained 0.388 to 0.882 μM enzyme solution and 0 to 100 mM substrate.

The optimal pH, temperature, and salt concentration were determined using *p*NP butyrate (0.8 mM final concentration) as the substrate; pH values between 4.0 and 9.0, temperatures between 4 and 70°C, and sodium chloride concentrations ranging from 0.1 to 1.2 M were tested. All of the following buffers were tested at 20 mM: sodium citrate (pH 4.0 to 4.5), sodium acetate (pH 5.0 to 6.0), 2-(*N*-morpholino)ethanesulfonic acid (pH 5.5 to 6.0), HEPES (pH 7.0 to 8.0), piperazine-*N*,*N*′-bis(ethanesulfonic acid) (pH 6.0 to 7.0), K/Na phosphate (pH 7.5), Tris-HCl (pH 8.5), and glycine (pH 9.0 to 9.5). The pH was always adjusted at 25°C. The pH and temperature profiles were obtained at 30°C and pH 7.0, respectively. The optimal anion concentration was obtained at 30°C and pH 7.0. Under our experimental conditions, the absorption coefficients for *p*-nitrophenol were measured for each indicated temperature and pH and ranged from 132 (for pH 4.0) to 28,381 (for pH 9.5) $M^{-1} \cdot cm^{-1}$. These

values were considered for calculating activity for each pH and temperature, as described above.

Oligonucleotide usage pattern analysis. DNA sequences of contigs were searched for oligonucleotide compositional similarity against all sequenced bacterial chromosomes, plasmids, and phages by using the GOHTAM web tool (36). Frequencies of tetranucleotides for compositional genome comparison were calculated as described earlier (37).

Three-dimensional (3D) model analysis. Suitable protein structures to be used as templates for modeling were searched using the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/). The Swiss Model server (http://swissmodel.expasy.org/) was used to generate the enzyme models. DeepView-Swiss-PdbViewer (http://spdbv.vital-it.ch/) was used to analyze the structures and generate the figures.

Sequence read accession number. The sequence reads determined in the course of this work are available at NCBI under accession number SRA059294.

RESULTS AND DISCUSSION

Study site description. Lake Arreo (42°46′N, 2°59′W; altitude, 655 m) is one of the deepest karstic lakes on the Iberian Peninsula (maximum depth, 24.8 m) (38, 39) (see Fig. S1 in the supplemental material). Chemically, the lake is subsaline, with an electrical conductivity range of 703 to 1,727 μS/cm during the last 20 years. Most of the sediment showed anoxic conditions, except for a thin surface layer, which has a Ca-(Mg)-(Na)-SO_4-HCO_3-(Cl) ionic composition (40, 41). The main water physicochemical data from the sampling point were the following: conductivity, 1.079 μS/cm; temperature, 6.9°C; dissolved oxygen content, 11.5 mg/liter (101%); pH 8.0. The mean values of major ionic concentrations (mg/liter) are the following: sulfates (354.9), bicarbonates (269.1), chloride (99.6), calcium (188.4), magnesium (37.5), sodium (51.52), and potassium (14.08).

Anoxic sediment samples (10 cm deep) were taken, and total DNA was purified. Size-fractionated DNA was ligated into the pCCFOS fosmid vector and further cloned in *E. coli* EPI300-$T1^R$ to create a fosmid clone library harboring approximately 1 Gbp of community genomes.

Metagenome library screening and sequence analysis. Approximately 11,520 fosmid clones from the metagenomic library of Lake Arreo sediment (nearly 342 Mbp of metagenome DNA) were screened using plate-based screens for hydrolytic activity against α-naphthyl acetate, a model esterase substrate (29). Ten positive clones (incidence of positive clones, 1/1,152) were identified. Fosmid DNA isolated from each clone was mixed in equal amounts and sequenced as a pool using a Roche GS FLX DNA sequencer (1/16 plate), which produced 77,841 reads with an average length per read of 454.69 bp. Accordingly, a total of 38.5 Mbp of raw DNA sequences were obtained, which were assembled into 735,686 bp (19 contigs with lengths ranging from 1,854 to 43,416 bp), with an average GC content of 58.75%. A total of 378 open reading frames were identified, which were analyzed and compared to the sequences available in the databases. Seven out of 10 identified genes encoding the predicted esterases from the α/β hydrolase superfamily (named LAE1 to LAE7) were expressed and produced in soluble form in *E. coli* BL21(DE3). The corresponding gene products were purified (see Fig. S2 in the supplemental material), and their activities were characterized using 15 model substrates and an array of 86 additional structurally diverse esters (see Fig. S3 in the supplemental material).

Sequence analysis of α/β hydrolases from Lake Arreo. Analysis of sequences of the Lake Arreo proteins showed that they all belong to the α/β hydrolase superfamily, with amino acid sequence identity ranging from 31 to 57%. As determined by Matcher (EMBOSS package), among Lake Arreo α/β hydrolases, LAE4 and LAE7 were the most similar enzymes (71.4% sequence identity), whereas LAE5 and LAE6 were the most different at the sequence level (18.2% identity) (see Table S3 in the supplemental material). The LAE1 protein was structurally most similar to thermophilic esterase from *Thermotoga maritima* (31%; Protein Data Bank [PDB] code 3DOH_A [42]) and α/β hydrolase from *Agrobacterium tumefaciens* (24%; PDB code 2r8b); LAE2 was most similar to GDSL-like brain platelet-activating factor acetylhydrolases (SGNH_hydrolase subfamily) (31%, PDB code 1FXW [43], and 28%, PDB code 1ES9 [44]); LAE3 was most similar to thermostable esterase/lipase from uncultured bacterium (36%; PDB code 3V9A_A) and hormone-sensitive esterase (35%; PDB code 3fak [45]); LAE4 was most similar to esterase from *Pyrobaculum calidifontis* (25%; PDB code 2YH2_A [46]) and bacterial α/β hydrolase (19%; PDB code 3k2i); LAE5 was most similar to the GDSL family lipolytic protein (SGNH_hydrolase subfamily) from *Alicyclobacillus acidocaldarius* subsp. *acidocaldarius* Dsm 446 (54%; PDB code 3RJT_B); LAE6 was most similar to thermophilic and thermostable carboxylesterase from the metagenomic library (41%; PDB code 2c7b [47]); and LAE7 was most similar to GDSL-like brain platelet-activating factor acetylhydrolase (SGNH_hydrolase subfamily) (32%; PDB code 1ES9 [44]).

The deduced molecular masses and estimated pI values of Lake Arreo α/β hydrolases range from 24,328.83 to 34,002.96 Da and from 4.78 to 10.04, respectively. Suitable templates have been found for each enzyme, and 3D models of the corresponding protein structures are represented in Fig. S4 in the supplemental material. The analysis of the seven enzymes, which can be classified in the microbial lipase/esterase families described by Arpigny and Jaeger (48), allows the distinction of three structural groups. The first group (LAE2, LAE5, and LAE7) are GDSL esterases/lipases (49). A characteristic feature of GDSL enzymes is that the serine-containing motif is close to the protein N terminus. However, whereas the presumptive nucleophile of LAE5 corresponds to Ser15, that of LAE2 and LAE7 correspond to Ser143 and Ser130, respectively. The structural analysis of the LAE2 and LAE7 enzymes shows that they are bimodular enzymes in which the α/β hydrolase fold of the esterase catalytic unit is preceded at an N-terminal position by a 100- to 120-amino-acid module whose structure could not be modeled due to the absence of suitable templates. A remarkable feature of this module is the presence of a collagen-like sequence with a glycine/proline-rich stretch (GPGG PGGPRGGGFGAPPTPPGP in LAE2). These sequences, which are found in diverse bacterial proteins, are considered to be signatures of phage origin (50); this agrees with the analysis of DNA fragments using a genome linguistics approach (see below). Whatever the case, the fact that the catalytic Ser located closer to the N terminus in LAE5 than in LAE2 and LAE7 could make the catalytic site more accessible to substrates; this was further confirmed (see below) by showing that LAE5 accepted a higher number of substrates and by showing that although all three enzymes hydrolyzed *p*NP acetate and *p*NP propionate, LAE5 did prefer *p*NP propionate, while LAE2 and LAE7 preferred shorter *p*NP acetate (Table 1). Additionally, the fact that LAE5 retained circa 70% of the activity at 4°C, while LAE2 and LAE7 retained only 42 and 28%, respectively (Fig. 1 and see below), suggests a link between the differences in active center location and optimization of

TABLE 1 Specific activities of the wild-type α/β hydrolases from Lake Arreo against a set of model α-naphthyl, *p*NP, and triacylglycerol ester substrates

Substrate	Activity[a] (U/mg)						
	LAE1	LAE2	LAE3	LAE4	LAE5	LAE6	LAE7
α-Naphthyl acetate	ND	111.04	ND	1.63	83.36	124.8	99.22
α-Naphthyl propionate	3.58	16.01	ND	8.34	24.35	55.1	11.73
α-Naphthyl butyrate	3.36	0.92	0.74	3.93	1.23	28.3	1.21
*p*NP acetate	0.133	5.57	23.80	0.05	1.23	34.3	0.64
*p*NP propionate	0.365	0.29	0.84	0.15	2.05	135.5	0.13
*p*NP butyrate	1.072	0.27	0.29	0.11	0.01	105.2	0.02
*p*NP octanoate	0.129	ND	0.04	ND	ND	2.6	ND
*p*NP decanoate	0.028	ND	0.02	ND	ND	1.0	ND
*p*NP laurate	ND	ND	ND	ND	ND	0.2	ND
Triacetin	0.015	1.07	0.01	1.01	1.14	18.7	1.19
Tripropionin	0.080	ND	0.01	2.33	0.02	14.9	0.39
Tributyrin	0.005	ND	0.09	0.83	0.00	11.3	0.14
Tricaprin	0.001	ND	ND	0.06	0.00	5.5	ND
Tricaprylin	ND	ND	ND	ND	ND	3.4	ND
Trilaurin	ND	ND	ND	ND	ND	ND	ND

[a] For α-naphthyl and triacylglycerol esters, reaction mixtures contained 2 μg pure enzyme (a lower amount of enzyme [0.4 μg] was used for LAE6), 2 mM substrate, 0.45 mM *p*-nitrophenol, and 2.5% (vol/vol) acetonitrile in 2 mM BES buffer (pH 7.2), at 30°C, in a total volume of 150 μl. For *p*NP esters, reaction mixtures contained the same amount of enzymes and 0.8 mM *p*NP ester in 20 mM HEPES buffer (pH 7.0), at 30°C, in a total volume of 190 μl. In all cases, the reactions were monitored every 2 min during 15 min, and the absorbance of *p*-nitrophenol at 410 nm (for *p*NP esters) or 404 nm (for substrates other than *p*NP esters) was measured. In all cases, 1 U of enzyme activity was defined as the amount of enzyme hydrolyzing 1 μmol of substrate in 1 min under the assay conditions. Three independent experiments were performed for each parameter, and mean values are given; the standard deviations were less than 5%. ND, no activity detected under the experimental conditions assayed.

activity at low temperatures. Further structural analysis is in progress to confirm this hypothesis. The second group corresponds to common esterases/lipases with typical α/β hydrolase folds and GxSxG motif (LAE1, LAE4, and LAE6). LAE1 belongs to family I and LAE4 and LAE6 to family VI; with a molecular mass in the range 23 to 26 kDa, the enzymes in family VI are among the smallest esterases known, in agreement with the theoretical molecular mass of LAE4, which is the smallest (24 kDa) among the Lake Arreo esterases. The third group, represented by LAE3, belongs to the hormone-sensitive lipase/esterase (HSL) type (51) and also has a GxSxG motif. It is thought that HSL hydrolases retained high activity at low temperature, although sequence analysis indicates that temperature adaptation is not responsible for such sequence conservation (48); the fact that LAE3 did show maximal activity at 16°C (Fig. 1) agrees with this hypothesis.

According to sequence and 3D model analyses (see Fig. S4 in the supplemental material), the catalytic triads were tentatively identified: Ser166, Glu217, and His249 (in LAE1), Ser140, Glu292, and His295 (in LAE2), Ser143, Glu237, and His267 (in LAE3), Ser144, Asp248, and His281 (in LAE4), Ser15, Asp192, and His195 (in LAE5), Ser161, Asp256, and His286 (in LAE6), and Ser130, Glu282, and His285 (in LAE7).

Optimization of activity and anion stimulation of esterases/lipases from Lake Arreo. As shown in Fig. 1, purified proteins showed high hydrolytic activity against the model esterase substrate *p*NP butyrate at 16 to 40°C and pH 8.0 to 8.5, with the hydrolases LAE3, LAE5, and LAE6 being the most active at temperatures as low as 4°C, as they retained from 64 to 78% of the activity compared to the activity level at the optimal temperatures. Hydrolytic activities were stimulated (1.5 to 2.2 times) by the addition of NaCl to the reaction mixture (0.1 to 1.2 M NaCl) for all enzymes but one (LAE3), which was inhibited from 1.6 (at 0.1 M) to 3.8 (at 1.2 M) times. This is consistent with the fact that the Lake Arreo basin was formed by dissolved evaporates; thus, it represents a subsaline environment (52). The fact that all but one of the esterases was activated by NaCl indicates that enzyme properties resembled habitat-specific characteristics, and that activation by chloride is common for enzymes from Lake Arreo. Similar observations have been reported for the cold-active and anion-activated carboxyl esterase OLEI01171 from the oil-degrading marine bacterium *Oleispira antarctica* (53).

Using model substrates that included 3 α-naphthyl, 6 *p*NP, and 6 triacylglycerol esters and according to specific activity (U/mg) determinations (Table 1), we observed that LAE6 did show the highest capacity to accept longer esters (up to tricaprin, *p*NP laurate, and α-naphthyl butyrate), which is in line with the highest lipase character of this enzyme compared to other Lake Arreo α/β hydrolases. *p*NP esters were preferred substrates for LAE3, whereas LAE1, LAE2, LAE4, LAE5, and LAE7 preferentially hydrolyzed α-naphthyl esters. LAE6 showed similar activity for both substrates. In all cases, the activity toward triacylglycerols was significantly lower (from 3.5-fold for LAE4 to 100-fold for LAE2 compared to α-naphthyl esters). Overall, LAE6 (approximately 125 U/mg for α-naphthyl acetate) was the most active enzyme, whereas LAE3 showed the lowest activity (0.74 U/mg with α-naphthyl butyrate) (Table 1).

Substrate fingerprinting of esterases/lipases from Lake Arreo. The substrate range of the purified esterases was characterized using a battery of 86 different esters (see Fig. S3 in the supplemental material) that included 25 halogenated alkyl and aryl esters, 34 alkyl esters, 12 aryl esters, 10 hydroxycinammic esters, 2 epoxides, 1 lactone, and 2 carbohydrate esters, using a colorimetric method in which *p*-nitrophenol was used as a pH indicator (35). The fingerprints of LAE1 to LAE7 are shown in Fig. 2 and 3, which showed that 52 out of the 86 (or 61%) esters used in the present study were accepted as substrates (for substrates not hydrolyzed by any of the enzymes, see Fig. S3 in the supplemental material).

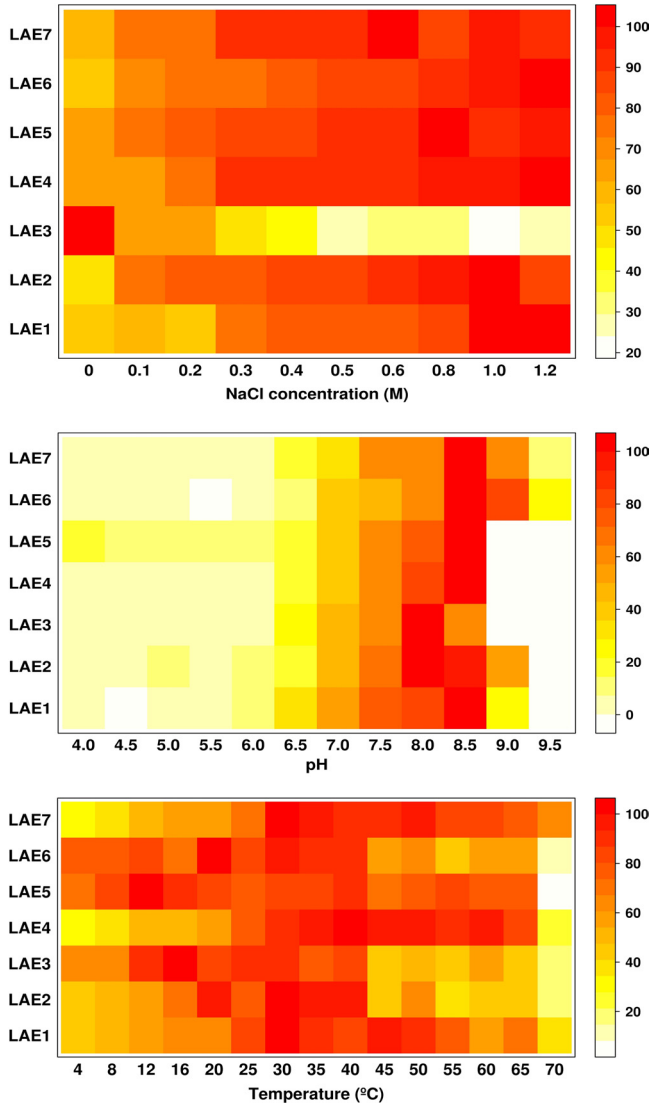


FIG 1 Effect of NaCl concentration (upper), pH (middle), and temperature (lower) on the hydrolytic activity of Lake Arreo enzymes. The heat map colors represent the relative percentages of specific activity (U/mg) compared to the maximum (100%) within each enzyme, using *p*NP butyrate as the substrate. The effect of NaCl concentration was measured in 20 mM HEPES buffer (pH 7.0) at 30°C and varied from 0 to 1.2 M; the pH dependence was tested in the range of pH 4.0 to 9.5 at 30°C in 20 mM buffers; the temperature dependence in the range of 4 to 70°C at pH 7.0 was also determined. Reaction conditions were as described in Table 1 and Materials and Methods.

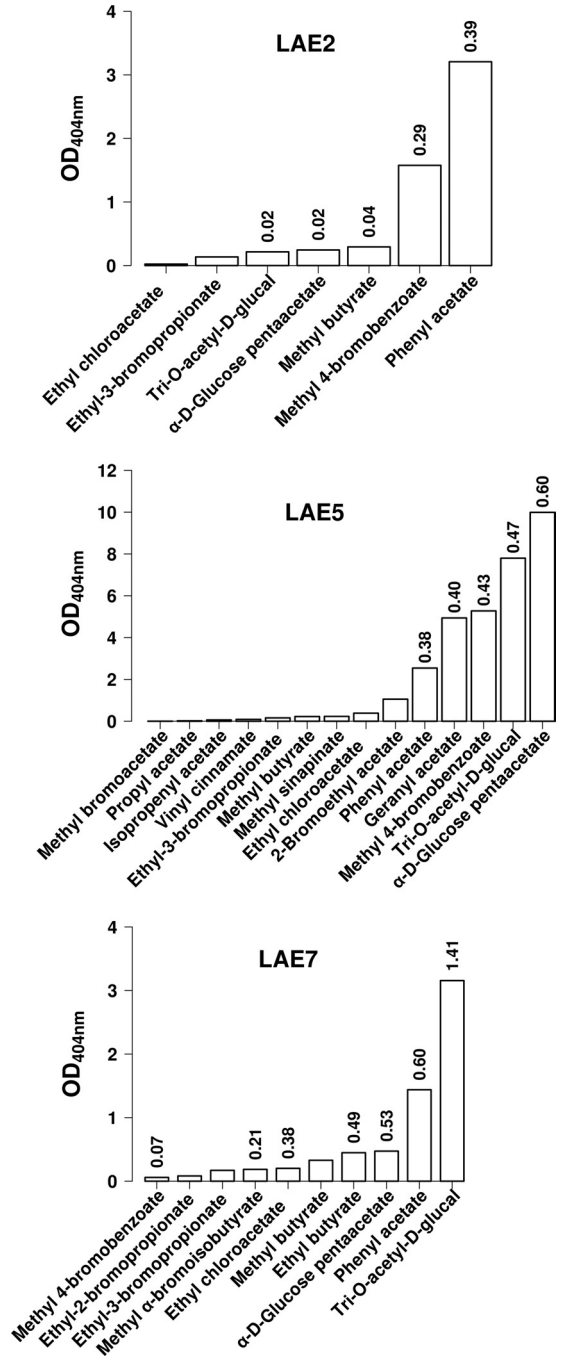
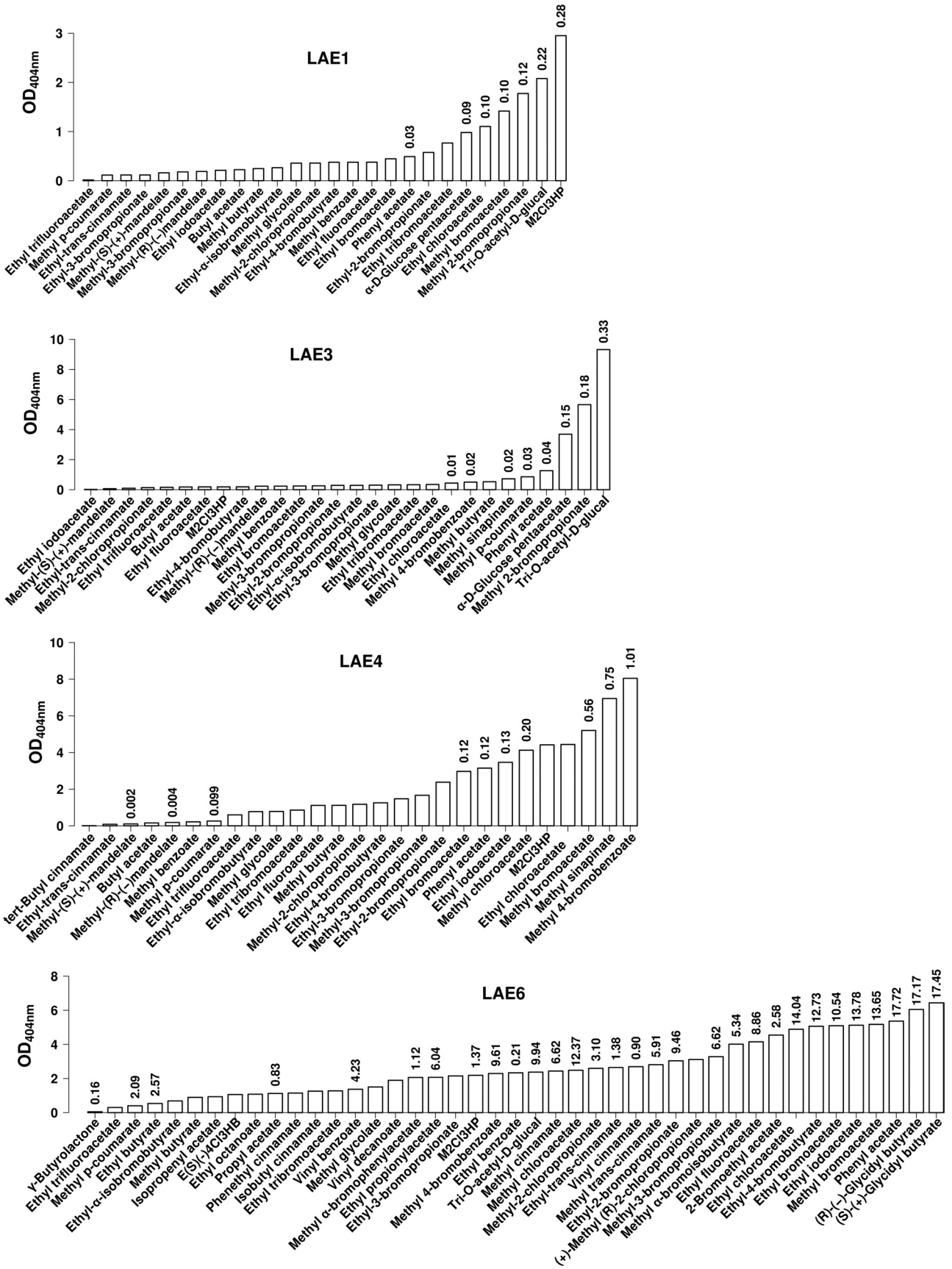


FIG 2 Substrate profiles of the wild-type LAE2, LAE5, and LAE7 α/β hydrolases from Lake Arreo against a set of structurally diverse ester substrates. LAE2, LAE5, and LAE7 were the α/β hydrolases characterized by a restricted substrate spectrum. Reactions (30°C and pH 7.2) were performed as described for the α-naphthyl and triacylglycerol esters (Table 1) and were monitored during 15 min, and the absorbance of *p*-nitrophenol at 404 nm was recorded and plotted. For best and/or representative substrates, specific activities (U/mg) were calculated as described in Table 1 and Materials and Methods and are shown on the top of bars. In all cases, three independent experiments were performed for each parameter, and graphs were plotted using mean values; the standard deviations were less than 5%.

Using the ester library substrates (Fig. 2 and 3), we observed that LAE6 exhibits the broader substrate spectrum, being able to hydrolyze 43 substrates, followed by LAE3 (28 substrates), LAE4 (27 substrates), LAE1 (26 substrates), and, to lesser extents, LAE5 (14 esters), LAE7 (10 esters), and LAE2 (7 esters). The majority of compounds were accepted by at least two or more enzymes; however, 14 substrates were LAE6 specific (Fig. 3), including ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate, (+)-methyl-(*R*)-2-chloropropionate, methyl α-bromophenylacetate, methyl α-bromoisobutyrate, ethyl octanoate, vinyl decanoate, ethyl benzoate, vinyl benzoate, methyl cinnamate, methyl *trans*-cinnamate, isobutyl cinnamate, phenethyl cinnamate, (*R*)-(−)-glycidyl butyrate, (*S*)-(+)-


LAE1
OD404nm
Ethyl trifluoroacetate
Methyl p-coumarate
Ethyl-trans-cinnamate
Ethyl-3-bromopropionate
Methyl-(S)-(+)-mandelate
Methyl-3-bromopropionate
Methyl-(R)-(−)-mandelate
Ethyl iodoacetate
Butyl acetate
Methyl butyrate
Ethyl-α-isobromobutyrate
Methyl glycolate
Methyl-2-chloropropionate
Ethyl-4-bromobutyrate
Methyl benzoate
Ethyl fluoroacetate
Ethyl bromoacetate
Phenyl acetate
Ethyl-2-bromopropionate
Ethyl tribromoacetate
α-D-Glucose pentaacetate
Ethyl chloroacetate
Methyl bromoacetate
Methyl 2-bromopropionate
Tri-O-acetyl-D-glucal
M2Cl3HP
0.03
0.09
0.10
0.10
0.12
0.22
0.28
LAE3
OD404nm
Ethyl iodoacetate
Methyl-(S)-(+)-mandelate
Ethyl-trans-cinnamate
Methyl-2-chloropropionate
Ethyl trifluoroacetate
Butyl acetate
Ethyl fluoroacetate
M2Cl3HP
Ethyl-4-bromobutyrate
Methyl-(R)-(−)-mandelate
Methyl benzoate
Ethyl bromoacetate
Methyl-3-bromopropionate
Ethyl-2-bromopropionate
Ethyl-α-isobromobutyrate
Ethyl-3-bromopropionate
Methyl glycolate
Ethyl tribromoacetate
Methyl bromoacetate
Ethyl chloroacetate
Methyl 4-bromobenzoate
Methyl butyrate
Methyl sinapinate
Methyl p-coumarate
Phenyl acetate
α-D-Glucose pentaacetate
Methyl 2-bromopropionate
Tri-O-acetyl-D-glucal
0.01
0.02
0.02
0.03
0.04
0.15
0.18
0.33
LAE4
OD404nm
tert-Butyl cinnamate
Ethyl-trans-cinnamate
Methyl-(S)-(+)-mandelate
Butyl acetate
Methyl-(R)-(−)-mandelate
Methyl benzoate
Methyl p-coumarate
Ethyl trifluoroacetate
Ethyl-α-isobromobutyrate
Methyl glycolate
Ethyl tribromoacetate
Ethyl fluoroacetate
Methyl butyrate
Methyl-2-chloropropionate
Ethyl-4-bromobutyrate
Methyl-3-bromopropionate
Ethyl-2-bromopropionate
Ethyl bromoacetate
Phenyl acetate
Ethyl iodoacetate
Methyl chloroacetate
Ethyl chloroacetate
M2Cl3HP
Methyl bromoacetate
Methyl sinapinate
Methyl 4-bromobenzoate
0.002
0.004
0.099
0.12
0.12
0.13
0.20
0.56
0.75
1.01
LAE6
OD404nm
γ-Butyrolactone
Ethyl trifluoroacetate
Methyl p-coumarate
Ethyl butyrate
Ethyl-α-isobromobutyrate
Methyl butyrate
Isopropenyl acetate
E(S)(-)4Cl3HB
Ethyl octanoate
Propyl acetate
Phenethyl cinnamate
Isobutyl cinnamate
Ethyl tribromoacetate
Vinyl benzoate
Vinyl glycolate
Vinyl decanoate
Methyl α-bromophenylacetate
Ethyl propionylacetate
Ethyl-3-bromopropionate
M2Cl3HP
Methyl 4-bromobenzoate
Ethyl benzoate
Tri-O-acetyl-D-glucal
Methyl cinnamate
Methyl chloroacetate
Methyl-2-chloropropionate
Ethyl-trans-cinnamate
Vinyl cinnamate
Methyl trans-cinnamate
Ethyl-2-bromopropionate
(+)-Methyl (R)-2-chloropropionate
Methyl-3-bromopropionate
Methyl α-bromoisobutyrate
Ethyl fluoroacetate
2-Bromoethyl acetate
Ethyl chloroacetate
Ethyl-4-bromobutyrate
Ethyl bromoacetate
Ethyl iodoacetate
Phenyl acetate
(R)-(−)-Glycidyl butyrate
Methyl bromoacetate
(S)-(+)-Glycidyl butyrate
0.16
2.09
2.57
0.83
4.23
1.12
6.04
1.37
9.61
0.21
9.94
6.62
12.37
3.10
1.38
0.90
5.91
9.46
6.62
5.34
8.86
2.58
14.04
12.73
10.54
13.78
13.65
17.72
17.17
17.45

glycidyl butyrate, and γ-butyrolactone, and the flavor terpene compound geranyl acetate was accepted as the substrate only by LAE5 (Fig. 2). Specific activity (U/mg) determinations were performed for the best representative substrates (Fig. 2 and 3). At 30°C and pH 7.2, methyl-2-chloro-3-hydroxypropionate (for LAE1, 0.28 U/mg), phenyl acetate (for LAE2, 0.39 U/mg), tri-O-acetyl-D-glucal (for LAE3, 0.33 U/mg; for LAE7, 1.41 U/mg), methyl 4-bromobenzoate (for LAE4, 1.01 U/mg), α-D-glucose penta-acetate (for LAE5, 0.60 U/mg), and (*S*)-(+)-glycidyl butyrate (for LAE6, 17.45 U/mg) were the preferred substrates.

The ability to hydrolyze nonhalogenated and halogenated (including those containing bromo-, chloro-, fluor-, and iodo-) alkyl and aryl esters was demonstrated for all enzymes (Fig. 2 and 3). Of special significance is that long alkyl and aryl esters with vinyl and isoprenyl substituents, such as vinyl decanoate, vinyl benzoate, and isoprenyl acetate, were mainly hydrolyzed by LAE6 (up to 4.23 U/mg) (Fig. 3); thus, this enzyme may be applied in transesterification reactions using vinyl esters, unlike the other enzymes from Lake Arreo. In addition, LAE1 was the only enzyme that did not accept aromatic halogenated esters, such as methyl 4-bromobenzoate (Fig. 3), although it was able to hydrolyze the nonhalogenated substrate (methyl benzoate). As shown in Fig. 2 and 3, four enzymes exhibited the capacity to accept the carbohydrate ester α-D-glucose penta-acetate (LAE2, LAE3, LAE5, and LAE7), and all but one (LAE4) showed activity for tri-O-acetyl-glucal. Finally, four enzymes were able to hydrolyze one or several cinnamate esters (LAE1, LAE3, LAE4, and LAE5), four accept *p*-coumarate esters (LAE1, LAE3, LAE4, and LAE6), and LAE3 and LAE5 were the only ones accepting methyl sinnapinate esters; no activity toward methyl ferulate was observed for any of the enzymes. Of special significance is that LAE4 (Fig. 3) was the only enzyme that showed activity on cinnamate esters of tertiary alcohols, such as tert-butyl cinnamate, although it was unable to accept primary and secondary alcohol substituents, such as methyl or isobutyl substituents. Further, LAE6 exhibited activity toward cinnamate esters with methyl, ethyl, isobutyl, vinyl, and the large aromatic phenethyl substituents, for which low or now activity was detected for the majority of other esterases/lipases (Fig. 3).

In relation to the enantioselective character of Lake Arreo hydrolases, we further observed that LAE1, LAE3, and LAE4 did show the capacity to hydrolyze both enantiomers of methyl mandelate, whereas LAE6 was the only one accepting both enantiomers of glycidyl butyrate (Fig. 2 and 3); other chiral esters, such as menthyl, neomenthyl, or lactate esters, were not substrates for any of the enzymes. According to the Quick E assay (34) and k_{cat}/K_m determinations for separate enantiomers (see Table S4 in the supplemental material), apparent enantiomeric ratios (E_{app} values) were calculated. At 30°C and pH 7.2, E_{app} varied from 816 (for LAE3) to 15 (for LAE6), 8.3 (for LAE1), and 2.0 (for LAE4). LAE1 and LAE3 did show enantiopreference for methyl-(*R*)-(−)-mandelate, whereas LAE4 preferred methyl-(*S*)-(+)-mandelate and LAE6 preferred (*R*)-(−)-glycidyl butyrate. In addition, LAE6 further exhibited activity toward γ-butyrolactone, although the activity for this substrate was 108-fold lower than that of (*S*)-(+)-glycidyl butyrate, which was the best substrate (Fig. 3).

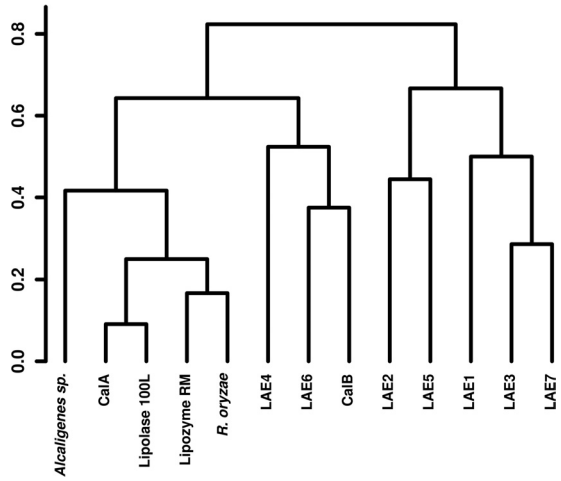


FIG 4 Clustering of substrate spectrum profile of Lake Arreo enzymes and commercial preparations, applying the Pearson's correlation to calculate the distances. Hierarchical clustering was derived from binomial distribution based on the presence or absence of activity against particular substrates (activity data not shown), measured as described in Table 1, using 2-μg protein extracts.

Our results indicate that esterases/lipases from Lake Arreo are characterized by high activities and different substrate profiles and enantioselectivities; therefore, they are potentially useful for various biotechnological applications. The differences in hydrolytic capacities (using the half-saturation [Michaelis] coefficient [K_m], the catalytic rate constant [k_{cat}], the catalytic efficiency [k_{cat}/K_m] values, and/or specific activities; see Table S4 in the supplemental material) with distinct acid and alcohol substituents of different lengths and nature further indicate differences in active sites. Having said that, it should be highlighted that, compared to previously reported esterases/lipases (see Table S1 in the supplemental material), some of the substrate profiles described here are biologically relevant, such as the ability of a single esterase/lipase such as LAE6 to hydrolyze triacylglycerols, halogenated and nonhalogenated alkyl and aryl esters, cinnamoyl and carbohydrate esters, lactones, and chiral epoxides to a similar extent, which has not been previously described. Enzymes LAE1 and LAE2, which belong to the same contig (i.e., are produced by the same bacterium; see below), have similar optimal parameters (pH 8.0 to 8.5, 30°C, and up to 1.2 M NaCl; Fig. 1) but have quite distinct substrate profiles (Fig. 3 and 4). Thus, only 6 out of 27 substrates were hydrolyzed by both enzymes, with LAE1 possessing the broader substrate spectrum; these distinct functionalities suggest complementary metabolic and ecological capacities *in vivo*, together with different biotechnological capacities.

Comparison to commercial esterases and lipases. Lipases and esterases are one of the most important classes of hydrolytic enzymes in industrial settings (2, 54). There is a wide range of available preparations that can be tested for particular applications and for comparative studies (3, 4, 55); in this context, it is interesting to evaluate and compare the substrate spectrum of newly isolated

FIG 3 Substrate specificity of the wide-substrate-spectrum LAE1, LAE3, LAE4, and LAE6 α/β hydrolases from Lake Arreo against a set of structurally diverse ester substrates. LAE1, LAE3, LAE4, and LAE6 were the hydrolases characterized by the widest substrate spectrum. Reaction conditions (30°C and pH 7.2) and activity parameter determinations were as described in Table 1 and the legend to Fig. 2. M2Cl3HP, methyl-2-chloro-3-hydroxypropionate.

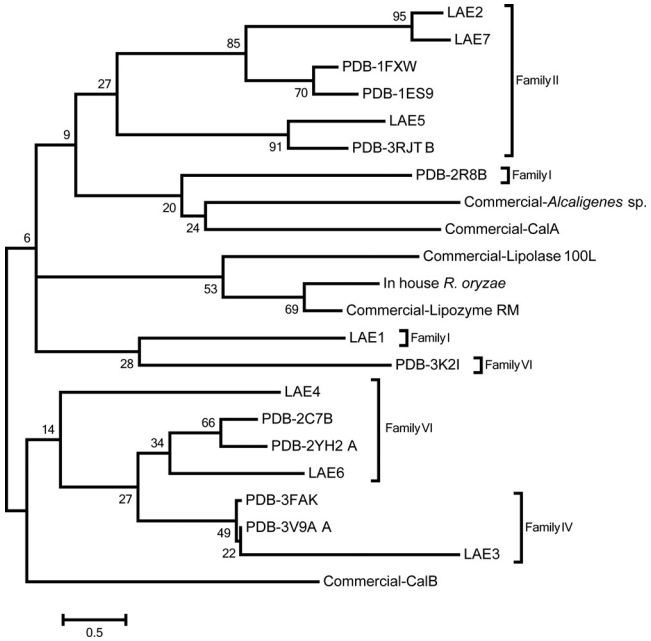


FIG 5 Dendrogram of protein sequence similarity relationships between newly identified and reference esterases/lipases. Enzyme families are depicted according to the Arpigny and Jaeger classification (48).

enzymes to that of commercially available and most common esterases and lipases. Here, we used five esterase/lipase commercial preparations, namely, Novozym 735 (lipase CalA from *Candida antarctica*), Novozym CALB L (lipase CalB from *C. antarctica*), Lipolase 100L (lipase from *Thermomyces lanuginosa*), Lipozyme RM-Novozym 388L (lipase from *Rhizomocur miehei*), and lipase from *Alcaligenes* sp. Lipase from *Rhizopus oryzae*, obtained and prepared as described elsewhere (28), was also added to the study.

The clustering analysis, generated from a binomial distribution based on the presence or absence of activity (at 30°C and pH 7.2) against the set of 101 different esters used here (Fig. 4), suggested that six of the Arreo Lake α/β hydrolases were functionally closer to each other than to the other preparations which clustered together. Lipases from *Rhizopus oryzae* and *Rhizomocur miehei* also showed similar profiles and clustered together with lipases from *T. lanuginosa* and CalA from *C. antarctica*. Interestingly, the lipase LAE6 clusters more closely with the lipase CALB L from *C. antarctica*, with both forming a separate group, which further indicates that both enzymes possess similar substrate spectra; this is of special significance, as Novozym CALB L is one of the most common commercially available lipases (Novozymes A/S, Bagsvaerd, Denmark). However, some minor differences were observed: (i) LAE6 hydrolyzed tri-*O*-acetyl-D-glucal (9.9 U/mg) and methyl 4-bromobenzoate (9.6 U/mg), for which no activity with CalB was detected, and (ii) CalB hydrolyzed methyl 2-bromopropionate (0.74 U/mg), for which negligible activity was found for LAE6. This suggests structural factors as determinants for substrate specificity when comparing both lipases.

To analyze the sequence similarity of known commercial and newly discovered esterases/lipases belonging to different families, a dendrogram was created in MEGA5 (56) by the maximum-likelihood algorithms (Fig. 5). The dendrogram was further used to evaluate whether the placing of the different enzymes in the phylogenetic tree was related to their placing in the activity profile-based tree (Fig. 4). The sequences of reference and commercial esterases were obtained from the NCBI and PDB databases. The protein sequences were aligned by MUSCLE (57) and edited manually in JalView 2.5.1 (58), correlating with the alignment quality histogram. Due to an extreme diversity of selected sequences of enzymes, some of which were bacterial (PDB and Lake Arreo sequences) and others fungal (5 out of 6 sequences of commercial enzymes), the final alignment was reduced to 161 amino acid residues by removal of ambiguously aligned regions constituting 70% of the initial alignment. Analysis of the alignment by MEGA5 suggested WAG+G as the best evolutionary model. Robustness of the dendrogram was analyzed by the Bootstrap analysis based on 100 replicates of the initial data set. As shown in Fig. 4 and 5, we made the following observations: (i) commercial lipases CalA, Lipolase 100L, Lipozyme RM, *Alcaligenes* sp., and *R. oryzae* formed a separate cluster at the sequence level, as was also found for the activity clustering; (ii) similarly, lipases/esterases CalB, LAE4, and LAE6 clustered together at both the sequence and activity levels, which is of special significance considering the extreme diversity of selected sequences; (iii) LAE1, LAE2, LAE5, and LAE7 formed a separate group from LAE4 and LAE6 at both levels; and (iv) no clear correlation between sequence and activity was observed for LAE3. Accordingly, under the experimental setting applied here, the results suggest that, to some extent, an association between the phylogenetic/sequence positioning and the activity relationships exists.

Analysis of metagenomic DNA fragments using genome linguistics approach. Compositional similarity between the metagenomic fragments and the sequences of sequenced bacterial genomes and plasmids was analyzed by the comparison of frequencies of tetranucleotides in DNA sequences. A comprehensive analysis of the GOHTAM and BLAST analysis results is shown in the supplemental material. Five studied metagenomic DNA fragments containing genes encoding α/β hydrolases (excluding the LAE6-containing contig 19, which was too short for the compositional analysis) shared a significant level of tetranucleotide usage pattern similarity that indicated their origination from related organisms, which may be of *Burkholderia*/*Ralstonia* (*Betaproteobacteria*) and/or *Rhizobium*/*Methylobacterium* (*Alphaproteobacteria*) lineages. A search against the GOHTAM database showed compositional similarities to multiple bacterial genomes, plasmids, and phages. The protein blast of contig-encoded proteins against plasmid proteins also showed many hits but no gene syntenies. The DNA fragment (contig 2) containing LAE1 and LAE2 enzymes is most similar to several plasmids from *Cupriavidus* and *Ralstonia*, and it shares one major facilitator superfamily (MFS_1) gene and an integrase fragment with a genomic island from *Ralstonia pickettii*. Accordingly, we can conclude that this fragment most likely is a mobilome associated with either a plasmid or an integrated genomic island of a betaproteobacterium related to *Cupriavidus* and *Ralstonia*. Several genes of this fragment share similarity with the corresponding genes of *Arthrobacter aurescens* TC1 plasmid TC2 (NC_008713) and *Shewanella baltica* OS155 plasmid pSbal01 (NC_009035). The DNA fragment (contig 3) containing LAE3 enzyme shows some compositional similarity to alphaproteobacterial *Methylobacterium* plasmids, and it shares several genes with *Nitrobacter hamburgensis* X14 plasmid 1 (NC_007959). It also contains a gene for transposase IS4. The DNA fragment (contig 4) containing LAE7 enzyme re-

sembles betaproteobacterial *Burkholderia* phages and contains phage major capsid proteins. The DNA fragment (contig 6) containing LAE4 enzyme has weak but consistent compositional similarity to alphaproteobacterial *Rhizobium* plasmids and multiple blast hits against *Rhizobium leguminosarum* bv. trifolii WSM1325 plasmid pR132504 (NC_012852), *Rhizobium leguminosarum* bv. trifolii WSM1325 plasmid pR132505 (NC_012854), *Rhizobium etli* CFN 42 plasmid p42c (NC_007764), and *Azospirillum* sp. strain B510 plasmid pAB510f (NC_013860). The DNA fragment (contig 7) containing LAE5 enzyme shows weak compositional similarity to the gammaproteobacterial *Azotobacter* genome and alphaproteobacterial *Azospirillum* plasmids and weak but persistent blast hits against several bacterial plasmids. The DNA fragment (contig 19) containing LAE6 enzyme is too short for the compositional analysis, but its genes are similar to those in the alphaproteobacterial *Sphingomonas* sp. strain KA1 plasmid pCAR3 (NC_008308). According to this analysis, DNA fragments containing Lake Arreo esterases may be plasmids or plasmid-borne genomic islands from proteobacteria, although contig 4, containing LAE7, most likely is a prophage.

Taken together, in the present study an esterase-driven assay, based on a well-established screen with α-naphthyl acetate, was used to identify hydrolytic activity in 10 *E. coli* clones. These clones harbor DNA plasmid fragments from a microbial community from an evaporite karstic lake (Lake Arreo). Lake Arreo is one of the few relatively deep (maximum depth [z_{max}] = 24 m) karstic lakes in Spain, and it developed in gypsum formations. Our study of α/β hydrolases mined from Lake Arreo suggests that they display habitat-specific characteristics, and that cold-adapted or psychrophilic proteobacteria occupy this ecological niche, with average temperatures ranging from 4.7 to 19.8°C, as they produced low-temperature active and anion-activated enzymes with unusual substrate specificities. The production of highly promiscuous hydrolases with broad substrate profiles may have important metabolic and ecological implications, such as that the function of these enzymes lies in substrate scavenging in low-temperature and substrate-poor karstic environments. A more comprehensive structural survey of enzymes from Lake Arreo, currently in progress, will determine whether or not this hypothesis is correct. Having said that, recent developments for the production of esterase and lipase improved commercial preparations have been achieved, and many factors account for the cost-effective utilization of esterases and lipases (2). This study provides experimental evidence that enzymes from evaporite bacterial metagenomes are of great interest for biotechnological processes, because they are salt-tolerant and are active at low temperatures and against a broad constellation of structurally diverse esters. This should be further evaluated and analyzed with commercial preparations under application conditions in a plethora of reactions according to the substrate specificity reported here. For example, preliminary tests confirmed that LAE6 preparations had performed well for esterification and transesterification reactions of sucrose with methyl and vinyl esters in several organic solvents; also, LAE3 preparations have been found preliminarily to be suitable for the production of enantiomers, such as (*R*)-mandelic acid, for synthetic purposes. Finally, this study may open research avenues into comparative catalysis models and structural-functional studies and confirms the necessity of isolating and characterizing new enzymes from environmental metagenomes.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support provided by the Spanish Ministry of Economy and Competitiveness (project CSD2007-00005), the European Community project MAGICPAH (FP7-KBBE-2009-245226), the European Regional Development Fund (ERDF), and the Government of Canada through Genome Canada, Ontario Genomics Institute, and Ontario Research Fund (2009-OGI-ABC-1405 and ORF-GL2-01-004). M.-E.G. thanks the CSIC for a JAE fellowship.

We thank Javier Tamames for his excellent support for taxonomic analysis. We thank Francisco J. Plou for his excellent support in the use of commercial lipase preparations used in the present study.

REFERENCES

1. **McQueen DA, Schottel JL.** 1987. Purification and characterization of a novel extracellular esterase from pathogenic *Streptomyces scabies* that is inducible by zinc. J. Bacteriol. **169**:1967–1971.
2. **Casas-Godoy L, Duquesne S, Bordes F, Sandoval G, Marty A.** 2012. Lipases: an overview. Methods Mol. Biol. **861**:3–30.
3. **Fan X, Niehus X, Sandoval G.** 2012. Lipases as biocatalyst for biodiesel production. Methods Mol. Biol. **861**:471–483.
4. **Hudlicky T, Reed JW.** 2009. Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. Chem. Soc. Rev. **38**:3117–3132.
5. **Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR.** 2004. The Pfam protein families database. Nucleic Acids Res. **32**:D138–D141.
6. **Desai N, Antonopoulos D, Gilbert JA, Glass EM, Meyer F.** 2012. From genomics to metagenomics. Curr. Opin. Biotechnol. **23**:72–76.
7. **Galperin MY.** 2008. Sorting out the mix in microbial genomics. Environ. Microbiol. **10**:3187–3192.
8. **Sharma VK, Kumar N, Prakash T, Taylor TD.** 2010. MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. Nucleic Acids Res. **38**:D468–D472.
9. **Gomez-Alvarez V, Teal TK, Schmidt TM.** 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. **3**:1314–1317.
10. **Fernández-Arrojo L, Guazzaroni ME, López-Cortés N, Beloqui A, Ferrer M.** 2010. Metagenomic era for biocatalyst identification. Curr. Opin. Biotechnol. **21**:725–733.
11. **Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni ME, Polaina J, Strittmatter AW, Reva O, Waliczek A, Yakimov MM, Golyshina OV, Ferrer M, Golyshin PN.** 2010. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl. Environ. Microbiol. **76**:5934–5946.
12. **Beloqui A, Pita M, Polaina J, Martínez-Arias A, Golyshina OV, Zumárraga M, Yakimov MM, García-Arellano H, Alcalde M, Fernández VM, Elborough K, Andreu JM, Ballesteros A, Plou FJ, Timmis KN, Ferrer M, Golyshin P-N.** 2006. Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties structural analysis and phylogenetic relationships. J. Biol. Chem. **281**:22933–22942.
13. **Beloqui A, Polaina J, Vieites JM, Reyes-Duarte D, Torres R, Golyshina OV, Chernikova TN, Waliczek A, Aharoni A, Yakimov MM, Timmis KN, Golyshin PN, Ferrer M.** 2010. Novel hybrid esterase-haloacid dehalogenase enzyme. Chembiochem **11**:1975–1978.
14. **Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD.** 2012. The great screen anomaly—a new frontier in product discovery through functional metagenomics. Appl. Microbiol. Biotechnol. **93**:1005–1020.
15. **Ellrott K, Jaroszewski L, Li W, Wooley JC, Godzik A.** 2010. Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. PLoS Comput. Biol. **6**:e1000798. doi:10.1371/journal.pcbi.1000798.
16. **Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins Dos Santos VA, Yakimov MM, Timmis KN, Golyshin PN.** 2005. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. Chem. Biol. **12**:895–904.
17. **Gabor E, Niehaus F, Aehle W, Eck J.** 2012. Zooming in on metagenomics: molecular microdiversity of subtilisin Carlsberg in soil. J. Mol. Biol. **418**:16–20.
18. **Lei Y, Peng H, Wang Y, Liu Y, Han F, Xiao Y, Gao Y.** 2012. Preferential

and rapid degradation of raw rice starch by an α-amylase of glycoside hydrolase subfamily GH13_37. Appl. Microbiol. Biotechnol. **94**:1577–1584.
19. **Nagarajan S.** 2012. New tools for exploring "old friends-microbial lipases." Appl. Biochem. Biotechnol. **168**(5):1163–1196.
20. **Park SY, Shin HJ, Kim GJ.** 2011. Screening and identification of a novel esterase EstPE from a metagenomic DNA library. J. Microbiol. **49**:7–14.
21. **Pham VD, Palden T, DeLong EF.** 2007. Large-scale screens of metagenomic libraries. J. Vis. Exp. **4**:201.
22. **Vieites JM, Ghazi A, Beloqui A, Polaina Jreu Golyshina JMOV, Nechitaylo TY, Waliczek A, Yakimov MM, Golyshin PN, Ferrer M.** 2010. Inter-conversion of catalytic abilities in a bifunctional carboxyl/feruloyl-esterase from earthworm gut metagenome. Microb. Biotechnol. **3**:48–58.
23. **Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Joachimiak MP, van Belle C, Chandonia JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC.** 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol. **5**:e16. doi:10.1371/journal.pbio.0050016.
24. **Beier K, Jones CM, Mohit V, Hallin S, Betrtilsson S.** 2011. Global phylogenography of chitinase genes in aquatic metagenomes. Appl. Environ. Microbiol. **77**:1101–1106.
25. **Chu X, He H, Guo C, Sun B.** 2008. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. Appl. Microbiol. Biotechnol. **80**:615–625.
26. **Elend C, Schmeisser C, Leggewie C, Babiak P, Carballeira JD, Steele HL, Reymond JL, Jaeger KE, Streit WR.** 2006. Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl. Environ. Microbiol. **72**:3637–3645.
27. **Sambrook J, Russell DW.** 2001. Molecular cloning: a laboratory manual, 3rd ed, p 6.22. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. **Guillén M, Benaiges MD, Valero F.** 2011. Comparison of the biochemical properties of a recombinant lipase extract from *Rhizopus oryzae* expressed in *Pichia pastoris* with a native extract. Biochem. Eng. J. **54**:117–123.
29. **Reyes-Duarte D, Ferrer M, García-Arellano H.** 2012. Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. Methods Mol. Biol. **861**:101–113.
30. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26**:1107–1115.
31. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
32. **Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni ME, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN.** 2012. Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. PLoS One **7**:e38134. doi:10.1371/journal.pone.0038134.
33. **Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680–685.
34. **Bradford MM.** 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
35. **Janes LE, Löwendahl C, Kazlauskas RJ.** 1998. Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. Chem. Eur. J. **4**:2317–2324.
36. **Ménigaud S, Mallet L, Picord G, Churlaud C, Borrel A, Deschavanne P.** 2012. GOHTAM: a website for genomic origin of horizontal transfers, alignment and metagenomics. Bioinformatics **28**:1270–1271.
37. **Ganesan H, Rakitianskaia AS, Davenport CF, Tümmler B, Reva ON.** 2008. The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. BMC Bioinformatics **9**:333. doi:10.1186/1471-2105-9-333.
38. **Corella JP, El Amrani A, Sigró J, Morellón M, Rico E, Valero-Garcés BL.** 2011. Recent evolution of Lake Arreo, northern Spain: influences of land use change and climate. J. Paleolimnol. **46**:469–485.
39. **Rico E, Chicote A, González ME, Montes M.** 1995. Batimetría y análisis morfométrico del Lago Arreo (N. España). Limnética **11**:55–58.
40. **Chicote A.** 2004. Limnología y ecología microbiana de un lago kárstico evaporítico: el lago de Arreo (Norte de España). Ph.D. thesis. Universidad Autónoma de Madrid, Madrid, Spain.
41. **González-Mozo ME, Chicote A, Rico Montes EC.** 2000. Limnological characterizetion of an evaporite karstic lake in Spain (Lake Arreo). Limnética **18**:91–98.
42. **Levisson M, Sun L, Hendriks S, Swinkels P, Akveld T, Bultema JB, Barendregt A, van den Heuvel RH, Dijkstra BW, van der Oost J, Kengen SW.** 2009. Crystal structure and biochemical properties of a novel thermostable esterase containing an immunoglobulin-like domain. J. Mol. Biol. **385**:949–962.
43. **Sheffield PJ, McMullen TW, Li J, Ho YS, Garrard SM, Derewenda U, Derewenda ZS.** 2001. Preparation and crystal structure of the recombinant α_1/α_2 catalytic heterodimer of bovine brain platelet-activating factor acetylhydrolase Ib. Protein Eng. **14**:513–519.
44. **McMullen TW, Li J, Sheffield PJ, Aoki J, Martin TW, Arai H, Inoue K, Derewenda ZS.** 2000. The functional implications of the dimerization of the catalytic subunits of the mammalian brain platelet-activating factor acetylhydrolase (Ib). Protein Eng. **13**:865–871.
45. **Nam KH, Kim M, Kim S, Priyadarshi A, Lee WH, Hwang KY.** 2009. Structural and functional analysis of a novel EstE5 belonging to the subfamily of hormone-sensitive lipase. Biochem. Biophys. Res. Commun. **379**:553–556.
46. **Hotta Y, Ezaki S, Atomi H, Imanaka T.** 2002. Extremely stable and versatile carboxylesterase from a hyperthermophilic archaeon. Appl. Environ. Microbiol. **68**:3925–3931.
47. **Byun J, Rhee J, Kim ND, Yoon J, Kim D, Koh E, Oh J, Cho H.** 2007. Crystal structure of hyperthermophilic esterase EstE1 and the relationship between its dimerization and thermostability properties. BMC Struct. Biol. **7**:47. doi:10.1186/1472-6807-7-47.
48. **Arpigny JL, Jaeger KE.** 1999. Bacterial lipolytic enzymes: classification and properties. Biochem. J. **343**:177–183.
49. **Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF.** 2004. GDSL family of serine esterases/lipases. Prog. Lipid Res. **43**:534–552.
50. **Ghosh N, McKillop TJ, Jowitt TA, Howard M, Davies H, Holmes DF, Roberts IS, Bella J.** 2012. Collagen-like proteins in pathogenic *E. coli* strains. PLoS One **7**:e37872. doi:10.1371/journal.pone.0037872.
51. **Mandrich L, Merone L, Manco G.** 2009. Structural and kinetic overview of the carboxylesterase EST2 from *Alicyclobacillus acidocaldarius*: a comparison with the other members of the HSL family. Protein Pept. Lett. **16**:1189–1200.
52. **Martín-Rubio M, Rodriguez-Lazaro J, Anadón P, Robles F, Utrilla R, Vázquez A.** 2005. Factors affecting the distribution of recent lacustrine ostracoda from the Caicedo de Yuso-Arreo Lake (Western Ebro Basin, Spain). Palaeogeogr. Palaeoclimatol. Palaeoecol. **225**:118–133.
53. **Lemak S, Tchigvintsev A, Petit P, Flick R, Singer AU, Brown G, Evdokimova E, Egorova O, Gonzalez CF, Chernikova TN, Yakimov MM, Kube M, Reinhardt R, Golyshin PN, Savchenko A, Yakunin AF.** 2012. Structure and activity of the cold-active and anion-activated carboxyl esterase OLEI01171 from the oil-degrading marine bacterium *Oleispira antarctica*. Biochem. J. **4452**:193–203.
54. **Biely P.** 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnol. Adv. **30**:1575–1588.
55. **Goswami D, Basu JK, De S.** 2013. Lipase applications in oil hydrolysis with a case study on castor oil: a review. Crit. Rev. Biotechnol. **33**:81–96.
56. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.
57. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5**:113. doi:10.1186/1471-2105-5-113.
58. **Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.** 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–1191.

CAPÍTULO 3
ESTERASAS DEL MAR DE BARENTS

RESUMEN DEL CONTENIDO

Este segundo capítulo se centra en la identificación y caracterización de nuevas hidrolasas de la super-familia de las α/β-hidrolasas, altamente promiscuas. En particular, las enzimas se han aislado de dos fuentes: *i*) una librería de clones metagenómicos de comunidades microbianas presentes en una muestra de agua marina del Mar de Barents (Isla de Kolguev), contaminada con hidrocarburos; y *ii*) una bacteria degradadora de compuestos aromáticos policíclicos (*Cycloclasticus* sp. ME7). Para ello se han utilizado técnicas metagenómicas y genómicas para la identificación de marcos de lectura abiertas que codifican las enzimas de interés. Posteriormente se han empleado técnicas de biología molecular para la expresión en la bacteria *E. coli* de los genes que codifican las enzimas de interés, así como la posterior producción a pequeña escala y purificación con la ayuda de colas de 6-histidinas. El empleo de técnicas bioquímicas, incluidos ensayos espectrofotométricos para la determinación de parámetros óptimos, especificidad de sustrato y cálculo de actividades específicas y constantes cinéticas, permitió posteriormente la caracterización de las mismas. Se han realizado estudios comparativos a nivel de secuencia. Finalmente, y a diferencia del capítulo anterior, se han realizado experimentos de mutagénesis dirigida y cristalización para la búsqueda de residuos que proporcionen capacidades altamente promiscuas.

De forma resumida, en este estudio se describen dos hidrolasas de la super-familia de las α/β-hidrolasas obtenidas tras rastrear con un sustrato modelo (acetato de naftilo) una librería de 14,000 fagos lambda (48 Gpb) de una muestra de agua marina del Mar de Barents. Pese a que las enzimas mostraban homología con esterasas y lipasas descritas en las bases de datos (50-70%), a nivel estructural también presentaban homología (22-28%) con hidrolasas que actúan sobre enlaces C-C, en particular, aquellas que actúan sobre derivados de la degradación de bifenilo (2-hidroxi-6-oxo-6-fenilhexa-2,4-dienoato; *HOPHD*) y catecol (2-hidroxi-6-oxohepta-2,4-dienoato; *HOHD*). La caracterización bioquímica de ambas enzimas reveló la capacidad de las mismas para hidrolizar, en diferente escala, un amplio espectro de sustratos que incluyen 35 ésteres diferentes (derivados de *p*-nitrofenilo, triglicéridos, ésteres de ácidos grasos halogenados y no halogenados, ésteres aromáticos, ésteres de carbohidratos y ácido cinámico, y lactonas), así como *HOPHD* y *HOHD*, hecho éste inédito en la literatura. Posteriormente esta amplia promiscuidad se demostró que ocurría en otras hidrolasas similares, en particular en 3 hidrolasas de la bacteria *Cycloclasticus* sp. ME7. El análisis estructural y la realización de mutagénesis dirigida en residuos cercanos al centro activo permitió la identificación de tres residuos (S32, V130 y W144) capaces de modular la especificidad de sustrato inusual de una de estas proteínas, CCSP0084.

Los resultados presentados en este capítulo arrojan nueva luz sobre la diversidad entre los miembros de la familia de hidrolasas α/β, y proporciona información para mejorar la anotación de proteínas en las bases de datos, y estudios de catálisis comparativa y estudios de modelos evolutivos. Los resultados pueden abrir nuevas vías de investigación relacionadas con la búsqueda de motivos que proporcionen capacidades altamente promiscuas. Las enzimas similares a las descritas en este Capítulo y que aparecen recogidas en la bibliografía presentan poco interés industrial y biotecnológico debido a su restringida especificidad de sustrato, ya que solo son capaces de hidrolizar *HOHD* and *HOHPD*. Sin embargo, las enzimas identificadas en este trabajo poseen un alto grado de promiscuidad siendo capaces de resolver múltiples transformaciones de interés industrial, como resoluciones quirales y transformación de lactonas. Finalmente, el estudio sugiere que las condiciones ambientales, y en particular la actividad antropogénica, pueden favorecer procesos evolutivos que hacen que, a través de mutaciones puntuales, las enzimas puedan modificar su especificidad de sustrato para acceder a nuevas moléculas introducidas de forma antropogénica en el medio ambiente.

El CD adjunto a esta Memoria contiene Material Suplementario relacionado con este Capítulo.

online data

Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family

María ALCAIDE*[1], Jesús TORNÉS*[1], Peter J. STOGIOS†, Xiaohui XU†, Christoph GERTLER‡, Rosa DI LEO†, Rafael BARGIELA*, Álvaro LAFRAYA*, María-Eugenia GUAZZARONI*[2], Nieves LÓPEZ-CORTÉS*, Tatyana N. CHERNIKOVA‡, Olga V. GOLYSHINA‡, Taras Y. NECHITAYLO§, Iris PLUMEIER‖, Dietmar H. PIEPER‖, Michail M. YAKIMOV¶, Alexei SAVCHENKO†[1,3], Peter N. GOLYSHIN‡[1,3] and Manuel FERRER*[1,3]

*Spanish National Research Council (CSIC), Institute of Catalysis, 28049 Madrid, Spain, †Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada, M5S 3E5, ‡School of Biological Sciences, Bangor University, Gwynedd LL57 2UW, U.K., §Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany, ‖Helmholtz Zentrum für Infektionsforschung (HZI), Microbial Interactions and Processes Research Group, 38124 Braunschweig, Germany, and ¶Institute for Coastal Marine Environment (IAMC), CNR, 98122 Messina, Italy

Several members of the C-C MCP (*meta*-cleavage product) hydrolase family demonstrate an unusual ability to hydrolyse esters as well as the MCPs (including those from mono- and bi-cyclic aromatics). Although the molecular mechanisms responsible for such substrate promiscuity are starting to emerge, the full understanding of these complex enzymes is far from complete. In the present paper, we describe six distinct α/β hydrolases identified through genomic approaches, four of which demonstrate the unprecedented characteristic of activity towards a broad spectrum of substrates, including *p*-nitrophenyl, halogenated, fatty acyl, aryl, glycerol, cinnamoyl and carbohydrate esters, lactones, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate and 2-hydroxy-6-oxohepta-2,4-dienoate. Using structural analysis and site-directed mutagenesis we have identified the three residues (Ser^{32}, Val^{130} and Trp^{144}) that determine the unusual substrate specificity of one of these proteins, CCSP0084. The results may open up new research avenues into comparative catalytic models, structural and mechanistic studies, and biotechnological applications of MCP hydrolases.

Key words: C-C bond, carboxylesterase, hydrolase, *meta*-cleavage, metagenomics, promiscuity.

INTRODUCTION

MCP (*meta*-cleavage product) hydrolases from the α/β hydrolase family hydrolyse C-C bonds of vinylogous 1,5-diketones [1]. They have been shown to possess a strict substrate specificity restricted to the catabolism of diverse ring fission products of aromatics. Therefore, they represent a bottleneck for aerobic catabolism of aromatics by bacteria [2]. Most MCP hydrolases can be classified into three groups on the basis of substrate specificity [1]. Group I and II hydrolases preferentially cleave fission products of bicyclic [i.e. HOPHD (2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate)] and monocyclic [i.e. HOHD (2-hydroxy-6-oxohepta-2,4-dienoate)] aromatics respectively, whereas group III prefer heteroaromatics. However, enzymes from *Sphingomonas* sp. strain RW1 [3], and the three best characterized enzymes $BphD_{LA\text{-}4}$ and $MfphA_{LA\text{-}4}$, from *Dyella ginsengisoli* LA-4 [4,5], and BphD, from *Burkholderia xenovorans* [6,7], possessed intermediate features. MfphA preferred (390-fold) HOHD as compared with HOPHD, whereas $BphD_{LA\text{-}4}$ and BphD preferentially hydrolysed HOPHD (427 and 4600-fold respectively).

Research on the reaction mechanism of MCP hydrolases has unambiguously revealed that the catalytic process occurs via a covalent mechanism of catalysis at the carbonyl carbon of an α/β-unsaturated system by generating a serinate for nucleophilic attack [6,7]. However, it is still a challenge to tune the substrate specificity of MCP hydrolases, especially interchanging MCP hydrolases from groups I and II, as it has been presumed that the substrate specificity may be affected by multiple residues [5]. In this context, MCP hydrolases have been described as having three distinct regions in their active site, a hydrophilic P-region (proximal to the presumed entrance to the substrate-binding channel), a hydrophobic D-region (distal to the entrance) [8–11] and an NP subsite [5]. The P-region contains highly conserved residues mediating interactions with the hydrophilic/dienoate end of the substrates, whereas the D-region is more variable and interacts with hydrophobic C6 substrates in a group I- or group II/III-specific manner. Analysis of the crystal structure of BphD further indicates that the NP subsite is involved in binding the phenyl ring of the substrates [5]. Some recent investigations have suggested a number of residues, among these three distinct regions, that may play a key role in determining substrate specificity, either by reshaping the active pocket or by modulating the orientation of the substrate. Indeed, five active-site residues are conserved in all of the MCP hydrolases characterized (including the CCSP0084, CCSP0528 and CCSP2178 proteins

Abbreviations used: HOHD, 2-hydroxy-6-oxohepta-2,4-dienoate; HOPHD, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate; MCP, *meta*-cleavage product; αNA, α-naphthyl acetate; NCBI, National Center for Biotechnology Information; PAH, polycyclic aromatic hydrocarbon; *p*NP, *p*-nitrophenyl.

[1] These authors contributed equally to this work.

[2] Present address: Universidade de São Paulo, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Departamento de Química, Avenida Bandeirantes 3900 Monte Alegre, 14049-901 Ribeirao Preto, SP, Brasil.

[3] Correspondence may be addressed to these authors (email alexei.savchenko@utoronto.ca, p.golyshin@bangor.ac.uk or mferrer@icp.csic.es)

The nucleotide sequence data reported for fosmid clones and sequences of *Cycloclasticus* sp. ME7 hydrolases will appear in the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under accession numbers JX133669, JX133670, JX133672–JX133675.

The structural co-ordinates reported for *Cycloclasticus* sp. ME7 CCSP0084 α/β hydrolase will appear in the PDB under accession code 4I3F.

characterized in the present study), but not in other α/β-hydrolases (including the K_05-6, K_06-5 and CCSP0211 proteins characterized in the present study): Arg^{180}, Asn^{103}, Phe^{162}, Cys^{256} and Trp^{259} (CCSP0084 numbering; see Supplementary Table S1 at http://www.biochemj.org/bj/454/bj4540157add.htm); however, replacement of any of the five residues reduces the rate of C-C bond hydrolysis in BphD [6,7], indicating that these residues are primarily involved in substrate recognition and their role in substrate preference could not be unambiguously defined. Mutations at residues in the P-region (i.e. Leu^{140} and Trp^{144} in CCSP0084) have been suggested to undergo large conformational changes in the catalytic cycle, although available crystal structures are insufficient to clarify their role in substrate preference [12]. Residue Ser^{32} (CCSP0084 numbering), which forms the putative oxyanion hole, has been shown to modulate the characteristics of the MCP hydrolase CumD, but it is not involved in determination of substrate specificity [13]. Phe^{162} and Arg^{180} (CCSP0084 numbering) in MhpC, located in the lid domain, are involved in substrate binding, but do not affect the recognition of MCP-substrate-containing phenyl groups [14]. Finally, computational design and characterization of novel MfphA–BphD hybrids, created by domain swapping, have demonstrated that residue Trp^{144} (CCSP0084 numbering) in the NP subsite may affect the orientation of the substrate through steric hindrance with the aromatic ring of HOPHD [5].

In addition to the ability to hydrolyse fission products of bicyclic and monocyclic aromatics, four MCP hydrolases have been also shown to mediate C-O bond hydrolysis *in vitro*, including $BphD_{LA-4}$ and $MfphA_{LA-4}$ from *Dyella ginsengisoli* LA-4 [4,5], MhpC from *Escherichia coli* [15,16] and BphD [6,7]. $MfphA_{LA-4}$ is able to hydrolyse HOHD, HOPHD and the ester *p*NP (*p*-nitrophenyl) butyrate; $BphD_{LA-4}$ additionally hydrolyses the ester analogue *p*NP benzoate; MhpC is able to hydrolyse ethyl adipate, thioethyladipate and ethyl 2-hydroxypentadienoate; and BphD catalyses the hydrolysis of HOPHD and *p*NP benzoate. The fact that both C-O and C-C substrates are chemically distinct may reflect differences in the mechanisms of hydrolysis, in agreement with the versatility of the catalytic triad in MCP hydrolases [6,7]. Differences in the esterase/thioesterase/MCP hydrolase catalytic function of MhpC have been reported by site-directed mutagenesis [16]. Thus mutations at residues Phe^{162} (CCSP0084 numbering), located in the lid domain, and Trp^{259} (CCSP0084 numbering) to a minor extent, caused decreased C-C cleavage activity while increasing the catalytic preference for thioesterase activity. This agrees with previous observations that both residues differentially affected the recognition of aromatic rings in MCP substrates [5].

In summary, although the catalytic mechanism in MCP hydrolases have been clarified and mutagenic studies have been useful to unravel the various residues that underpin substrate recognition, the finer details that dictate the co-evolution of MCP hydrolases from groups I and II/III as well as the co-evolution of dual esterase and MCP hydrolase activities in single proteins remain unclear. This last point is of special significance given that both enzymes share in many cases a high motility region, the lid domain, that has been shown to play a significant role in determining activity, selectivity and stability both in esterases and MCP hydrolases [5]. The purpose of the present study is two-fold: (i) to elucidate and compare the properties of novel α/β hydrolases with broad substrate specificity, that include the unusual ability to hydrolyse fission products of bicyclic and monocyclic aromatics as well as esters; and (ii) to suggest the molecular features responsible for the activity against an exceptional array of substrates. We believe that our study supports, for first time, that single mutations at a limited number of residues is a mechanism by which dual esterolytic/MCP hydrolytic phenotypes may evolve in single α/β hydrolases.

MATERIALS AND METHODS

General methods and GenBank® accession numbers

Full descriptions of the methods used for the 'naïve' activity screen, gene cloning, site-directed mutagenesis and protein expression and purification are available in the Supplementary Materials and methods section (at http://www.biochemj.org/bj/454/bj4540157add.htm). DNA sequences of fosmid clones and sequences of *Cycloclasticus* sp. ME7 hydrolases were deposited in GenBank® under accession numbers JX133669, JX133670, JX133672–JX133675.

Biochemical assays

Esterase activity using *p*NP esters were assayed at 410 nm as described previously [17], but the buffer used was 40 mM Hepes, pH 7.0. Esterase activity using structurally diverse esters other than *p*NP esters was determined at 540 nm using Phenol Red as a pH indicator and 5 mM Hepps, pH 8.0, as described by Janes et al. [18], with slight modifications [17]. MCP hydrolase activity using freshly prepared HOPHD and HOHD [19–21] was measured at 388 or 434 nm [19,20] respectively in reaction mixtures containing 50 mM potassium/sodium phosphate, pH 7.5.

For determination of kinetic parameters, these were calculated by non-linear curve fitting from the Lineweaver–Burk plot using GraphPad Prism software (version 4.00). Each well contained 0.388–0.882 μM enzyme solution and 0.012–100 mM ester substrate [from a stock solution in acetone (for *p*NP esters) or acetonitrile (for esters others than *p*NP esters)] or 0.01–0.2 mM HOPHD or HOHD (from a 0.2 mM stock solution). For specific activity determinations of CCSP0084 wild-type and mutants using butyl acetate, *p*NP propionate, HOPHD and HOHD, the following conditions were used, using 2 μg of pure enzymes. For butyl acetate, reaction mixtures contained 2 mM substrate (from a 200 mM stock solution in acetonitrile), 0.45 mM Phenol Red and 2.5 % acetonitrile (v/v) in 5 mM Hepps, pH 8.0. For *p*NP propionate, reactions contained 0.8 mM *p*NP ester (from a 80 mM stock solution in acetone) in 20 mM Hepes, pH 7.0. For determination of activity towards HOHD and HOPHD, reactions contained 50 mM potassium/sodium phosphate, pH 7.5, and 0.2 mM HOPHD or HOHD.

If not stated otherwise, reactions were followed at 40 °C every 2 min for a total time of 15 min (except for highly hydrolysed substrates, with which 1 min incubations and 0.2 min intervals were used), in 96-well microtitre plates (total volume of 200 μl) on a Synergy HT Multi-Mode Microplate Reader (BioTek). One unit of enzyme activity was defined as the amount of enzyme hydrolysing 1 μmol of substrate in 1 min under the assay conditions. In all activity assays, three independent experiments were performed for each parameter, substrate and enzyme, and graphs were plotted using the means $\pm$ S.D. It should be highlighted that all values were corrected for non-enzymatic hydrolysis (background rate).

The optimal pH and temperature were determined according to specific activity determinations using *p*NP propionate as described above. pH values between 4.0 and 9.0 and temperatures between 4 and 70 °C were tested. All of the following buffers were tested at 40 mM: sodium citrate (pH 4.0–4.5), sodium acetate (pH 5.0–6.0), Mes (pH 5.5–6.0), Hepes (pH 7.0–8.0), Pipes (pH 6.0–7.0), potassium/sodium phosphate (pH 7.5), Tris/

HCl (pH 8.5) and glycine (pH 9.0–9.5). The pH was always adjusted at 25 °C in all cases. The pH and temperature profiles were obtained at 40 °C and pH 7.0 respectively.

CCSP0084 purification, crystallization and structure determination

The CCSP0084 enzyme was expressed and purified according to the procedures described by Lai et al. [22]. The His_6-tagged enzyme was crystallized by the sitting drop method using Intelliplate 96-well plates and a mosquito® liquid handling robot (TTP LabTech), which mixed 0.5 μl of protein at 17 mg/ml and 0.5 μl of reservoir solution. The reservoir solution was 0.1 M Hepes, pH 7.5, 0.2 M magnesium chloride and 25 % (w/v) PEG [poly(ethylene glycol)] 3350. The crystal was cryo-protected with reservoir solution supplemented with paratone-N oil prior to flash freezing in an Oxford Cryosystems cryostream. Diffraction data were collected at 100 K and the Cu-Kα wavelength was collected at the Structural Genomics Consortium (Toronto) using a Rigaku FR-E Superbright rotating anode with a Rigaku R-AXIS HTC detector. Diffraction data were reduced with XDS [23] and Scala [24] to a high resolution limit of 1.69 Å (1 Å = 0.1 nm).

The structure was determined by Molecular Replacement using a search model of the CCSP0084 sequence, based on the structure of the MCP hydrolase CumD (PDB code 1UK8) [25], generated by the Phyre2 fold recognition server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id = index) and using Phenix.phaser [26]. The structure was manually inspected and modified with Coot [27] and refined with Phenix.refine [28]. *B*-factors were refined as anisotropic for protein atoms and isotropic for non-protein atoms. TLS (Translation–Libration–Screw-rotation) parameterization groups corresponding to residues 0–131, 132–189 and 190–282 were added, as determined by the TLSMD server [29]. Water atoms were added using Phenix.refine, Coot and manual inspection; ions were added using manual inspection. The suitability of Ramachandran angles were validated with Phenix and the PDB Validation Server {residues in most favoured regions: 91.4 %, additionally allowed; 8.2 %, generously allowed; 0.4 %, disallowed (corresponds to catalytic Ser^{104} in the strained turn/nucleophilic elbow structure, conserved in α/β hydrolases [30]}. The final atomic model includes residues 1–282 of the enzyme, plus an extra N-terminal glycine residue (residue 0) from the expression vector. Cys^{70} appears to have been covalently modified and the nature of the modification could not be identified and therefore was not modelled. This residue is far from the active site and is not thought to be functionally important.

Electrostatic surface representation of the crystal structure of CCSP0084 was calculated using the PDB2PQR server [31], using the AMBER forcefield and otherwise default settings. Multiple sequence alignments were conducted using the ClustalW2 tool (http://www.ebi.ac.uk/Tools/msa/clustalw2/), integrated into BioEdit 7.0.9.1 software [32] and Geno3D [33]. The final structure and structural factors were deposited into the PDB under the accession code 4I3F.

RESULTS AND DISCUSSION

Molecular and biochemical features of two dual ester/MCP hydrolases from community genomes

Subsets of 14 000 phage clones from the metagenomic library of Kolguev Island (created from crude-oil enrichment cultures established with sea water at sampling point 68.45 N 49.2 E) [34], which harbours nearly 48 Gbp of community genomes were scored for their ability to hydrolyse αNA (α-naphthyl acetate) and indoxyl acetate [35]. Two positive clones, designated k_05 and k_06, were identified as being active against both substrates. The inserts of k_05 (9087 bp) and k_06 (6636 bp) were sequenced, analysed and compared with the sequences available in the NCBI (National Center for Biotechnology Information) non-redundant public database [36,37]. Among the encoded 11 (in k_05) and six (in k_06) predicted protein-encoding genes, two predicted esterases with the α/β hydrolase fold were identified, which were named based on the source fragment ID number and the number of the corresponding open reading frame: K_05-6 and K_06-5.

BLAST searches of the NCBI non-redundant database with the protein sequence of K_05-6 (319 amino acids; molecular mass 35 487 Da; pI 5.04) and K_06-5 (316 amino acids; molecular mass 34 301 Da; pI 8.44) showed high identity (up to 50/25 %) and similarity (up to 70/36 %) with various gammaproteobacterial and betaproteobacterial carboxylesterases and lipases (Figure 1) respectively. Both enzymes can be classified in the microbial HSL (hormone-sensitive lipase/esterase) family V type, described by Arpigny and Jaeger [38], and contained a -GxSxG- motif and presumptive Ser-Asp-His catalytic triads (Supplementary Table S1). A BLAST search against the PDB database further implied that K_05-6 and K_06-5 are also structurally most similar to the group I MCP hydrolases MhpC (PDB code 1U2E) [11], CarC (PDB code 1J1I) [10], CumD (PDB code 1UK6) [9,12] and BphD (PDB code 2RHT) [39–41] with pairwise sequence identities between 22–28 %, although both K_05-6 and K_06-5 formed separate clusters at the sequence level (Figure 1). Instead, they were closer to carbon-heteroatom hydrolases from the heterogeneous group IV [1].

After expression in *E. coli* and purification, kinetic parameters were determined for K_05-6 and K_06-5 against a battery of 35 structurally diverse esters (for esterolytic activity) and the two ring cleavage products HOHD and HOPHD (for MCP hydrolytic activity). Both enzymes were able to efficiently hydrolyse *p*NP esters ranging from acetate to decanoate (but not longer esters), with *p*NP octanoate being the preferred substrate (Figure 2 and Supplementary Figure S1 at http://www.biochemj.org/bj/454/bj4540157add.htm); however, whereas the current analysis also confirmed activity of K_05-6 on methyl-bromo-isobutyrate, methyl-bromo-acetate, methyl caproate, ethyl butyrate and tributyrin, no other esters tested were hydrolysed by K_06-5. Conversion of both MCP substrates was also evident (Figure 2 and Supplementary Figure S1). According to catalytic efficiencies for the best ester and MCP substrates, under our experimental conditions, K_05-6 showed similar activity of MCP hydrolytic to esterolytic activity (ratio $\sim$1), whereas K_06-5 was characterized by a higher ratio of MCP hydrolytic to esterolytic activity ($\sim$8.5). In addition, although the K_05-6 enzyme prefers the HOHD substrate, K_06-5 prefers HOPHD (Figure 2). Temperature and pH optima, under our experimental conditions, are provided in Supplementary Table S2 (at http://www.biochemj.org/bj/454/bj4540157add.htm).

Cycloclasticus sp. ME7 as a case: exploring substrate specificity of other α/β hydrolases

Members of the genus *Cycloclasticus* (Gammaproteobacteria, Piscirickettsiaceae) have been identified as key micro-organisms for the aerobic breakdown of PAHs (polycyclic aromatic hydrocarbons) [42]. The genus name refers to its 'ring-breaking' activity, i.e. its ability to degrade the PAH ring to form simpler products [43]. A 'Mediterranean' species, named *Cycloclasticus* sp. strain ME7, has been isolated from contaminated sediments

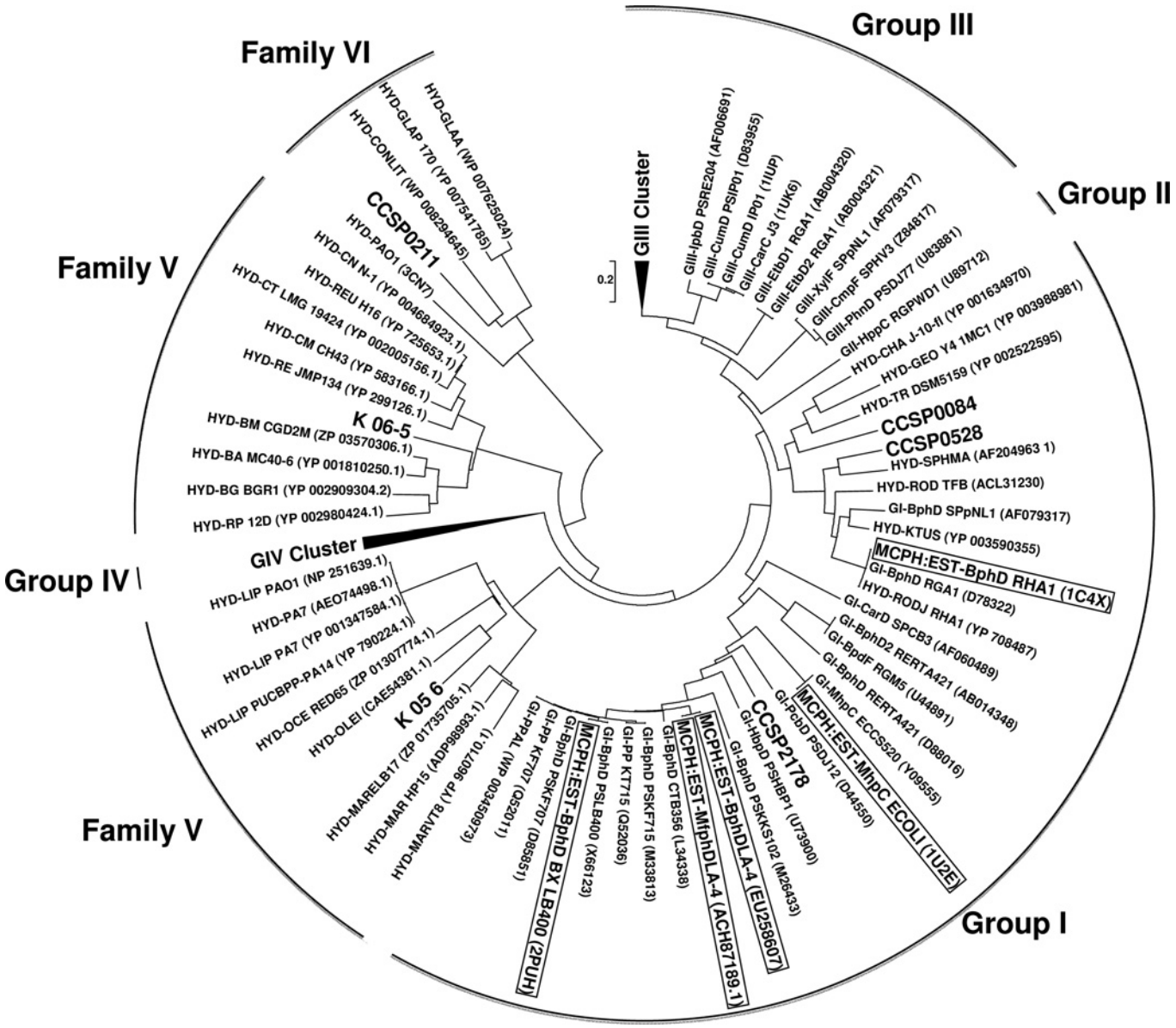


Figure 1 The unrooted circular neighbour-joining tree indicating phylogenetic positions of polypeptide sequences of enzymes characterized in the present study (in bold) and reference hydrolases

GenBank® and PDB accession numbers are indicated in brackets. For dendogram construction details, see the Supplementary Materials and methods section (at http://www.biochemj.org/bj/454/bj4540157add.htm). Clusters GIV and GIII include GIV-DhlA XAGJ10 (M26950), GIV-Eph COC12 (AJ224332), GIV-EphX2 HS (L05779), GIV-Nap BS168 (AB001488), GIV-PcaD ACADP1 (L05770), GIV-CatD ACADP1 (AF009224) and GIII-NahN PSAN10 (AF039534), GIII-XylF PSpWW0 (M64747), GIII-DmpD PSCF600 (X52805), GIII-AtdD ACYAA (AB008831), GIII-McbF RAJS705 (AJ006307), GIII-TecF BUPS12 (U78099) and GIII-TodF PSF1 (Y18245) respectively, where the code in parentheses is the accession number. Scale bar represents 0.2 substitutions per position. MCP hydrolase (groups I, II, III and IV) and lipase/esterase (Family V and VI) families are depicted according to the Hernáez et al. [1] and Arpigny and Jaeger [38] classifications respectively. MCP hydrolases which have been shown to hydrolyse both fission products of bicyclic and/or monocyclic aromatics as well as esters *in vitro* are boxed.

of the Messina (Italy) harbour and its genome sequenced (NCBI accession number CP005996). Four (CCSP0084, CCSP0211, CCSP0528 and CCSP2178) out of six genes encoding potential α/β hydrolases that share significant homology with the K_05-6 and K_06-5 proteins ($\geqslant$31 % amino acid sequence identity), were cloned, expressed and biochemically characterized; all four *E. coli* clones expressing those genes demonstrated esterase activity on agar plates containing αNA as a substrate.

BLAST searches against the PDB revealed that CCSP0528 (287 amino acids; molecular mass 32 498 Da; pI 5.78) is most closely related to BphD from the *Rhodococcus jostii* RHA1 (PDB 1C4X; 56 % identity and 76 % similarity) [44]. The sequence of CCSP0084 (282 amino acids; molecular mass 31 547 Da; pI 5.36) most closely matches the sequence of CumD (PDB 2D0D; 38 % identity and 59 % similarity) [9]. The sequence of CCSP2178 (282 amino acids; molecular mass 31 354 Da; pI 5.37) best matches the sequence of BphD from *Burkholderia xenovorans* LB400 (PDB 2PUH; 62 % identity and 78 % similarity) [41]. Finally, CCSP0211 (223 amino acids; molecular mass 24 190 Da; pI 4.66) is most similar to the carboxylesterase PA3859 from *Pseudomonas aeruginosa* (PDB 3CN7; 41 % identity, 55 % similarity) [45]. In addition, CCSP0084, CCSP0211, CCSP0528

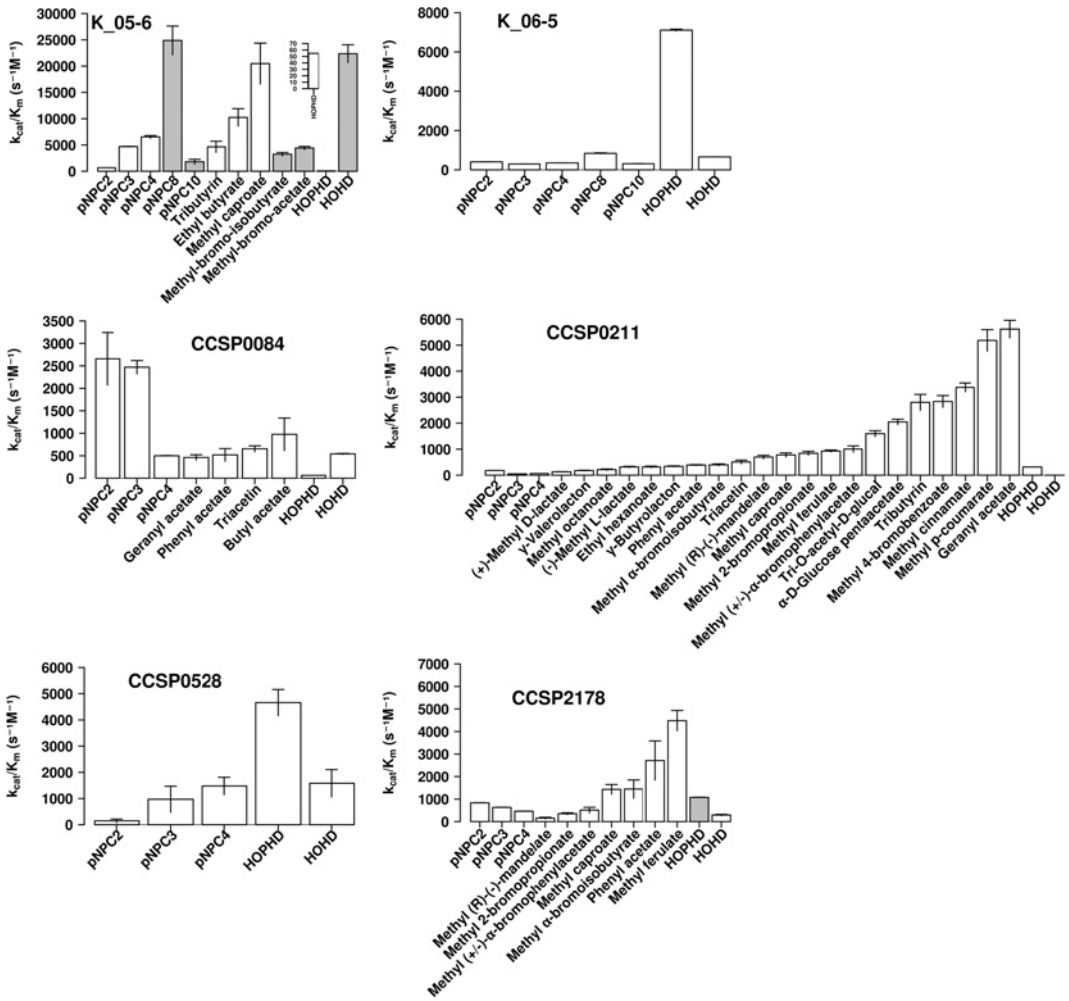


Figure 2 Substrate profiles of the wild-type α/β hydrolases from Kolguev Island metagenome and *Cycloclasticus* sp. ME7 against a set of structurally diverse esters and ring fission products

The catalytic efficiencies (k_{cat}/K_m) were calculated as described in the Materials and methods section in triplicate at 40 °C, and 20 mM Hepes, pH 7.0 (for *p*NP esters), 5 mM Hepps, pH 8.0 (for esters others than *p*NP esters), and 50 mM potassium/sodium phosphate, pH 7.5 (for HOHPD and HOHD). Mean values are given with the standard deviations. Note: grey-filled bars indicate those substrates whose catalytic efficiencies are significantly much higher and thus the values given are in $min^{-1} \cdot M^{-1}$ instead of $s^{-1} \cdot M^{-1}$. For significantly low data, a scaled-up insert is provided.

and CCSP2178 share 31, 37, 53 and 53 % amino acid sequence identity with K_05-6 and 57, 44, 66 and 47 % with K_06-5. Presumptive catalytic triads of all four hydrolases were identified, which are listed in Supplementary Table S1. The positioning of the four sequences was examined in a phylogenetic tree. As shown in Figure 1, we made the following observations: (i) CCSP0084, CCSP0528 and CCSP2178 cluster together with group I MCP hydrolases [1]; and (ii) CCSP0021 forms a separate group, most closely related to K_06-5 at the sequence level than to any known MCP hydrolase. Moreover, sequence analysis confirmed that CCSP0211 belongs to the lipase/esterase family VI [38]; with a molecular mass in the range 23–26 kDa, the enzymes in Family VI are among the smallest esterases known [38], in agreement with the theoretical molecular mass of CCSP0211, which is the smallest (24.2 kDa) among the hydrolases described in the present paper.

All purified *Cycloclasticus* sp. ME7 proteins exhibit significant activity (for optimal parameters, see Supplementary Table S2) towards *p*NP esters ranging from acetate to butyrate, but not longer esters (Figure 2 and Supplementary Figure S1). CCSP0528 is unable to hydrolyse any of the 30 esters that were tested further. In contrast, the CCSP2178 enzyme hydrolyses phenyl acetate, methyl caproate, methyl ferulate, methyl (*R*)-(−)-mandelate and three halogenated fatty acid esters; methyl ferulate was the preferred substrate (Figure 2 and Supplementary Figure S1). The CCSP0084 enzyme shows considerable activity towards phenyl acetate, butyl acetate, triacetin and geranyl acetate; *p*NP acetate was thus the preferred substrate. Finally, CCSP0211 exhibits remarkable activity towards a set of 21 out of 30 additional ester substrates tested (Figure 2 and Supplementary Figure S1), including four halogenated fatty acid esters, three cinnamoyl-like esters, two carbohydrate esters, two lactones, two triacylglycerols, three alkyl esters [including (+)-methyl-D-lactate] and two aryl esters [including methyl (*R*)-(−)-mandelate]; geranyl acetate is the preferred substrate. Note that it is both CCSP2178 and CCSP0211 were also enantio-selective as no appreciable activity, under our assay conditions, was detected for methyl (*S*)-(+)-mandelate (for CCSP2178 and CCSP0211) and (−)-methyl-L-lactate (for CCSP0211).

Enzymes CCSP0084, CCSP0528 and CCSP2178 also exhibit significant activity towards both HOHD and HOPHD (Figure 2 and Supplementary Figure S1), with activity ratios of ∼8.6,

$\sim$0.34 and $\sim$0.005 respectively. Along with esters, CCSP0211 is unable to hydrolyse HOHD, but it hydrolyses HOPHD (Figure 2 and Supplementary Figure S1). Considering the best MCP and ester substrates, CCSP2178 and CCSP0528 showed an improved ratio of MCP hydrolytic to esterolytic activity ($\sim$14.3 and $\sim$3.2 respectively), whereas CCSP0211 and CCSP0084 did show preference for esterolytic activity (ratio of $\sim$0.06 and $\sim$0.2 respectively).

On the basis of these results, CCSP0084, CCSP0211, CCSP0528 and CCSP2178 could all be considered as MCP hydrolases with broad substrate specificity, similar to the K_05-6 and K_06-5 proteins (a schematic representation of the substrates being hydrolysed is further provided in Supplementary Figure S2 at http://www.biochemj.org/bj/454/bj4540157add.htm). Moreover, whereas CCSP0528 and K_06-5 exhibit the most restrictive substrate range, CCSP0211, CCSP2178 and K_05-6 have broader substrate specificity, which is unprecedented as compared with other previously reported esterases and MCP hydrolases of the α/β hydrolase fold [4–7,15,16]. Indeed, the present study demonstrated the capacity of a 'presumptive' (on the basis of BLAST similarities) carboxylesterase such as CCSP0211 to perform both esterase and MCP hydrolase activities, in agreement with the fact that family VI esterases have been demonstrated to show wide substrate specificity [17,38]. Additionally, five 'presumptive' (according to BLAST and/or PDB similarities) MCP hydrolases, K_05-6, K_06-5, CCSP0084, CCSP0528 and CCSP2178, were also shown to posses a similar dual character, albeit they differ in their specific substrate profiles. This in turns suggests that 'presumptive' homologous MCP hydrolases may be highly efficient carboxylesterases with unusual and exceptionally broad substrate specificity; however, further experimental evidence is required to test this hypothesis. Degradation of monocyclic and bicyclic fission products was shown to be preferably supported by K_05-6/CCSP0084 and K_06-5/CCSP0211/CCSP2178-related sequences respectively, whereas sequences similar to CCSP0528 may be equally efficient for both. Noteably, the results of the present study suggest that no clear association exists between the phylogenetic/sequence positioning (Figure 1) and the activity relationships (Figure 2). However, we noticed that all previously described MCP hydrolases capable of mediating C-O bond hydrolysis *in vitro* are positioned within the Group I, where three of our enzymes are also clustered.

Structural analysis of *Cycloclasticus* sp. ME7 CCSP0084

To gain more molecular insight into the activity of the *Cycloclasticus* sp. ME7 α/β-hydrolases, we determined the crystal structure of the CCSP0084 enzyme (Supplementary Table S3 at http://www.biochemj.org/bj/454/bj4540157add.htm). An attempt at crystallizing CCPS0211, CCSP0528 and CCSP2178 was carried out, but no crystals of sufficient quality were obtained for these enzymes. The CCSP0084 crystal structure was determined to 1.69 Å resolution and was solved by Molecular Replacement using the structure of MCP hydrolase CumD (PDB 1UK8). Structural analysis confirmed that CCSP0084 adopts the α/β hydrolase fold typical of other enzymes in this class (Figure 3A) with the core α/β domain and an additional 4 α-helical subdomain (spanning residues 135–200), which is positioned as a lid partially covering access to the active-site cavity (Figure 3A).

Overall, the structure of CCSP0084 most closely resembles the structures of previously characterized MCP hydrolases, such as CumD [9], MhpC [11], CarC [10] and BhpD [41] (Supplementary Figure S3 at http://www.biochemj.org/bj/454/bj4540157add.htm). The pairwise RMSD value for superimposition of CCSP0084 with CumD structure is 0.8 Å over 183 matching Cα atoms spanning both core and lid domains.

Analysis of the CCSP0084 active side cavity reveals a well-defined positively charged channel leading into the presumed active site featuring the typical α/β-hydrolase catalytic triad residues Ser^{104}, Asp^{230} and His^{258} and an oxyanion hole (backbone amides of Ser^{32} and Gly^{33}) (Figure 3C and Supplementary Figure S3B). Alanine substitution of Ser^{104} abrogated the activity (results not shown) for both esterase and MCP substrates (on the basis of specific activity determinations using purified variants), which confirmed the proposed role of this residue as an essential nucleophile in catalysis towards both types of substrates. The oxyanion hole in CCSP0084 structure is occupied by a chloride atom that most likely originated from the enzyme's purification buffer, a phenomenon also observed in the case of the CumD structure (PDB code 2D0D) [13]. Another chloride atom is found in the active site of CCSP0084 (Figure 3C). The position of this chloride atom corresponds to that of the carboxylate of 3-Cl HOPHD bound to BhpD (PDB code 2RHT) [39].

The P-region of the CCSP0084 active site is made up of residues highly conserved across reported MCP hydrolases and three of the hydrolases characterized in the present study (Supplementary Table S1) included Gly^{31}, Gly^{33}, Asn^{41}, Asn^{103} and Arg^{180}, suggesting that these CCSP0084 residues contribute to the hydrogen bond network with substrates such as HOHD, similar to the network observed in the BphD/HOHD complex structure (Figure 3C). In contrast, the CCSP0084 residues defining the D-region demonstrate significant variation when compared with the equivalent regions in the other MCP hydrolases (Figure 3B). As a result, the active site of CCSP0084 appears more constricted when compared with the 3-Cl HOPHD-bound BphD structure, especially in the region of BphD that interacts with the 6-oxo-6-phenyl region of the substrate (Figure 3C, in surface representation). In particular, CCSP0084 residues Val^{130}, Trp^{144} and Ile^{233} in the D-region of the active site are responsible for constricting the size of this region of the active site. These features may explain the preference of CCSP0084 for substrates such as HOHD over HOPHD or other molecules that feature a bulky C^6 aromatic group (Figure 2). Note that a comparative analysis further indicates that the D-region of CumD (which also prefers fission products from monocyclic aromatics [13]) is not constricted. There are sequence differences between CumD and CCSP0084 in the D-region, but two prominent residues that may be responsible for this are CumD Ala^{129} against CCSP0084 Val^{130} and CumD Val^{227} against CCSP0084 Ile^{233}. Modelling of these two CCSP0084 residues into CumD has a large effect on constricting the size of the D-region of this enzyme (Supplementary Figure S4 at http://www.biochemj.org/bj/454/bj4540157add.htm). Subsequently we analysed the molecular architecture of the CCSP0084 active site with respect to its ability to catalyse the hydrolysis of diverse esterase substrates. On the basis of the assessment that the D-region of the CCSP0084 active site appears too constricted to accomodate the *p*NP moiety or long acyl chains, we suggest that these larger regions of the substrates are orientated towards the P-region of the active site. This particular feature of the CCSP0084 D-region may be responsible for the lower catalytic efficiency of this enzyme toward the butyrate and larger *p*NP compounds (Figure 2). Along the same line, the specific molecular environment of the CCSP0084 P-region of the active site may restrict substrate accessibility to butyl acetate (or smaller) esters, explaining the lower catalytic efficiency values for geranyl acetate and glyceryl triacetate, which contain longer alkyl chains.

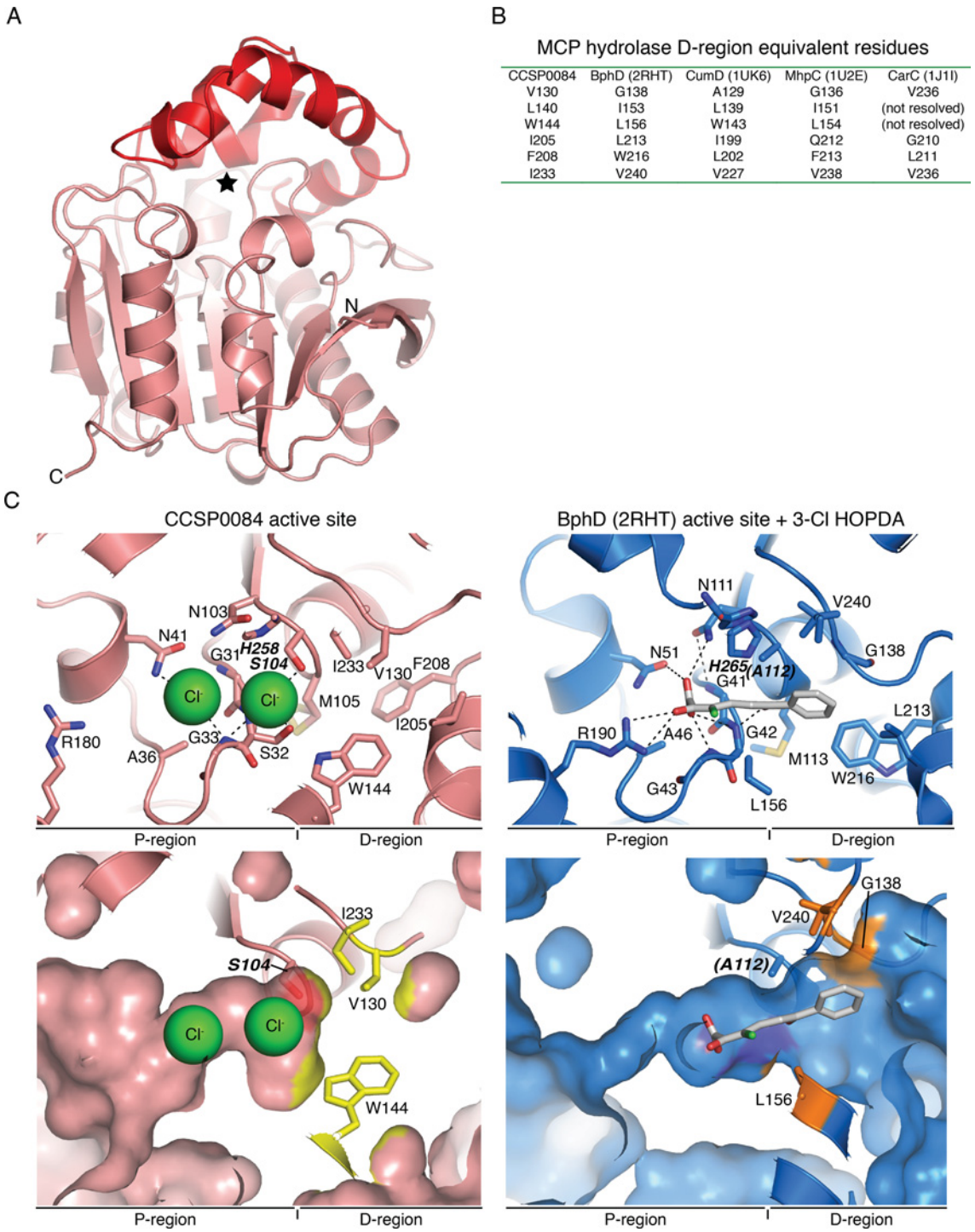


Figure 3 Crystal structure of CCSP0084

(**A**) Overall structure of the enzyme, with the inserted α-helical subdomain coloured red. The star indicates the position of the active site (PDB code 4I3F). (**B**) Comparison of residues found in the D-region of the active site of MCP hydrolases. (**C**) Comparison of the active site of CCSP0084 (left-hand side) with the structural homologue BphD in complex with the substrate analogue 3-Cl HOPHD (PDB code 2RHT, right-hand side). In the top images, all side chains forming interactions with substrates (from PDB code 2RHT) are shown as sticks. In the bottom images, enzymes are shown in surface representation to illustrate the difference in accessibility of the active-site clefts. Residues coloured yellow (in CCSP0084) and orange (in 2RHT) differ in sequence, are positioned to alter the accessibility of the active-site cleft and are suggested to play a role in substrate specificity. Chloride ions in CCSP0084 are shown as green spheres.

We subsequently investigated the role of Val^{130}, Trp^{144} and Ile^{233} residues in substrate recognition by site-directed mutagenesis (Figure 4 and Supplementary Table S4 at http://www.biochemj.org/bj/454/bj4540157add.htm), on the basis of specific activity determinations using purified variants. The V130G mutation, chosen on the basis of the presence of a glycine residue at the corresponding position in BphD enzyme (PDB code 2RHT), abrogated CCSP0084 activity nearly completely towards both esterase and MCP substrates, suggesting that this residue is critical for recognition of both types of substrates.

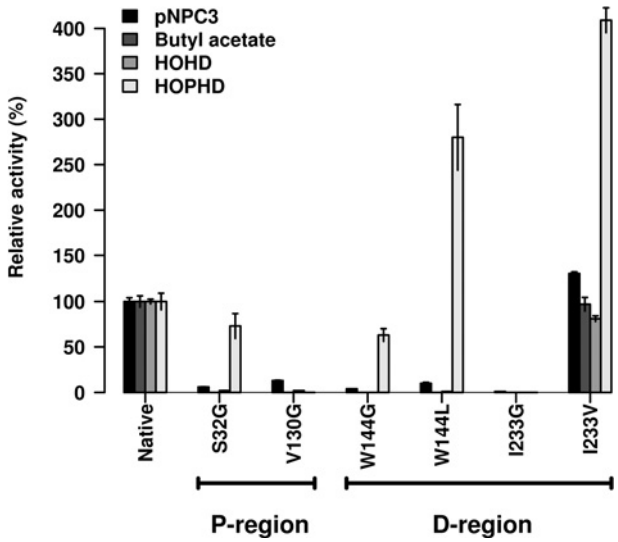


Figure 4 Esterolytic and MCP hydrolytic specific activities for CCSP0084 mutant variants determined using *p*NP propionate, butyl acetate, HOHD and HOPHD

Conditions were as described in the Materials and methods section and Figure 2. Three independent experiments were performed for each parameter and results are given as means ± S.D. Relative specific activity was represented, with 100 % referring to the activity of wild-type CCSP0084 enzyme: 2612 ± 108 (for *p*NP propionate; at 40 °C and pH 7.0), 2025 ± 126 (for butyl acetate; at 40 °C and pH 8.0), 15828 ± 411 (for HOHD; at 40 °C and pH 7.5) and 2.2 ± 0.2 (for HOPHD; at 40 °C and pH 7.5) units · g^{-1}.

It is of significance that the longer butyl acetate and HOPHD substrates were more susceptible (0 % relative activity) to the V130G mutation than *p*NP propionate (~13 % relative activity) and HOHD (~2.1 % relative activity). The W144G and W144L substitutions had a similar major effect on CCSP0084 esterase activity, with activity towards butyl acetate (0 % relative activity) being again more sensitive to these mutations than towards the *p*NP propionate (<10 % relative activity). In contrast, although the W144G and W144L CCSP0084 variants were completely or nearly completely inactive (~1.2 % relative activity) towards the HOHD substrate, they demonstrated comparable (~70 % relative activity for W144G) or improved (~280 % relative activity for W144L) activity for the HOPHD substrate. These results agree with the fact that BphD, which features a leucine residue at the corresponding position, demonstrates a high level of activity for *p*NP benzoate and HOPHD, but does not hydrolyse either alkyl esters or HOHD [6,7]. Finally, the I233V mutation, chosen based on the presence of a valine residue at the corresponding position of BphD, did not adversely affect CCSP0084 activity towards ester and MCP substrates. Remarkably, the CCSP0084 I233V variant demonstrates significantly (~409 %) higher activity towards the HOPHD substrate than the wild-type enzyme. By contrast, a more intrusive I233G substitution suppresses the enzymes's activity nearly completely in the case of both esterase and MCP substrates.

In the CCSP0084 structure, the Ser^{32} side chain forms an interaction with the chloride ion positioned in the oxyanion hole. This position of Ser^{32} would enable interactions with the additional oxygen of potential substrates. In line with this hypothesis, the mutation of Ser^{32} to glycine severely abrogated activity towards both ester and HOHD substrates (Figure 4). On the other hand, the S32G variant retains significant (~70 % relative activity) activity for the HOPHD substrate. Combined with structural analysis these data suggest that the serine residue side chain plays a role in catalysis as well as in substrate preference.

To conclude, the present study provides structural and biochemical evidence of the important roles played by specific active-site residues in the dual esterolytic/MCP hydrolytic activity of α/β hydrolases. Using genomics approaches we have identified six α/β hydrolases with an unusually broad activity spectrum, including multiple ester and MCP substrates. Four of these enzymes feature a particularly diverse activity profile, which includes the ability to hydrolyse substrates that include *p*NP, halogenated, fatty acyl, aryl, glycerol, cinnamoyl and carbohydrate esters, lactones, HOPHD and/or HOHD. Using structural and mutational analysis we identified specific positions in the active site (occupied by residues Ser^{32}, Val^{130} and Trp^{144} in CCSP0084) that are critical for controlling the dual activity and the substrate preference of this type of enzyme. Our results demonstrate that CCSP0084 activity towards butyl acetate is more sensitive to alteration of Ser^{32}, Val^{130} and Trp^{144} residues than activity towards *p*NP propionate, suggesting that these residues are involved in interactions with ester substrates with longer alkyl chains. In addition, these three residues were found to significantly alter the activity toward the MCP substrates, with substrates containing a phenyl ring being most susceptible to alteration of CCSP0084 Val^{130}. This prompted us to hypothesize that this residue is involved in hydrophobic interactions with the phenyl ring of the MCP substrates. In agreement with this, the superimposition of the S112/HOPHD complex with the structure of CCSP0084 (Figure 3C) places the phenyl ring of the substrate in proximity to Val^{130} of CCSP0084.

The structural and biochemical analyses of six α/β hydrolases shown in the present study shed new light on the diversity among members of important family of α/β hydrolases, providing insights for improving protein annotation in databases [46] and comparative catalysis and evolutionary model studies. Our data suggest that hydrolases with such a broad substrate range may persist in microbial communities to a much greater extent than previously anticipated, expanding the pool of substrates available to micro-organisms that mineralize organic carbon. Hence this type of enzyme may contribute to the global carbon cycling processes and for chanelling complex substrates into the common catabolic pathways, including recalcitrant organic pollutants. This was further proven by genomic context examination, which showed that genes encoding both CCSP0084 and CCSP0528 were located close to genomic signatures associated with degradation: CCSP0084 is surrounded by aromatic dioxygenases (large and small subunits) and an alkane 1-monoxygenase (in opposite directions), and CCSP0528 is surrounded by an N-methylhydantoinase A/acetone carboxylase (β-subunit), a 2-keto-4-pentenoate hydratase (probably involved in benzoate and xylene degradation) and a 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase (in the same direction). No pollutant-degrading genomic signatures were associated with CCSP0211 [surrounded by a metal-dependent hydrolase (sulfatase-like) and a diaminopimelate decarboxylase in opposite directions] and CCSP2178 (surrounded by isopropylmalate dehydratase subunits and an α/β hydrolase family protein in opposite directions), suggesting that the participation in biodegradation of these enzymes could not be anticipated by genomic analysis, but rather by biochemical testing. Additionally, we observed further that a significant proportion of proteins (annotated as hydrolases or acyltransferases), which share a high level of sequence identity with the proteins reported in the present study, possibly originated from organisms with known biodegradation capacities or were located in clusters containing potential genomic signature associated with degradation (results not shown).

Finally, although it is true that we have evaluated the activity of the enzymes reported in the present study with a wide array of substrates, it is also plausible that some of the observed activities are characteristics of other MCP hydrolases and esterases/lipases. In this context, further experimental evidence is required to test whether some related characterized MCP hydrolases, such as BphD, MhpC or CumD, do possess similar broad activities. Additionally, homologous, but not characterized, α/β hydrolases may also be included in future studies. This last issue should be of special significance given the fact that three of the hydrolases (K_05-6, K_06-5 and CCSP0211), reported in the present study as capable of mediating C-O and C-C hydrolysis, do form a separate cluster (most similar to Family V and VI of esterases/lipases) as compared with, and do not contain highly conserved residues (see Supplementary Table S1) characteristics of, known characterized MCP hydrolases.

AUTHOR CONTRIBUTION

María Alcaide, Jesús Tornés, Peter Stogios, Xiaohui Xu, Christoph Gertler, Rosa Di Leo, María-Eugenia Guazzaroni, Nieves López-Cortés, Tatyana Chernikova, Olga Golyshina, Iris Plumeier, Alexei Savchenko, Peter Golyshin and Manuel Ferrer designed the experiments. María Alcaide, Jesús Tornés, Peter Stogios, Rafael Bargiela, Álvaro Lafraya, Taras Nechitaylo, Dietmar Pieper, Michail Yakimov, Alexei Savchenko and Manuel Ferrer analysed the data. Manuel Ferrer, Alexei Savchenko and Peter Golyshin wrote the paper.

ACKNOWLEDGEMENTS

We thank Dr Julio Polaina for his excellent support in relation to the classification analysis of K_05-6, K_06-5 and CCSP0211 hydrolases.

FUNDING

The authors gratefully acknowledge the financial support provided by the European Community projects MAGICPAH (Molecular Approaches and MetaGenomic Investigations for optimizing Clean-up of PAH contaminated sites) [number FP7-KBBE-2009-245226], ULIXES (Unravelling and exploiding Mediterranean Sea microbial diversity and ecology for xenobiotics' and pollutants' clean up) [number FP7-KBBE-2010-266473] and MAMBA (Marine Metagenomics for New Biotechnological Applications) [number FP7-KBBE-2008-226977]. This work was further funded by the Spanish Ministry of Economy and Competitiveness (formerly MICINN) [grant number BIO2011-25012]. M.-E.G. thanks the CSIC (Consejo Superior de Investigaciones Científicas) for a JAE (Junta para la Ampliación de Estudios) fellowship.

REFERENCES

1 Hernáez, M. J., Andújar, E., Ríos, J. L., Kaschabek, S. R., Reineke, W. and Santero, E. (2000) Identification of a serine hydrolase which cleaves the alicyclic ring of tetralin. J. Bacteriol. **182**, 5448–5453
2 Seah, S. Y., Labbé, G., Nerdinger, S., Johnson, M. R., Snieckus, V. and Eltis, L. D. (2000) Identification of a serine hydrolase as a key determinant in the microbial degradation of polychlorinated biphenyls. J. Biol. Chem. **275**, 15701–15708
3 Bünz, P. V., Falchetto, R. and Cook, A. M. (1993) Purification of two isofunctional hydrolases (EC 3.7.1.8) in the degradative pathway for dibenzofuran in *Sphingomonas* sp. strain RW1. Biodegradation **4**, 171–178
4 Li, A., Qu, Y. Y., Zhou, J. T., Ma, F., Zhou, H. and Shi, S. N. (2010) Characterization of a novel meta-fission product hydrolase from *Dyella ginsengisoli* LA-4. Process Biochem. **45**, 94–100
5 Zhou, H., Qu, Y. Y., Kong, C. L., Wu, Y. G., Zhu, K., Yang, J. and Zhou, J. T. (2012) Promiscuous esterase activities of the C-C hydrolases from *Dyella ginsengisoli*. Biotechnol. Lett. **34**, 1107–1113
6 Ruzzini, A. C., Ghosh, S., Horsman, G. P., Foster, L. J., Bolin, J. T. and Eltis, L. D. (2012) Identification of an acyl-enzyme intermediate in a meta-cleavage product hydrolase reveals the versatility of the catalytic triad. J. Am. Chem. Soc. **134**, 4615–4624
7 Ruzzini, A. C., Horsman, G. P. and Eltis, L. D. (2012) The catalytic serine of meta-cleavage product hydrolases is activated differently for C-O bond cleavage than for C-C bond cleavage. Biochemistry **51**, 5831–5840
8 Nandhagopal, N., Yamada, A., Hatta, T., Masai, E., Fukuda, M., Mitsui, Y. and Senda, T. (2001) Crystal structure of 2-hydroxyl-6-oxo-6-phenylhexa-2,4-dienoic acid (HPDA) hydrolase (BphD enzyme) from the *Rhodococcus* sp. strain RHA1 of the PCB degradation pathway. J. Mol. Biol. **309**, 1139–1151
9 Fushinobu, S., Saku, T., Hidaka, M., Jun, S. Y., Nojiri, H., Yamane, H., Shoun, H., Omori, T. and Wakagi, T. (2002) Crystal structures of a meta-cleavage product hydrolase from *Pseudomonas fluorescens* IP01 (CumD) complexed with cleavage products. Protein Sci. **11**, 2184–2195
10 Habe, H., Morii, K., Fushinobu, S., Nam, J. W., Ayabe, Y., Yoshida, T., Wakagi, T., Yamane, H., Nojiri, H. and Omori, T. (2003) Crystal structure of a histidine-tagged serine hydrolase involved in the carbazole degradation (CarC enzyme). Biochem. Biophys. Res. Comm. **303**, 631–639
11 Dunn, G., Montgomery, M. G., Mohammed, F., Coker, A., Cooper, J. B., Robertson, T., Garcia, J. L., Bugg, T. D. and Wood, S. P. (2005) The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism. J. Mol. Biol. **346**, 253–265
12 Fushinobu, S., Jun, S. Y., Hidaka, M., Nojiri, H., Yamane, H., Shoun, H., Omori, T. and Wakagi, T. (2005) A series of crystal structures of a meta-cleavage product hydrolase from *Pseudomonas fluorescens* IP01 (CumD) complexed with various cleavage products. Biosci. Biotechnol. Biochem. **69**, 491–498
13 Jun, S. Y., Fushinobu, S., Nojiri, H., Omori, T., Shoun, H. and Wakagi, T. (2006) Improving the catalytic efficiency of a *meta*-cleavage product hydrolase (CumD) from *Pseudomonas fluorescens* IP01. Biochim. Biophys. Acta **1764**, 1159–1166
14 Li, C., Li, J. J., Montgomery, M. G., Wood, S. P. and Bugg, T. D. (2006) Catalytic role for arginine 188 in the C-C hydrolase catalytic mechanism for *Escherichia coli* MhpC and *Burkholderia xenovorans* LB400 BphD. Biochemistry **45**, 12470–12479
15 Lam, W. W. and Bugg, T. D. (1997) Purification, characterization, and stereochemical analysis of a C-C hydrolase: 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid 5,6-hydrolase. Biochemistry **36**, 12242–12251
16 Li, C., Hassler, M. and Bugg, T. D. (2008) Catalytic promiscuity in the α/β-hydrolase superfamily: hydroxamic acid formation, C-C bond formation, ester and thioester hydrolysis in the C-C hydrolase family. ChemBioChem **9**, 71–76
17 Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F. et al. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo – a metagenomic approach. Appl. Environ. Microbiol. **79**, 3553–3562
18 Janes, L. E., Löwendahl, C. and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. Chem. Eur. J. **4**, 2317–2324
19 Happe, B., Eltis, L. D., Poth, H., Hedderich, R. and Timmis, K. N. (1993) Characterization of 2,2′,3-trihydroxybiphenyl dioxygenase, an extradiol dioxygenase from the dibenzofuran- and dibenzo-p-dioxin-degrading bacterium *Sphingomonas* sp. strain RW1. J. Bacteriol **175**, 7313–7320
20 Junca, H., Plumeier, I., Hecht, H. J. and Pieper, D. H. (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. Microbiology **150**, 4181–4187
21 McKay, D. B., Prucha, M., Reineke, W., Timmis, K. N. and Pieper, D. H. (2003) Substrate specificity and expression of three 2,3-dihydroxybiphenyl 1,2-dioxygenases from *Rhodococcus globerulus* strain P6. J. Bacteriol. **185**, 2944–2951
22 Lai, K. K., Stogios, P. J., Vu, C., Xu, X., Cui, H., Molloy, S., Savchenko, A., Yakunin, A. and Gonzalez, C. F. (2011) An inserted α/β subdomain shapes the catalytic pocket of *Lactobacillus johnsonii* cinnamoyl esterase. PLoS ONE **6**, e23269
23 Kabsch, W. (2010) XDS. Acta Crystallogr., Sect. D: Biol. Crystallogr. **66**, 125–132
24 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr., Sect. D: Biol. Crystallogr. **62**, 72–82
25 Argiriadi, M. A., Morisseau, C., Hammock, B. D. and Christianson, D. W. (1999) Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase. Proc. Natl. Acad. Sci. U.S.A. **96**, 10637–10642
26 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Cryst. **40**, 658–674
27 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr., Sect. D: Biol. Crystallogr. **60**, 2126–2132
28 Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., Sect. D: Biol. Crystallogr. **66**, 213–221
29 Painter, J. and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr., Sect. D: Biol. Crystallogr. **62**, 439–450
30 Heikinheimo, P., Goldman, A., Jeffries, C. and Ollis, D. L. (1999) Of barn owls and bankers: a lush variety of α/β hydrolases. Structure **7**, R141–R146
31 Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. **32**, W665–W667

32 Hall, T. A. (1999) Bioedit: a user-friendly biological sequence aligment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**, 95–98
33 Combet, C., Jambon, M., Deléage, G. and Geourjon, C. (2002) Geno3D: automatic comparative molecular modelling of protein. Bioinformatics **18**, 213–214
34 Beloqui, A., Polaina, J., Vieites, J. M., Reyes-Duarte, D., Torres, R., Golyshina, O. V., Chernikova, T. N., Waliczek, A., Aharoni, A., Yakimov, M. M. et al. (2010) Novel hybrid esterase-haloacid dehalogenase enzyme. ChemBioChem **11**, 1975–1978
35 Reyes-Duarte, D, Ferrer, M. and García-Arellano, H. (2012) Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. Methods Mol. Biol. **861**, 101–113
36 Lukashin, A. V. and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26**, 1107–1115
37 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402
38 Arpigny, J. L. and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. Biochem. J. **343**, 177–183
39 Bhowmik, S., Horsman, G. P., Bolin, J. T. and Eltis, L. D. (2007) The molecular basis for inhibition of BphD, a C-C bond hydrolase involved in polychlorinated biphenyls degradation: large 3-substituents prevent tautomerization. J. Biol. Chem. **282**, 36377–36385
40 Horsman, G. P., Ke, J., Dai, S., Seah, S. Y., Bolin, J. T. and Eltis, L. D. (2006) Kinetic and structural insight into the mechanism of BphD, a C-C bond hydrolase from the biphenyl degradation pathway. Biochemistry **45**, 11071–11086
41 Horsman, G. P., Bhowmik, S., Seah, S. Y., Kumar, P., Bolin, J. T. and Eltis, L. D. (2007) The tautomeric half-reaction of BphD, a C-C bond hydrolase. Kinetic and structural evidence supporting a key role for histidine 265 of the catalytic triad. J. Biol. Chem. **282**, 19894–19904
42 Geiselbrecht, A. D., Hedlund, B. P., Tichi, M. A. and Staley, J. T. (1998) Isolation of marine polycyclic aromatic hydrocarbon (PAH)-degrading *Cycloclasticus* strains from the Gulf of Mexico and comparison of their PAH degradation ability with that of puget sound *Cycloclasticus* strains. Appl. Environ. Microbiol. **64**, 4703–4710
43 Kasai, Y., Kishira, H. and Harayama, S. (2002) Bacteria belonging to the genus *Cycloclasticus* play a primary role in the degradation of aromatic hydrocarbons released in a marine environment. Appl. Environ. Microbiol. **68**, 5625–5633
44 Nandhagopal, N., Senda, T., Hatta, T., Yamada, A., Masai, E., Fukuda, M. and Mitsui, Y. (1997) Three-dimensional structure of microbial 2-hydroxyl-6-oxo-6-phenylhexa-2,4-dienoic acid (Hpda) hydrolase (Bphd enzyme) from *Rhodococcus* sp. strain Rha1, in the Pcb degradation pathway. Proc. Jpn. Acad., Ser. B **73**, 154–157
45 Pesaresi, A. and Lamba, D. (2005) Crystallization, X-ray diffraction analysis and phasing of carboxylesterase PA3859 from *Pseudomonas aeruginosa*. Biochim. Biophys. Acta **1752**, 197–201
46 Galperin, M. Y. and Koonin, E. V. (2012) Divergence and convergence in enzyme evolution. J. Biol. Chem. **287**, 21–28

Received 22 April 2013/10 June 2013; accepted 10 June 2013
Published as BJ Immediate Publication 10 June 2013, doi:10.1042/BJ20130552

CAPÍTULO 4
ESTERASAS DEL MAR MEDITERRÁNEO

RESUMEN DEL CONTENIDO

Este tercer capítulo proporciona información sobre las características bioquímicas y estructurales de enzimas aisladas de ambientes marinos profundos sometidos a condiciones poli-extremas de temperatura, presión y salinidad. Esto supone un salto cuantitativo frente a los estudios de los dos capítulos anteriores ya que las muestras analizadas procedían de ambientes moderadamente extremos. En particular, se han utilizado muestras procedentes de 4 fosas marinas hipersalinas profundas (3,040-3,342 m) y una fosa marina no hipersalina (4,908 m de profundidad) del Mar Mediterráneo, caracterizadas por una temperatura moderada constante de 13.0-16.5ºC, y un amplio rango de salinidad (39-348 unidades prácticas de salinidad (g/L)). Se ha empleado también a modo comparativo una muestra de una fumarola hidrotermal de la Isla de San Pablo (100 m de profundidad y 65ºC). Para ello se han utilizado técnicas metagenómicas para la identificación de clones con actividad esterasa y la posterior identificación de marcos de lectura abiertas que codifican las enzimas de interés. Posteriormente se han empleado técnicas de biología molecular para la expresión, producción a pequeña escala y purificación de las enzimas de interés. El empleo de técnicas bioquímicas, incluidos ensayos espectrofotométricos con un total de 210 sustratos diferentes, permitió posteriormente la caracterización de las mismas. Se han realizado estudios comparativos a nivel de secuencia de aminoácidos y tetra-nucleótidos. Así mismo se han realizado experimentos de cristalización, y finalmente, a diferencia de los capítulos, en este trabajo incorporamos estudios de dicroísmo circular para calcular la temperatura de desnaturalización de las proteínas y estudios de la actividad específica a altas presiones. A diferencia de los capítulos anteriores el estudio se ha realizado no solo con esterasas de la super-familia de las α/β-hidrolasas (5 en total), sino también con otros tres tipos de enzimas, en particular, glicosidasas (1), aldo-ceto reductasas (1) y lactato dehidrogenasas (1).

La masa molecular y el punto isoeléctrico de las enzimas oscila entre 26,802 y 84,279 Da, y 4.38 y 7.294, respectivamente. A nivel de secuencia las enzimas identificadas y analizadas presentan homologías entre el 41% y el 99%. El pH óptimo varía entre 6.0 y 8.5, la temperatura óptima de actividad oscila entre los 16 y los 70ºC y las concentraciones de sal (NaCl) para actividad óptima varían entre 0 y 4.0 M. La temperatura de desnaturalización oscilaba entre 40.3 y 71.4ºC. La mayoría de las enzimas procede de bacterias del filo Tenericutes (5), Proteobacteria (2) y Firmicutes (1). Entre los resultados más relevantes en este trabajo hemos demostrado que en ambientes marinos profundos hipersalinos, la adaptación de las proteínas a la presión está relacionada a la resistencia térmica (p valor = 0.0036). Por lo tanto, la salinidad puede incrementar la ventana de temperatura óptima de las enzimas en ambientes profundos. Como ejemplo, la fosa *Medee*, la mayor fosa marina profunda hiper-salina del Mar Mediterráneo, cuya temperatura no ha superado nunca los 16.5ºC, contiene una enzima halotermopiezofílica cuyas temperaturas óptima de actividad y de desnaturalización son 70ºC y 71.4ºC, respectivamente. Se trata por tanto de la primera enzima con estas características descrita en un ambiente no hiper-termófilo.

Por otro lado, se ha determinado la estructura cristalográfica de 5 de las enzimas, siendo el mayor conjunto de proteínas cristalizadas de ambientes marinos profundos publicados hasta la fecha. El análisis estructural revela que la adaptación de las proteínas a condiciones poli-extremas de presión, salinidad y temperatura se basa en mecanismos moleculares diferentes a los conocidos. Así mismo, se ha demostrado que las enzimas descritas en este trabajo tienen una disposición de los centros activos muy diferentes a las enzimas homólogas descritas en la bibliografía, lo que explica las diferencias a nivel de especificidad de sustrato.

El CD adjunto a esta Memoria contiene Material Suplementario relacionado con este Capítulo.

Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats

María Alcaide,[1‡] Peter J. Stogios,[2‡] Álvaro Lafraya,[1†] Anatoli Tchigvintsev,[2] Robert Flick,[2] Rafael Bargiela,[1] Tatyana N. Chernikova,[3] Oleg N. Reva,[4] Tran Hai,[3] Christian C. Leggewie,[5] Nadine Katzke,[6] Violetta La Cono,[7] Ruth Matesanz,[8] Mohamed Jebbar,[9] Karl-Erich Jaeger,[6] Michail M. Yakimov,[7] Alexander F. Yakunin,[2] Peter N. Golyshin,[3] Olga V. Golyshina,[3*] Alexei Savchenko,[2**] Manuel Ferrer[1*] and The MAMBA Consortium**

[1]*Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28049, Spain.*

[2]*Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada.*

[3]*School of Biological Sciences, University of Bangor, Gwynedd LL57 2UW, UK.*

[4]*Department of Biochemistry, University of Pretoria, Pretoria, South Africa.*

[5]*Evocatal GmbH, Monheim am Rhein 40789, Germany.*

[6]*Institute of Molecular Enzyme Technology Heinrich-Heine-University Düsseldorf and Institute of Bio- and Geosciences IBG-1: Biotechnology Forschungszentrum Jülich GmbH, Jülich D-52426, Germany.*

[7]*Institute for Coastal Marine Environment, CNR, Messina 98122, Italy.*

[8]*Centro Investigaciones Biológicas, CSIC, Madrid 28040, Spain.*

[9]*Université de Bretagne Occidentale, Laboratoire de Microbiologie des Environnements Extrêmes-UMR 6197 (CNRS-Ifremer-UBO), Institut Universitaire Européen de la Mer, Plouzané, France.*

Received 7 August, 2014; revised 1 October, 2014; accepted 2 October, 2014. For correspondence. *E-mail mferrer@icp.csic.es; Tel. +34915854872; Fax +34915854760. **E-mail alexei.savchenko@utoronto.ca; Tel. 416 978 3925; Fax 416 978 8605. ***E-mail: o.golyshina@bangor.ac.uk; Tel. 4401248383629; Fax 4401248382569. †Present address: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP), Institut National des Sciences Appliquées, 135 Avenue de Rangueil, F-31077 Toulouse Cedex 04, France. ‡The authors have contributed equally to this work.

Summary

The present study provides a deeper view of protein functionality as a function of temperature, salt and pressure in deep-sea habitats. A set of eight different enzymes from five distinct deep-sea (3040–4908 m depth), moderately warm (14.0–16.5°C) biotopes, characterized by a wide range of salinities (39–348 practical salinity units), were investigated for this purpose. An enzyme from a 'superficial' marine hydrothermal habitat (65°C) was isolated and characterized for comparative purposes. We report here the first experimental evidence suggesting that in salt-saturated deep-sea habitats, the adaptation to high pressure is linked to high thermal resistance (*P* value = 0.0036). Salinity might therefore increase the temperature window for enzyme activity, and possibly microbial growth, in deep-sea habitats. As an example, Lake *Medee*, the largest hypersaline deep-sea anoxic lake of the Eastern Mediterranean Sea, where the water temperature is never higher than 16°C, was shown to contain halopiezophilic-like enzymes that are most active at 70°C and with denaturing temperatures of 71.4°C. The determination of the crystal structures of five proteins revealed unknown molecular mechanisms involved in protein adaptation to poly-extremes as well as distinct active site architectures and substrate preferences relative to other structurally characterized enzymes.

Introduction

The deep oceanic/sea regions (below 200 m depth) form the largest marine subsystem by volume and comprise 1.3×10^{18} m^{-3} or approximately 80% of the oceanic/sea volume (De Corte *et al.*, 2012). However, the bathy (1000–4000 m depth), abyssopelagic (4000–6000 m depth) and hadopelagic (below 6000 m depth) regions are by far the least explored systems on Earth, although they are the largest reservoirs of organic carbon in the biosphere and also home to largely enigmatic food webs (Nagata *et al.*, 2010). The habitability window in deep realms is shaped mostly as a function of salinity, pressure and temperature, and the individual and collective effects of these characteristics on life have been extensively

investigated (Harrison *et al*., 2013). Microorganisms are able to grow in a wide range of salt concentrations ranging from sea water (De Corte *et al*., 2012) to salt-saturated lakes (Daffonchio *et al*., 2006; Smedile *et al*., 2013; Yakimov *et al*., 2013), pressures up to 120 MPa (e.g. Zeng *et al*., 2009) and temperatures from 4°C (De Corte *et al*., 2012) to ~ 60–108°C in active chimney walls (Zeng *et al*., 2009; Eloe *et al*., 2011; Wang *et al*., 2011). However, attempts to define the collective influence of these environmental conditions on protein function are scarce, and in particular, the extent of poly-extremes remains mostly undefined. Poly-extremes are of particular importance as life adaptation is a focus of intense research interest, much of which is centred on dissecting the changes in the composition and genomic content of the communities under different environmental constraints (Daffonchio *et al*., 2006; Schlitzer, 2010; Eloe *et al*., 2011; La Cono *et al*., 2011; Smedile *et al*., 2013; Yakimov *et al*., 2013). One limitation of taxonomic and genomic data based on the analysis of total extracted DNA is that these studies cannot address whether an organism is alive or has succumbed to such multiple extreme conditions. Another major limitation of using small-subunit ribosomal gene surveys and shotgun data is the large number of organisms and genes that are anonymous; the majority have not yet been cultured or sequenced (Puspita *et al*., 2012; Akondi and Lakshmi, 2013). However, these data can be complemented with naïve screens, which directly analyse the enzymes of metagenomes (Martínez-Martínez *et al*., 2013 and references therein).

The importance of defining the combined effects of salt, pressure and temperature extremes on protein function contrasts with the limited information available in the literature. Therefore, as shown in Table S1, the majority of enzymes characterized thus far from deep-sea realms were mostly isolated either from slightly saline, cold (4°C) habitats or from slightly saline, high-temperature hydrothermal vents, and only one study described enzymes from salt-saturated basins (Ferrer *et al*., 2005). In addition, only six deep-sea proteins, all from single cultivated organisms (De Vos *et al*., 2007; Shirai *et al*., 2008; Xu *et al*., 2008; Shin *et al*., 2009; Sineva and Davydov, 2010; Pietra, 2012), have been structurally characterized; none of which originated from sites experiencing the three stressors (Table S1). This dearth of three-dimensional structures of proteins from deep-sea inhabitants, particularly from salt-saturated biotopes, precludes a thorough understanding of the structural adaptations necessary for life in poly-extremophilic environments and stifles the discovery and optimization of useful enzymes for structural-functional, engineering and industrial purposes. Nevertheless, it has been recently shown that DNA-based methodologies appear to be inaccurate approaches by which to study the 'adaptation signatures' in the brines of deep-sea salt-saturated lakes (Hallsworth *et al*., 2007; Yakimov *et al*., 2013). Rather, biochemical-based methodologies might be more accurate approaches by which to study such 'signatures', as enzyme activities represent the highest level of the functional hierarchy regardless of the heterogeneities that commonly appear at the DNA and amino acid levels.

This investigation takes a step beyond descriptive studies of microbial cultivation and gene repositories and the utilization of naïve screens, biochemical tests and structure determinations provided deeper insights into the combined effects of salinity, high pressure and temperature on marine enzymes. We used the bio-resources from two collaborative projects, BEEM (http://www.beem.utoronto.ca; funded by Genome Canada) and MAMBA (http://mamba.bangor.ac.uk; funded by the EU FP7 program), which are focused on investigating moderate-temperature deep-sea sites in the Mediterranean Sea. We particularly focused here on the salt-saturated deep-sea lakes *Medee*, *Bannock* and *Kryos* (Daffonchio *et al*., 2006; Yakimov *et al*., 2013) and the hadopelagic seawater column at Station Matapan-Vavilov Deep (Smedile *et al*., 2013) in the Eastern Mediterranean Sea (Fig. 1). A 'superficial' marine hydrothermal habitat was used for comparison. The extensive characterization of a set of nine different enzymes isolated from those sites by naïve screens and the determination of five crystal structures provided, to the best of our knowledge, the first experimental evidence linking pressure adaptation to thermal adaptation in salt-saturated habitats, by as yet unknown molecular mechanisms. The results are discussed in the context of expanding the thermal window for growth in deep-sea realms, and novel electrostatic charges and active site architectures are also examined.

Results and discussion

Deep-sea metagenome libraries, screening and general features of selected sequence-encoded esterases (ESTs)

Samples were collected from two distinct deep-sea marine environments. First, brine/interface fluid was collected from deep hypersaline anoxic basins (DHAB): Lake *Medee* (3040 m depth; 15.5°C), which is the largest known DHAB, Lake *Bannock* (3342 m depth; 14.5°C) and Lake *Kryos* (3340 m depth; 16.5°C) (Daffonchio *et al*., 2006; Yakimov *et al*., 2013). Note that brine fluids were collected from *Medee* and *Bannock* basins whereas alive interface was collected from Lake *Kryos*. Second, seawater was collected from the hadopelagic Station Matapan-Vavilov Deep (4908 m depth; 14.5°C), which is the deepest site of the Mediterranean Sea (Smedile *et al*.,

A

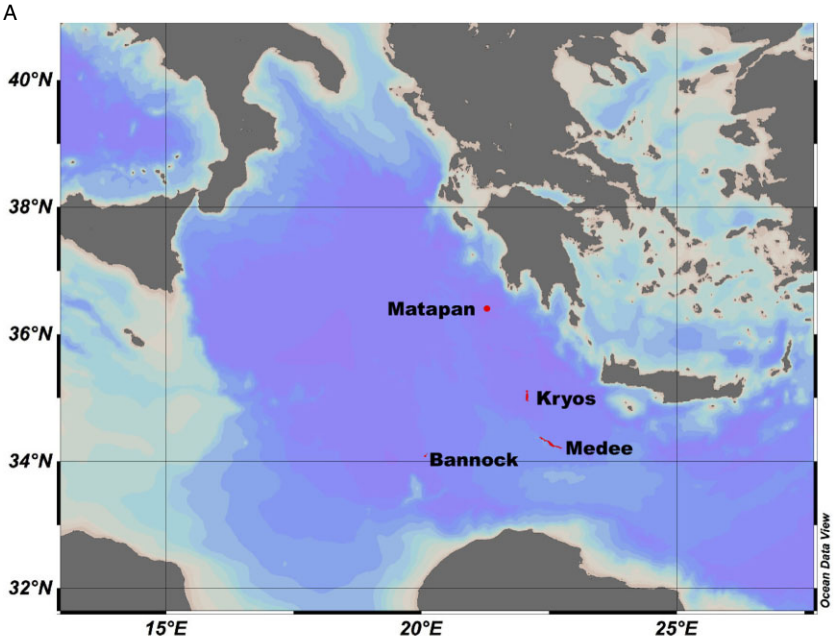


Fig. 1. Global map displaying the locations of the main deep-sea sites in the Mediterranean Sea investigated in this study. Hydrochemistry of selected deep-sea regions (A) is shown in (B) (for extensive details, see Table S2).

B

	Medee	Kryos	Bannock	Matapan
Depth (m):	3010	3340	3342	4908
Salinity (g kg^{-1}) :	347.80	313.92	263.20	38.6
Temperature (ºC) :	14.75–15.46	16.5	14.5	14.29

2013). At these sites, the water temperature is never below 13.0°C or higher than 16.5°C. Total DNA was extracted, and subsets of 40 024 clones from the four libraries generated in this study harbouring nearly 1.3 Gbp of community genomes were scored for the ability to hydrolyse α-naphthyl acetate (αNA) and tributyrin, as previously reported (Reyes-Duarte *et al.*, 2012), which is indicative of EST/lipase activity. A total of five unique clones were selected as active. Hydrochemistry of selected deep-sea regions and EST screening statistics are shown in Fig. 1 and Table S2. The inserts were sequenced, analysed and compared with the sequences available in the National Center for Biotechnology Information (NCBI) non-redundant public database (Hall, 1999). Five (one per active clone) predicted metagenome sequence (MGS)-encoding ESTs with the α/β hydrolase fold were identified and successfully produced as soluble proteins when expressed in *Escherichia coli*, and their properties were investigated. They were named based on the source ID followed by a serial number: MGS-M1 and MGS-M2 (from the *Medee* basin), MGS-B1 (from the *Bannock* basin), MGS-K1 (from the *Kryos* basin), and MGS-MT1 (from the Matapan basin).

According to BLAST searches of the NCBI non-redundant database, the five studied protein sequences were 44–62% similar to homologous proteins in the database (Table S3A–B). The deduced molecular masses and estimated *p*I values of these proteins ranged from 31.6 to 56.2 kDa and from 4.4 to 7.3 respectively. The pairwise amino acid sequence identity ranged from 8.3% to 21.8%; MGS-B1 and MGS-MT1 were the most similar enzymes (21.8% sequence identity), whereas MGS-K1 and MGS-M1 were the most divergent at the sequence level (8.3% sequence identity). The selected α/β hydrolases contain a classical Ser-Asp-His catalytic triad, but the catalytic elbow and oxyanion hole (i.e. the GXSXGG and H/N-GGG(A)/P-X motifs) often diverged from the consensus regions as identified by an extensive sequence analysis. Nonetheless, there was adequate sequence conservation among these catalytic motifs and the overall enzyme sequence to categorize the enzymes into the following accepted lipase/EST subfamilies (Kourist *et al.*, 2010): family IV (MGS-MT1, MGS-B1 and MGS-M1), family V (MGS-M2) and family VII (MGS-K1). The sequence-based features and amino acids participating in the predicted catalytic sites are listed in Table S3A–B.

Salt-saturated deep-sea brines contain biochemical signatures indicating adaptation to salinity and thermal extremes

According to the standard assay conditions described in Experimental procedures, and summarized in Table 1, the five proteins were fully characterized. We first confirmed that the purified proteins, which were most active at pH values ranging from 7.0 to 8.5 (Fig. S1), exhibited the expected EST activity, tested over a set of 101 structurally different esters. Extensive differences in activity level as well as substrate profiles and preferences were noticeable according to specific activity (units mg^{-1}) determinations (Appendix S1; Fig. S2). Using *p*-nitrophenyl propionate (*p*NP-propionate) as a model substrate, Lake *Medee* enzymes were found to be the least active enzymes, i.e. the enzyme from Matapan-Vavilov Deep, which was the most active enzyme, exhibited specific activities 900-fold greater than those of Lake *Medee* enzymes (Table 1).

The enzymatic activities of purified proteins were stimulated by the addition of NaCl and KCl to the reaction mixture (Fig. 2). The optimal concentration of Na^+/K^+ for activity was the lowest for MGS-B1 and MGS-K1 (optimal at 0.8–1.2 M), whereas the other enzymes were most active at concentrations greater than 3.0 M. At the optimal concentration, MGS-B1 from Lake *Bannock* exhibited the greatest increase in activity (14-fold) compared with reactions not containing salts. $MgCl_2$ triggered stronger effects on enzyme activities. However, while MGS-K1, MGS-B1 and MGS-M1 (in order of greatest inhibition to least inhibition by $MgCl_2$) were strongly inhibited by this salt at concentrations greater than 0–0.8 M, MGS-M2 and MGS-MT1 were stimulated with maximal activity at 2.4 and 1.6 M salt respectively. The activation of all enzymes by $Na^+/K^+/Mg^{2+}$ indicates that the properties of the ESTs herein reported reflect specific habitat characteristics and that activation by sodium, potassium, and, to a lesser extent, magnesium, may be common in enzymes from the deep-sea sites examined, independently of the habitat environmental constraints.

Three distinct profiles of temperature optima were further observed (Fig. 3). The first profile was a thermophilic-like profile, as exemplified by MGS-M2, in which the enzyme was most active at 70°C and retained ≤ 5% activity at 4–16°C (in the presence or absence of NaCl). Calculation of denaturing temperature (71.4°C) by circular dichroism (CD; Table 1) confirmed the high protein stability. The high optimal temperature of MGS-M2 was unexpected because this enzyme was isolated from Lake *Medee*, a permanently moderate-temperature site (15.5°C) (Yakimov *et al*., 2013). To prove that the MGS-M2 enzyme showed a temperature profile typical for thermophilic proteins, an EST with the α/β hydrolase fold (referred to as MGS-HA1), which originated from a clone library created from a superficial seawater sample at a hydrothermal vent at Saint Paul Island (100 m depth) and maintained at 60–65°C, was purified and characterized for comparative purposes. The enzyme was derived from a clone (out of 20 000 total clones) active towards αNA. The MGS-HA1 enzyme, which can be categorized into the lipase/EST family VI (Kourist *et al*., 2010) and most likely originated from *Geobacillus* as determined by BLAST homology search and GOHTAM (Tables S3A and S4), showed maximal activity at pH 8.0 (Fig. S1) and concentrations of $Na^+/K^+/Mg^{2+}$ of up to 1.6–3.2 M (Fig. 1). Its optimal temperature for activity (70–75°C) and residual activity at low temperatures, e.g. < 0.4% at 4°C (Fig. 3),

Table 1. Specific activity and protein denaturation temperature (T_d) of proteins as determined by circular dichroism.

Enzyme	Activity	Protein fold	Specific activity (units g^{-1})[a]	Standard assay conditions for activity determination [pH/T(°C)/NaCl (M)][a]	T_d (°C)[b]
MGS-M1	EST	α/β Hydrolase	105.73 ± 2.34	8.0/25/3.6	65.2
MGS-M2	EST	α/β Hydrolase	188.32 ± 8.76	8.0/70/4	71.4
MGS-B1	EST	α/β Hydrolase	24 077 ± 85	8.0/25/0.8	52.4
MGS-K1	EST	α/β Hydrolase	18 094 ± 270	7.0/30/0.8	40.3
MGS-MT1	EST	α/β Hydrolase	94 994 ± 460	8.0/40/2.8	55.7
MGS-M3	GLY	α/β Hydrolase	12 471 ± 895	8.0/45/ 0.4	59.3
MGS-M4	AKR	TIM-barrel	5371 ± 28	8.0/16/0.8	63.3
MGS-M5	LDH	Rossman	1072 ± 12	8.0/30/0	n.d
MGS-HA1	EST	α/β Hydrolase	408.2 ± 18.0	8.0/70/3.2	79.5

a. *p*NP-propionate (1 mM; for EST activity), *p*NP-β-D-glucose (30 mM; for GLY activity), sodium pyruvate (5 mM; for LDH activity) and methyl glyoxal (1 mM; for AKR activity) were used as standard assay substrates. The following buffers were used for activity determinations: 50 mM Tris-HCl for pH 8.0 and 50 mM 4-(2-hydroxyethylpiperazine-1-ethanesulphonic acid (HEPES) for pH 7.0, which were used as standard assay buffers. Note that activity determinations were performed at the optimal parameters and conditions specifically cited in the Experimental Procedures, and summarized in this Table.
b. Standard deviation < 0.1.
Activity and protein fold associated to each of the enzymes are also summarized. AKR, aldo-keto reductase; EST, esterase; GLY, glycosidase; LDH, lactate dehydrogenase; n.d., not determined.

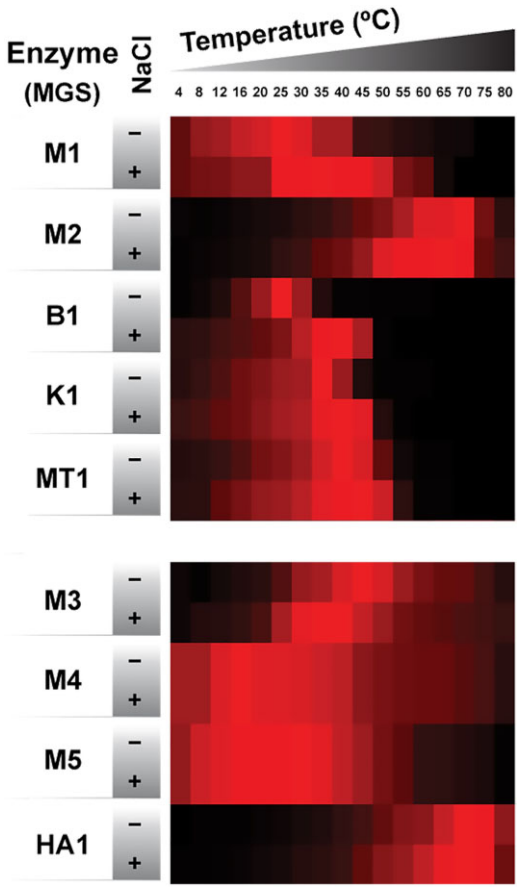


Fig. 2. Temperature profiles of the enzymes in the absence (–) or presence (+) of salt (NaCl). Profiles for deep-sea esterases are shown in the top panel, whereas those for other enzymes are shown in the bottom panel. The heat map colours represent the relative percentages of specific activity (units g^{-1}) compared with the maximum (100%) given in Table 1. The specific activities were calculated in triplicate (SD < 0.5%) using the standard assay substrates (see Table 1) and conditions described in the Experimental procedures. Note that due to the significant differences in specific activities in the presence and absence of salt, the assays were performed using the same units of enzymes on the basis of *p*NP-propionate transformation. This effect is particularly noticeable for the MGS-B1 esterase, which retains only 3.4% of its activity in the absence of salt compared with the optimum level (0.8 M NaCl). The colour code ranges from black (no activity) to intense red (100% activity). Heat maps were constructed in R (http://www.r-project.org) using the 'heatmap.2' function within the 'gplots' package.

allowed it to be categorized as a typical thermophilic enzyme, which was further evidenced by the determination of its denaturing temperature (79.5°C; Table 1). The comparative analysis of MGS-M2 and MGS-HA1 showed that MGS-M2 reassembled a thermophilic protein. The second profile was a mesophilic-like profile, as exemplified by MGS-MT1, in which the enzyme was most active at 40–45°C and retained less than 40% of its activity at ≥ 50°C (Fig. 3) in the presence or absence of NaCl; this result is in agreement with its denaturing temperature (55.7°C; Table 1). The third profile consisted of a psychrophilic-like profile, as exemplified by MGS-M1, MGS-B1 and MGS-K1, with enzymes that were most active at 25–40°C (Fig. 3). Notably, compared with mesophilic (MGS-MT1) and thermophilic (MGS-M2) ESTs that showed a salt-independent thermal profile (Fig. 3), a positive impact of salinity on thermal activation/stabilization in these three enzymes was observed in the presence of optimal concentrations of NaCl. This result was particularly noticeable for MGS-M1 and MGS-B1, as they displayed a shift in the optimal temperature from 25 to 35°C and from 25 to 40°C respectively. It was also true to a lesser extent for MGS-K1 (from 35 to 40°C). The higher stabilization levels for MGS-M1 and MGS-B1 agreed with their higher denaturing temperatures (65.2°C and 52.4°C respectively) compared with MGS-K1 (40.3°C). Together, the data demonstrated that deep-sea salt-saturated biotopes might contain enzymes adapted to work under multiple temperature extremes (80% activity retained in the range from 16 to 70°C), despite these sites being moderately warm (14.0–16.5°C).

Adaptation to high pressure is linked to high thermal resistance in deep-sea brines

As pressure is one of the most representative characteristics of deep-sea habitats, the influence of hydrostatic pressure (350 bar) on enzyme performance was further evaluated in high-pressure 2 ml reactors as described in the Experimental procedures, and presumptive links between the site and EST characteristics were evaluated. The data presented in Fig. 4 revealed that the relative percentage of activity at 350 bar compared with that at atmospheric pressure positively correlated with the optimal temperature for activity in the two ESTs from the salt-saturated Lake *Medee*; thus, the highest adaptation to pressure was obtained for the MGS-M2 EST, which had the highest temperature optima (70°C) compared with MGS-M1 (25°C). No such correlation could be evaluated for the other deep-sea sites, which showed different activation levels, as only one enzyme candidate was characterized per site; no additional enzymes could be tested because no additional active clones could be obtained in our library screen tests, or we were unable to produce additional soluble proteins from sequenced positive fosmids.

To prove that such a thermal-pressure correlation exists in Lake *Medee*, we polymerase chain reaction-amplified candidate genes from the two EST-positive clones (the ones containing MGS-M1 and MGS-M2 ESTs), and we were able to successfully express and produce in soluble form three proteins: a glycosidase (GLY; herein named MGS-M3), an aldo-keto reductase (AKR; MGS-M4) and an (*L*)-lactate dehydrogenase

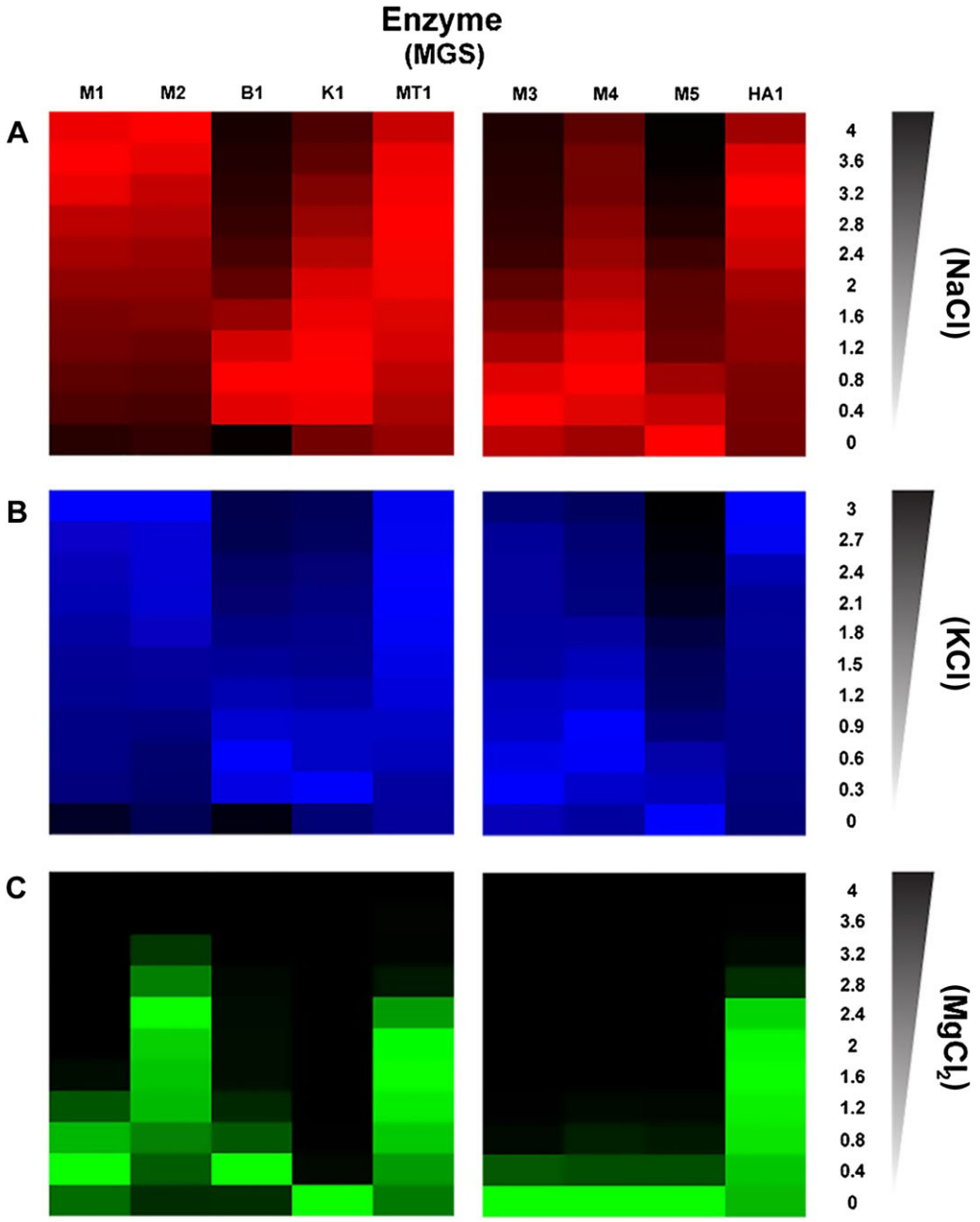


Fig. 3. Heat maps displaying the activities of enzymes at different concentrations of NaCl (A, red), KCl (B, blue) and $MgCl_2$ (C, green). Profiles for deep-sea esterases are shown in the left panel, whereas those for other enzymes are shown in the right panel. The heat map colours represent the relative percentages of specific activity (units g^{-1}) compared with the maximum activity (100%). The specific activities were calculated in triplicate [standard deviation (SD) < 0.5%] using the standard assay substrates (see Table 1) and conditions described in the Experimental procedures. The colour code ranges from black (no activity) to intense red, blue and green (100% activity). The 100% levels for NaCl/KCl/$MgCl_2$ are as follows: MGS-M1 (105.7/105.6/38.1 units/g), MGS-M2 (188.3/111.6/212.5 units g^{-1}), MGS-M3 (12 471/ 8142/6481 units g^{-1}), MGS-M4 (5371/4715/362.0 units g^{-1}), MGS-M5 (1072/1000/153.7 units g^{-1}), MGS-B1 (24 077/11 685/4496 units g^{-1}), MGS-K1 (18 094/17 829/8239 units g^{-1}), MGS-MT1 (94 994/88 259/117 186 units g^{-1}) and MGS-HA1 (408.2/388.4/246.1 units g^{-1}). Heat maps were constructed in R (http://www.r-project.org) using the 'heatmap.2' function within the 'GPLOTS' package.

(LDH; MGS-M5). According to the CAZY database (http://www.cazy.org; Cantarel *et al.*, 2009) and considering structural similarities, MGS-M3 is related to family 3 of GLYs, whereas MGS-M4 belongs to the gluconic reductase subfamily 5 of AKRs, and MGS-M5 matches lactate/malate dehydrogenases rather than other Rossman fold-containing enzymes. The sequence-based features and amino acids participating in the predicted

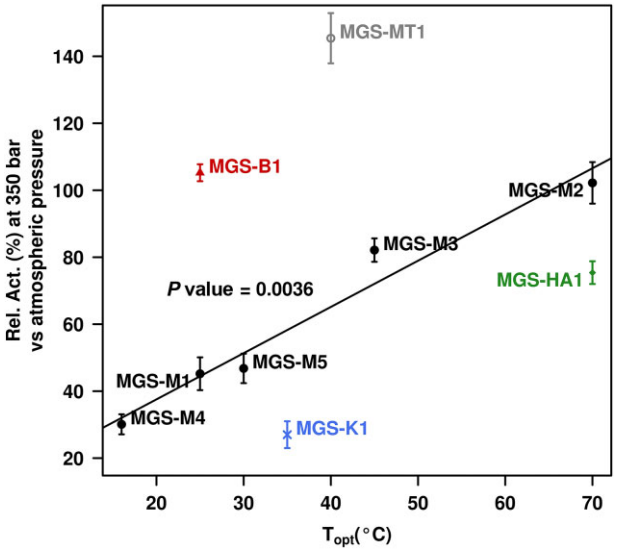


Fig. 4. Activity levels of enzymes as a function of pressure and optimal temperature for activity. The specific activities using the standard assay substrates and pH (see Table 1) and 25°C were calculated in triplicate [standard deviation (SD) is shown] at atmospheric pressure or 350 bar as described in the Experimental procedures. One hundred percent activity refers to the activity value at atmospheric pressure for each of the enzymes. The optimal temperature for activity of each of the enzymes studied is reported in Fig. 2. Only enzymes from *Medee* Lake were considered for correlation analysis (see Results and Discussion section for details), and thus the enzymes from other site are shown in different colours.

catalytic sites are listed in Table S3C. According to the standard assay conditions described in Experimental procedures (see also Table 1), the corresponding activities, found to be maximal at pH 8.0 (Fig. S1), were first confirmed against a total of 47 model substrates (Table 1; Appendix S1; Fig. S1). Activity measurement further showed that while MGS-M5 was inhibited by Na^+ and K^+ (Fig. 1), MGS-M3 and MGS-M4 were activated by low salt concentrations (0.3–0.4 M and 0.8–1.2 M respectively). $MgCl_2$ strongly inhibited all enzymes (Fig. 1). The influence of temperature and hydrostatic pressure (350 bar) on enzyme performance was further evaluated for these three enzymes, which confirmed that the highest resistance to pressure was obtained for enzymes with the highest temperature optima (Fig. 4).

Overall, we found that pressure adaptation, at least in Lake *Medee*, is linked to thermal adaptation and that the correlation (*P* value = 0.0036; *t*-test; Fig. 4) was accurate for five proteins with different types of fold, such as MGS-M1, MGS-M2 and MGS-M3 with an α/β-hydrolase fold, MGS-M4 with a triosephosphate isomerase TIM-barrel fold, and MGS-M4 with a Rossman fold (MGS-M4).

Analysis of temperature, salt and pressure adaptations in enzyme primary and tertiary structures

Given the diverse thermo- and halo-tolerance of the MGS proteins (for summary see Figs 2 and 3), we were interested in identifying primary sequence and tertiary structure elements that might reflect adaptations conferring these properties. It has been hypothesized that intracellular proteins of thermophilic, psychrophilic and halophilic organisms undergo multiple, distinct adaptations to retain activity in such environments. For example, enzymes of thermophilic origin may increase their stability to resist thermal-induced unfolding by increasing the size of their hydrophobic cores, increasing the number of disulphide bonds, forming additional salt-bridge interactions or increasing the number of charged residues on their exposed surfaces (Reed *et al.*, 2013). Enzymes of halophilic origin must increase their hydration to compensate for increased extracellular salt concentrations (Madern *et al.*, 2000; Delgado-García *et al.*, 2012; Reed *et al.*, 2013). Observed adaptations included an increased negative surface charge and/or lower isoelectric point (*p*I) caused by an increase in acidic residues and a decrease in lysine residues (Ferrer *et al.*, 2012), a decrease in aliphatic amino acids (Leu/Ile) and an increase in small hydrophobic amino acids (Gly/Ala/Val), and a decrease in the extent of buried non-polar amino acids. We searched for these adaptations by obtaining the crystal structures and/or primary sequences of the thermophilic enzyme MGS-M2, the psychrophilic/mesophilic enzymes MGS-M1, MGS-M4, MGS-M5 and MGS-MT1, the halophilic enzymes MGS-M1 to MGS-M4 and MGS-MT1 and comparing these enzymes with their closest homologues lacking these adaptations (Fig. 5; Appendix S1). We also analysed MGS-M5, which does not exhibit salt tolerance, as a baseline comparison. To remove any signal from substrate-binding amino acids correlated with functional diversification, we excluded residues close to the catalytic sites.

An analysis of 16 known temperature-dependent adaptations showed that the MGS-M2 enzyme contains six adaptive features typical of thermophilic proteins (Table S3D). However, the analysis also showed that the psychrophilic enzymes MGS-M1 and MGS-M4 contain 10 and 8 adaptations, respectively, that are typically observed in thermophiles. In addition, the mesophilic enzyme MGS-M4 contains nine of these so-called thermophilic adaptations. The presence of these thermophilic adaptations may explain the preservation of the activity of these enzymes at temperatures as high as 60–75°C (Fig. 2). These results suggest that other characteristics beyond those studied here are involved in conferring the temperature dependence of these MGS enzymes.

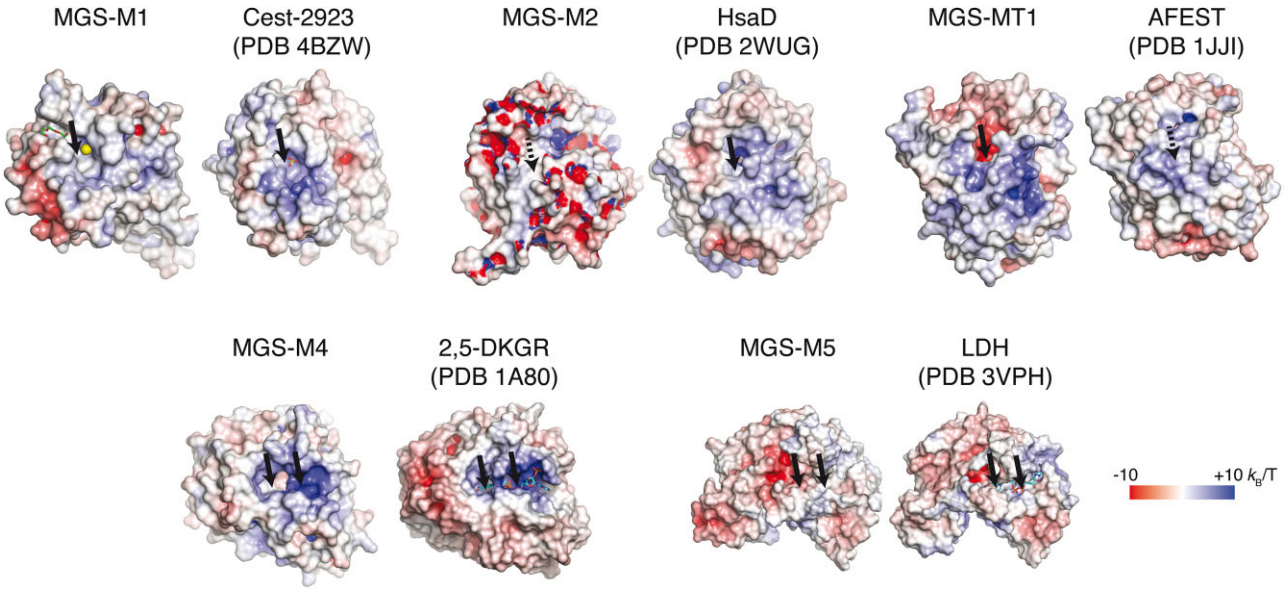


Fig. 5. Comparison of surface features of the MGS enzymes crystallized in this study and their structural homologues revealed by structure similarity searches. The solvent-exposed surfaces of the enzymes are shown, coloured by electrostatic potential as indicated by the scale at the bottom right (in units k_B/T = Boltzmann constant over temperature). The arrows refer to the locations of the catalytic serines (for esterases in the top row), with dashes indicating that this residue is hidden within the protein, or the NADPH/NADH and substrate binding sites (for enzymes in the bottom row).

As the activity of MGS-M4 was preserved even at 1 M KCl, we repeated the analysis of structural adaptations on this enzyme by comparison with a non-halophilic homologue, excluding a region surrounding the substrate-binding site. For this enzyme, the analysis identified only two of the seven adaptations (Table S3E, a decrease in overall *p*I and a decrease in lysine residues). This result suggests that this enzyme contains other unknown characteristics that preserve its activity in high-salt environments. Of the seven halophilic-dependent adaptations examined in the MGS enzymes, MGS-M1, MGS-MT1 and MGS-M2 contained five, three and two of these adaptations respectively (Table S3E). Thus, although the salt adaptation of MGS-M1 is well explained by previously identified adaptations, the salt adaptation of the MGS-M2 and MGS-MT1 enzymes is not. Notably, an increase in surface acidic residues is the most commonly observed adaptation in halophilic enzymes, but MGS-MT1 exhibited the opposite feature: it has a high *p*I value (7.29), reflecting a decrease in surface acidic residues. This evidence suggests that other undiscovered adaptations are present in this enzyme.

Piezophilic proteins appear to exhibit a smaller hydrophobic core, fewer large and more small amino acids in the hydrophobic core, increased multimerization and fewer intra-molecular ionic interactions (Michels and Clark, 1997). An analysis of these four known pressure-dependent adaptations (Table S3E) did not reveal that MGS-MT1 and MGS-M2 have more of these adaptations than the proteins that are not piezophilic (Fig. 4).

An analysis of the crystal structures further revealed that the conformation and/or overall electrostatic charge (Fig. 5; Fig. S3 and S4) were distinct relative to other structurally characterized homologues in four of the five proteins. Indeed, the overall electrostatic charge on the surface of MGS-M1 was similar to the charges of Cest-2923 and bacillibactin esterase (BES), with no obvious patches of concentrated charge; however, MGS-M1 contained a deletion of 14 amino acids relative to Cest-2923 (residues 141–150 in MGS-M1 and 158–181 in Cest-2923), which resulted in alterations in the conformation of the active site. MGS-M2 has different electrostatic characteristics and residue composition: the active sites of the closest structural homologue (HsaD) are noticeably positively charged, whereas those of MGS-M2 contain a mix of positive and negative features. The MGS-MT1 active site is long, deep, negatively charged, and open to solvent; these electrostatic characteristics are in sharp contrast to the properties of Archaeoglobus fulgidus esterase (AFEST). Compared with its closest structural homologues (2,5-DKGR), MGS-M4 contains a small (153 Å^3), slightly positively charged nicotinamide adenine dinucleotide phosphate (NADPH)-binding site and a smaller active site, which is missing some structural loops involved in substrate contacts. Finally, a comparison of the electrostatic surface and active site composition of the

MGS-M5 structure revealed that it closely resembled the structures of its closest homologues (LDH), including the four residues involved in interacting with the pyruvate/L-lactate ligands. Such differences are consistent with variations at the level of substrate profiles and preferences (Fig. S2), but they may also be implicated in yet unknown differential adaptations to environmental constraints. Full details about substrate fingerprint and crystal structures and active site architecture are given in Appendix S1.

Source organisms of investigated sequence-encoded ESTs

We finally investigated the presumptive microbial origin of the genes encoding enzymes under investigation (Table S4). A search against the GOHTAM database (Menigaud *et al*., 2012) revealed compositional similarities between the DNA fragments containing the genes for MGS-M1 to MGS-M5 and several bacterial strains that are most likely from uncharacterized species belonging to *Firmicutes*. Genes in the DNA fragment encoding MGS-M2 to MGS-M5 share similarity with the corresponding genes of the *Tenericutes*-like bacterium *Haloplasma contractile* (Antunes *et al*., 2008). This bacterium was recently isolated from deep-sea sediment samples and is a unique 'transiting form' of bacteria that has been placed on the phylogenetic tree between *Firmicutes* (*Bacilli* and *Clostridia*) and *Mollicutes* (*Mycoplasmas* and other strictly symbiotic bacteria). Interestingly, the closest neighbour to *H. contractile* is '*Candidatus lumbricincola*', an uncultured bacterium associated with earthworms (Nechitaylo *et al*., 2009). However, unlike *Mollicutes*, *H. contractile* grows axenically on 'simple' media, suggesting a free-living lifestyle (Antunes *et al*., 2011). MGS-B1 was not identified by GOHTAM, but the DNA fragment encoding this enzyme shares genes with *Betaproteobacteria* of the genus *Variovorax*. MGS-K1, which is part of an insertion sequence element or prophage, shares DNA similarity with *Aspergillus*, although TBLASTX analysis indicated some weak similarity with rare *Betaproteobacteria* species. Hence, similarly to other fragments, MGS-K1 could be derived from an unknown species. The DNA fragment encoding the MGS-MT1 enzyme exhibits compositional similarity to the *Alteromonas* genomes, most likely *Alteromonas macleodii* 'deep-ecotype', which dominates the Matapan-Vavilov Deep (Smedile *et al*., 2013).

Experimental procedures

Source of enzymes

Total DNA was extracted from selected sites using the G'NOME DNA extraction kit (BIO 101/Qbiogene, Morgan Irvine, CA, USA) according to the manufacturer's instructions for each of the microbial communities, and large-insert pCCFOS1 fosmid (for all but *Bannock* Lake) or bacteriophage lambda-based ZAP phagemid (*Bannock* Lake) libraries were generated (Ferrer *et al*., 2005; Alcaide *et al*., 2013) and scored for the ability to hydrolyse αNA and tributyrin (Reyes-Duarte *et al*., 2012). Positive clones were selected, and their DNA inserts were sequenced using a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Life Sequencing SL (Valencia, Spain) or completely sequenced using universal primers and subsequent primer walking. In all cases, upon completion of sequencing, the reads were assembled to generate non-redundant metasequences using Newbler GS De Novo Assembler v.2.3 (Roche). The GENEMARK software (Lukashin and Borodovsky, 1998) was employed to predict potential protein-coding regions (open reading frames with ≥ 20 amino acids) from the sequences of each assembled contig, and deduced proteins were screened via BLASTP and PSI-BLAST (Altschul *et al*., 1997). MGS-encoding enzymes were deposited in GenBank under the accession numbers KF831414 – KF831421.

General methods

The cloning, expression and purification of selected proteins using the p15TV-Lic vector and *E. coli* BL21(DE3) Codon Plus-RIL (for MGS-M4 and MGS-M5) and Ek/LIC 46 and *E. coli* BL22 (for MGS-M1, MGS-M2, MGS-B1, MGS-K1 and MGS-HA1) or *Rosseta* (for MGS-MT1) were performed as described elsewhere (Alcaide *et al*., 2013) using the primer pairs described in Table S5. Purity was assessed as > 98% by SDS-PAGE. Note that after purification, MGS-MT1, MGS-K1 and MGS-B1 were required to be maintained in 50 mM HEPES [4-(2-hydroxyethylpiperazine-1-ethanesulphonic acid] pH 7.0 containing 0.8 M NaCl to ensure protein stability. All chemicals used for enzymatic tests were of the purest grade available and were purchased from Fluka-Aldrich-Sigma Chemical Co. (St Louis, MO, USA) or Apin Chemicals (Oxon, UK) (Alcaide *et al*., 2013). EST activity was assayed using 1 mM *p*-nitrophenyl (*p*NP) esters (at 410 nm) and structurally diverse esters other than *p*NP esters (at 540 nm) using substrates and conditions as previously described (Alcaide *et al*., 2013). GLY activity was determined using 1 mM *p*NP sugars (at 410 nm) as described elsewhere (Del Pozo *et al*., 2012). LDH activity was assayed by a colorimetric assay in which the conversion of pyruvate to lactate was determined using a LDH kit (Sigma Chemical Co., St. Louis, MO, USA) to follow the release of NAD^+ at 340 nm in a reaction mixture containing 0.4 mM nicotinamide adenine dinucleotide (NADH) and 5 mM sodium pyruvate. AKR activity was determined routinely by monitoring the decrease in absorbance at 340 nm in an assay mixture containing the corresponding substrate (1 mM) and 0.4 mM NADPH. Unless otherwise stated, *p*NP-propionate (1 mM; for EST activity), *p*NP-β-D-glucose (30 mM; for GLY activity), sodium pyruvate (5 mM; for LDH activity) and methyl glyoxal (1 mM; for AKR activity) were used as standard assay substrates. The optimal pH, temperature and salt (NaCl) concentrations (see Table 1) were used for standard assay reactions and specific activity determinations (see Fig. S2). pH values between 4.0 and 11.0 (at the optimal temperature),

temperatures between 4 and 80°C (at the optimal pH), and NaCl, KCl and $MgCl_2$ concentrations of up to 4 M (at the optimal pH and temperature) were tested for optimal parameter determinations using the standard assay substrates (for details Table 1). For pH and temperature optima determinations, the assay buffers were supplemented with NaCl, which was used as model salt, at the optimum concentration or not supplemented (Table 1).

Unless stated otherwise, reactions were conducted using 0.005–2 μg of pure proteins; the absorbance was determined every 1 min for a total time of 15 min. All reactions were performed in triplicate, and one unit (U) of enzyme activity was defined as the amount of enzyme transforming 1 μmol of substrate in 1 min under the assay conditions. All values were corrected for non-enzymatic hydrolysis (background rate).

Pressure perturbation studies

The effect of hydrostatic pressure was further analysed by placing the reactions into a high-pressure incubating system that consists of 2 ml of high-pressure cell connected to a pressure generator (High Pressure Equipment, Erie, PA, USA) that is capable of generating a pressure of up to 10 000 psi. Briefly, 20 ml of standard assay buffer (see Table 1) containing 30 mM *p*NP-β-D-xylose (for MGS-M3), 1.0 mM *p*NP-propionate (for ESTs), 5 mM sodium pyruvate and 0.4 mM NADH (for MGS-M4) or 1.0 mM methyl glyoxal and 0.4 M NADH (for MGS-M5) was freshly prepared. Next, 0.32–540 μg of pure protein (depending on the enzyme) was added. Two millilitres of the reaction mixture was immediately transferred to high-pressure-maintaining reactors, and another 2 ml was held at atmospheric pressure. The experiments were performed at 25°C and at a salt (NaCl/KCl) concentration similar to that found at the sampling site [Fig. 1; 39–348 practical salinity units (PSUs)]. In all cases, reactions without protein were used as negative controls. After incubation for a total of 5–30 min (depending on the enzyme), the samples were depressurized, and the extent of the reaction was monitored by spectrophotometer measurements at 410 nm. All experiments were performed in triplicate.

Protein purification, crystallization and structure determination

The ESTs MGS-M1, MGS-M2 and MGS-MT1, the LDH MGS-M4, and the AKR MGS-M5 were expressed and purified according to the procedures described by Lai and colleagues (2011). The purified His_6-tagged MGS-M1, MGS-M2, MGS-MT1, MGS-M4 and MGS-M5 enzymes were crystallized using the sitting drop method, Intelliplate 96-well plates and a Mosquito liquid handling robot (TTP LabTech), which mixed 0.5 μl of protein between 21 to 27 mg ml^{-1} with 0.5 μl of the reservoir solution. The reservoir solutions were as follows: MGS-M1 – 20% potassium fluoride, 20% (w/v) polyethylene glycol (PEG) 3350; MGS-M2 – 0.1 M sodium HEPES pH 7.5, 1.4 M sodium citrate, thermolysin protease; MGS-MT1 – 0.1 M MES pH 6.0, 20% (w/v) PEG 10K; MGS-M4 – 0.1 M Tris pH 8.5, 0.2 M ammonium sulphate, 25% (w/v) PEG 3350, tobacco etch virus protease; MGS-M5 – 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate, 9% PEG8 K, trypsin protease. The crystals were cryo-protected with reservoir solution supplemented with either 12% glycerol, 15% ethylene glycol or Paratone-N oil prior to flash freezing in an Oxford Cryosystems Cryostream. Diffraction data were collected at 100K and the Cu Kα emission wavelength using a Rigaku HF-0007 rotating anode with a Rigaku R-AXIS IV++ detector. Diffraction data were reduced with either HKL3000 (Minor *et al.*, 2006) or XDS (Kabsch, 2010) and SCALA (Evans, 2006). The structures were determined by molecular replacement using the following Protein Databank (PDB) codes: MGS-M1 – 3HXK, MGS-M2 – 2XUA, MGS-MT1 – 3V9A, MGS-M4 – 4FZI, and MGS-M5 – 1LDN. Electrostatic surface representations were calculated using the PDB2PQR server (Dolinsky *et al.*, 2004) with the Assisted Model Building and Energy Refinement (AMBER) force field and otherwise default settings. Substrate-binding cavity volumes were calculated by the CASTp server (Dundas *et al.*, 2006). Structural homologues in the PDB were identified using the PDBeFold server (Krissinel and Henrick, 2004). For halophilic and thermophilic adaptation analysis, we excluded from all analyses residues within 10 Å of the catalytic serine in the α/β hydrolase enzymes or within 5 Å of the bound substrate in the AKR prostaglandin F synthase (PDB 1RY0) to remove sequence substitutions due to substrate-binding differences. Surface or core classification was assigned after calculation of the solvent-accessible surface area (SASA) by NACCESS (http://www.bioinf.manchester.ac.uk/naccess); residues with > 10 $Å^2$ SASA were assigned as surface residues. The difference in hydrophobic core burial was calculated by comparing the apolar buried surface area in the hydrophobic core using NACCESS with the unfolded state SASA calculated by the unfolded server (http://folding.chemistry.msstate.edu/utils/unfolded.html) (Creamer *et al.*, 1997). Multiple sequence alignments were constructed using the CLUSTALW2 tool (http://www.ebi.ac.uk/clustalw/index.html) integrated into the BIOEDIT 7.0.9.1 software (Hall, 1999) and GENO3D (Combet *et al.*, 2002). X-ray diffraction statistics can be found in Table S6. The structural coordinates reported will appear in the PDB under accession codes 4Q3K, 4Q3L, 4Q3M, 4Q3N and 4Q3O.

CD

CD spectra were acquired between 190 and 255 nm with a Jasco J-720 spectropolarimeter equipped with a Peltier temperature (from 4 to 95°C) controller, employing 0.1 cm of path cell, at 25°C. The protein concentration was determined spectrophotometrically (at 280 nm) according to the corresponding amino acid sequence (http://www.expasy.org/tools/protparam.html). Protein solutions were prepared in 50 mM HEPES buffer pH 7.0; for MGS-B1 and MGS-MT1, 0.8 M NaCl was added to the buffer to ensure protein stability during the assay. Spectra were analysed and denaturation temperatures were determined at 220 nm, as reported (Pace and Scholtz, 1997; Schmid, 1997).

Construction of a neighbour-joining tree and oligonucleotide usage pattern analysis

Multiple protein alignments were performed using the CLUSTALW program built into the BIOEDIT software version

7.0.9.0 (Hall, 1999). Phylogenetic analysis of protein sequences was conducted with the MEGA 4.0 software (Tamura *et al*., 2007) using the neighbour-joining tree method and sampling of 1000 trees for bootstrapping and Poisson correction. DNA sequences of contigs were searched against all sequenced bacterial chromosomes, plasmids and phages for oligonucleotide compositional similarity using the GOHTAM web tool (Menigaud *et al*., 2012).

Conclusions

Since their discovery in 1983 (De Lange and Ten Haven, 1983; MEDRIFF Consortium, 1995; Wallmann *et al*., 1997), the number of Mediterranean deep-sea hypersaline lakes that have been described has grown constantly (Chamot-Rooke *et al*., 2005; Yakimov *et al*., 2007; La Cono *et al*., 2011; Yakimov *et al*., 2013). The surface of these brine lakes lies 3.0–3.5 km below sea level, and the salinity of the brines is 5–13 times higher than that of seawater. These lakes are characterized by a moderate-temperature gradient that consistently ranges from 13 to 16.5°C. Microbial populations inhabiting such ancient ecosystems are adapted to operate under harsh physical and chemical conditions, particularly high salinities and high pressures (Daffonchio *et al*., 2006; Smedile *et al*., 2013; Yakimov *et al*., 2013). These conditions are incompatible with life for common marine microorganisms (Harrison *et al*., 2013); however, although increasing evidence suggests that these environmental constraints may impact organism and protein evolution and properties, how and why this process occurs remains to be fully elucidated. The main reason for this lack of knowledge is that the overwhelming majority of autochthonous microbiota resist cultivation, and only a few isolates and enrichments have been obtained from these basins thus far (Antunes *et al*., 2003; 2007; 2008; Albuquerque *et al*., 2012; Yakimov *et al*., 2013; Werner *et al*., 2014).

The biochemical knowledge generated in this study demonstrated that pressure had a marked and consistent effect on the temperature profiles of enzymes from microorganisms inhabiting deep-sea salt-saturated habitats. Thus, by examining the two variables of pressure resistance and optimal temperature, we noticed that the effect of salt, e.g. 348 practical salinity units (PSU), is demonstrated by the link between high pressure and high thermal adaptations. Crystal structure analysis of five enzymes further demonstrated that the salinity level and the protein sequence/structure may play additional significant roles in defining the temperature profile by unknown structural adaptation mechanisms. Manipulating these factors may allow expansion of the lower and upper thermal tolerance limits of microbes inhabiting deep-sea salt-saturated lakes. Notably, marine enzymes that are most active at temperatures as high as 85–130°C have only been identified in deep-sea hydrothermal vent chimneys, whereas the optimal temperatures range from 4 to 60°C for the other deep-sea enzymes reported thus far (Table S1). Nevertheless, although pressure-enhanced activity has been reported in some hyperthermophilic proteins (Michels and Clark, 1997), no examples of proteins that are most active at ≥ 70°C (resembling thermophiles), such as the MGS-M2 enzyme herein reported, have been identified in deep-sea regions other than deep-sea hydrothermal vent chimneys with salinities < 40 PSU.

We hypothesize that the data generated herein may help with the design of new cultivation strategies for the isolation of new thermophiles from moderately warm (14.0–16.5°C) salt-saturated deep-sea lakes. Furthermore, the present study reported the largest set of structures of deep-sea proteins from uncultivable bacteria inhabiting hypersaline lakes (348 PSU) and the hadopelagic water column of the Eastern Mediterranean Sea; these results may provide future implications for our understanding of deep-sea protein adaptation, reaction mechanisms and substrate preferences.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the European Community project MAMBA (FP7-KBBE-2008-226977). This work was further funded by grant BIO2011-25012 from the Spanish Ministry of Economy and Competitiveness (formerly MICINN). HT, OVG and PNG acknowledge European Commission for 'MicroB3' grant (FP7-OCEAN.2011-2 (contract Nr 287589)). This work received support from the Government of Canada through Genome Canada and the Ontario Genomics Institute (grant 2009-OGI-ABC-1405 to A.F.Y. and A.S.) and from the U.S. National Institutes of Health (grants GM074942 and GM094585 to A.S. through the Midwest Center for Structural Genomics). We thank Dr. C. Méndez-García for her excellent support for the preparation of heat map figures and Dr. T.Y. Nechitaylo for the sequence submission to databases. We also thank José M. Andreu and María A. Olivia (CIB-CSIC) for their excellent support in relation to the circular dichroism experiments and analyses. The authors declare that they have no competing interests.

References

Akondi, K.B., and Lakshmi, V.V. (2013) Emerging trends in genomic approaches for microbial bioprospecting. *OMICS* **17:** 61–70.

Albuquerque, D.M., Lopes, J.B., Segundo, L.F., Brandão, T.M., Ribeiro, M.N., Ramos, L., *et al*. (2012) Dehydrated brewery residue for pigs in the growth phase under high temperature conditions. *Rev Bras Zool* **41:** 1784–1788.

Alcaide, M., Tornes, J., Stogios, P.J., Xu, X., Gertler, C., Di Leo, R., *et al*. (2013) Single residues dictate the co-evolution of dual esterases – MCP hydrolases from the α/β hydrolase family. *Biochem J* **454:** 157–166.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Antunes, A., Eder, W., Fareleira, P., Santos, H., and Huber, R. (2003) *Salinisphaera shabanensis* gen. nov., sp. nov., a novel, moderately halophilic bacterium from the brine-seawater interface of the Shaban Deep, Red Sea. *Extremophiles* **7:** 29–34.

Antunes, A., França, L., Rainey, F.A., Huber, R., Nobre, M.F., Edwards, K.J., and da Costa, M.S. (2007) *Marinobacter salsuginis* sp. nov., isolated from the brine-seawater interface of the Shaban Deep, Red Sea. *Int J Syst Evol Microbiol* **57:** 1035–1040.

Antunes, A., Taborda, M., Huber, R., Moissl, C., Nobre, M.F., and da Costa, M.S. (2008) *Halorhabdus tiamatea* sp. nov., a non-pigmented, extremely halophilic archaeon from a deep-sea, hypersaline anoxic basin of the Red Sea, and emended description of the genus *Halorhabdus*. *Int J Syst Evol Microbiol* **58:** 215–220.

Antunes, A., Rainey, F.A., Wanner, G., Taborda, M., Pätzold, J., Nobre, M.F., *et al*. (2008) A new lineage of halophilic, wall-less, contractile bacteria from a brine-filled deep of the Red Sea. *J Bacteriol* **190:** 3580–3587.

Antunes, A., Alam, I., El Dorry, H., Siam, R., Robertson, A., Bajic, V.B., and Stingl, U. (2011) Genome sequence of *Haloplasma contractile*, an unusual contractile bacterium from a deep-sea anoxic brine lake. *J Bacteriol* **193:** 4551–4552.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233–D238.

Chamot-Rooke, N., Rabaute, A., and Kreemer, C. (2005) Western Mediterranean Ridge mud belt correlates with active shear strain at the prism-backstop geological contact. *Geology* **33:** 861–864.

Combet, C., Jambon, M., Deléage, G., and Geourjon, C. (2002) Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* **18:** 213–214.

Creamer, T.P., Srinivasan, R., and Rose, G.D. (1997) Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **36:** 2832–2835.

Daffonchio, D., Borin, S., Brusa, T., Brusetti, L., van der Wielen, P.W., Bolhuis, H., *et al*. (2006) Biodeep scientific party. Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. *Nature* **440:** 203–207.

De Corte, D., Sintes, E., Yokokawa, T., Reinthaler, T., and Herndl, G.J. (2012) Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. *ISME J* **6:** 1566–1577.

De Lange, G.J., and Ten Haven, H.L. (1983) Recent sapropel formations in the eastern Mediterranean. *Nature* **305:** 797–798.

De Vos, D., Xu, Y., Hulpiau, P., Vergauwen, B., and Van Beeumen, J.J. (2007) Structural investigation of cold activity and regulation of aspartate carbamoyltransferase from the extreme psychrophilic bacterium *Moritella profunda*. *J Mol Biol* **365:** 379–395.

Del Pozo, M.V., Fernández-Arrojo, L., Gil-Martínez, J., Montesinos, A., Chernikova, T.N., Nechitaylo, T.Y., *et al*. (2012) Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol Biofuels* **5:** 73.

Delgado-García, M., Valdivia-Urdiales, B., Aguilar-González, C.N., Contreras-Esquivel, J.C., and Rodríguez-Herrera, R. (2012) Halophilic hydrolases as a new tool for the biotechnological industries. *J Sci Food Agric* **92:** 2575–2780.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32:** W665–W667.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* **34:** W116–W118.

Eloe, E.A., Shulse, C.N., Fadrosh, D.W., Williamson, S.J., Allen, E.E., and Bartlett, D.H. (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environ Microbiol Rep* **3:** 449–458.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62:** 72–82.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A., Yakimov, M.M., *et al*. (2005) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* **12:** 895–904.

Ferrer, M., Werner, J., Chernikova, T.N., Bargiela, R., Fernández, L., La Cono, V., *et al*. (2012) Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study. *Environ Microbiol* **14:** 268–281.

Hall, T.A. (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.

Hallsworth, J.E., Yakimov, M.M., Golyshin, P.N., Gillion, J.L., Auria, D., de Lima, G., *et al*. (2007) Limits of life in $MgCl_2$-containing environments: chaotropicity defines the window. *Environ Microbiol* **9:** 801–813.

Harrison, J.P., Gheeraert, N., Tsigelnitskiy, D., and Cockell, C.S. (2013) The limits for life under multiple extremes. *Trends Microbiol* **21:** 204–212.

Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125–132.

Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S.K., Gall, M., *et al*. (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chembiochem* **11:** 1635–1643.

Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60:** 2256–2268.

La Cono, V., Smedile, F., Bortoluzzi, G., Arcadi, E., Maimone, G., Messina, E., *et al*. (2011) Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: prokaryotes and environmental settings. *Environ Microbiol* **13:** 2250–2268.

Lai, K.K., Stogios, P.J., Vu, C., Xu, X., Cui, H., Molloy, S., *et al*. (2011) An inserted α/β subdomain shapes the catalytic pocket of *Lactobacillus johnsonii* cinnamoyl esterase. *PLoS ONE* **6:** e23269.

Lukashin, A.V., and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26:** 1107–1115.

Madern, D., Ebel, C., and Zaccai, G. (2000) Halophilic adaptation of enzymes. *Extremophiles* **4:** 91–98.

Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., *et al*. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl Environ Microbiol* **79:** 3553–3562.

MEDRIFF Consortium (1995) Three brine lakes discovered in the seafloor of the eastern Mediterranean. *EOS Trans AGU* **76:** 313–318.

Menigaud, S., Mallet, L., Picord, G., Churlaud, C., Borrel, A., and Deschavanne, P. (2012) GOHTAM: a website for genomic origin of horizontal transfers, alignment and metagenomics. *Bioinformatics* **28:** 1270–1271.

Michels, P.C., and Clark, D.S. (1997) Pressure-enhanced activity and stability of a hyperthermophilic protease from a deep-sea methanogen. *Appl Environ Microbiol* **63:** 3985–3991.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62:** 859–866.

Nagata, T., Tamburini, C., Aristegui, J., Baltar, F., Bochdansky, A.B., Fonda-Umani, S., *et al*. (2010) Emerging concepts on microbial processes in the bathypelagic ocean – ecology, biogeochemistry and genomics. *Deep Sea Res Part II Top Stud Oceanogr* **57:** 1519–1536.

Nechitaylo, T.Y., Timmis, K.N., and Golyshin, P.N. (2009) *'Candidatus Lumbricincola*', a novel lineage of uncultured *Mollicutes* from earthworms of family Lumbricidae. *Environ Microbiol* **11:** 1016–1026.

Pace, N., and Scholtz, J.M. (1997) Measuring the conformational stability of a protein, Chapter 12. In *Protein Structure: A practical Approach*. Creighton, T.E. (ed.). Heidelberg, Germany: IRL Press at Oxford University Press, pp. 299–322.

Pietra, F. (2012) On 3LEZ, a deep-sea halophilic protein with in vitro class-a β-lactamase activity: molecular-dynamics, docking, and reactivity simulations. *Chem Biodivers* **9:** 2659–2684.

Puspita, I.D., Kamagata, Y., Tanaka, M., Asano, K., and Nakatsu, C.H. (2012) Are uncultivated bacteria really uncultivable? *Microbes Environ* **27:** 356–366.

Reed, C.J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. (2013) Protein adaptations in archaeal extremophiles. *Archaea* **2013:** 373275.

Reyes-Duarte, D., Ferrer, M., and García-Arellano, H. (2012) Functional-based screening methods for lipases esterases and phospholipases in metagenomic libraries. In *Lipases and Phospholipases: Methods and Protocols Methods in Molecular Biology*, Vol. 861. Sandoval, G. (ed.). New York, NY, USA: © Springer Science+Business Media, pp. 101–113.
doi:10.1007/978-1-61779-600-5_6.

Schlitzer, R. (2010) *Ocean data view*. SeaDataNet, Pan-European Infrastructure for Oceand & Marine Data Management, IFREMER/SISMER, Plouzane, France. URL http://odv.awi.de.

Schmid, F.Z. (1997) Optical spectroscopy to characterize protein conformation, Chapter 11. In *Protein Structure: A Practical Approach*. Creighton, T.E. (ed.). Oxford, UK: IRL Press, pp. 261–298.

Shin, D.S., Didonato, M., Barondeau, D.P., Hura, G.L., Hitomi, C., Berglund, J.A., *et al*. (2009) Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *J Mol Biol* **385:** 1534–1555.

Shirai, T., Hung, V.S., Morinaka, K., Kobayashi, T., and Ito, S. (2008) Crystal structure of GH13 alpha-glucosidase GSJ from one of the deepest sea bacteria. *Proteins* **73:** 126–133.

Sineva, E.V., and Davydov, D.R. (2010) Cytochrome P450 from *Photobacterium profundum* SS9, a piezophilic bacterium, exhibits a tightened control of water access to the active site. *Biochemistry* **49:** 10636–10646.

Smedile, F., Messina, E., La Cono, V., Tsoy, O., Monticelli, L.S., Borghini, M., *et al*. (2013) Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ Microbiol* **15:** 167–182.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24:** 1596–1599.

Wallmann, K., Suess, E., Westbrook, G.H., Winckler, G., and Cita, M.B. (1997) Salty brines on the Mediterranean Sea floor. *Nature* **387:** 31–32.

Wang, H., Gong, Y., Xie, W., Xiao, W., Wang, J., Zheng, Y., *et al*. (2011) Identification and characterization of a novel thermostable gh-57 gene from metagenomic fosmid library of the Juan de Fuca Ridge hydrothemal vent. *Appl Biochem Biotechnol* **164:** 1323–1338.

Werner, J., Ferrer, M., Michel, G., Mann, A.J., Huang, S., Juarez, S., *et al*. (2014) *Halorhabdus tiamatea*: proteogenomics and glycosidase activity measurements identify the first cultivated euryarchaeon from a deep-sea anoxic brine lake as potential polysaccharide degrader. *Environ Microbiol* **16:** 2525–2537.

Xu, M., Xiao, X., and Wang, F. (2008) Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles* **12:** 255–262.

Yakimov, M.M., Giuliano, L., Cappello, S., Denaro, R., and Golyshin, P.N. (2007) Microbial community of a hydrothermal mud vent underneath the deep-sea anoxic brine lake Urania (eastern Mediterranean). *Orig Life Evol Biosph* **37:** 177–188.

Yakimov, M.M., La Cono, V., Slepak, V.Z., La Spada, G., Arcadi, E., Messina, E., *et al*. (2013) Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci Rep* **3:** 3554.

Zeng, X., Birrien, J.L., Fouquet, Y., Cherkashov, G., Jebbar, M., Querellou, J., *et al*. (2009) *Pyrococcus* CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *ISME J* **3:** 873–876.

Supporting information

Additional Supporting information may be found in the online version of this article at the publisher's web-site:

Fig. S1. pH profiles of wild-type enzymes. The specific activities were calculated in triplicate as described in the Experimental procedures. The standard deviation (SD) is shown. The 100% activity is as shown in Table 1. Note: due to protein instability at low pH, the pH profile for MGS-M5 could not be obtained; preliminary test reactions indicated pH 8.0 (50 mM Tris-HCl) as being the most suitable buffer for activity

determinations, and this value was used as the standard buffer for this enzyme.

Fig. S2. Substrate profiles of the enzymes with a set of structurally diverse substrates. The specific activities were calculated in triplicate as described in the Experimental procedures, using the standard assay conditions (see also summary conditions in Table 1). Mean values (in log scale) are given. The standard deviation (SD) is not shown due to the logarithmic scale, but it is ≤ 0.23%. Note: using standard conditions for MGS-M3, no activity was detected using *p*NP-α-glucose, *p*NP-α-maltooligosaccharides (C2 to C6), *p*NP-α-galactose, *p*NP-β-galactose, *p*NP-α-xylose, *p*NP-β-arabinopyranose, *p*NP-α-rhamnose, *p*NP-α-mannose, *p*NP-β-mannose, *p*NP-α-fucose, *p*NP-β-glucuronide, carboxymethyl cellulose, laminarin, lichenan and crystalline cellulose.

Fig. S3. Comparison of the structures of MGS enzymes crystallized in this study and their structural homologues revealed by structure similarity searches. Enzymes are shown as cartoon representations. Arrows refer to the locations of the catalytic serine (for esterases in top row) or the NADPH/NADH and substrate-binding sites (for bottom row enzymes).

Fig. S4. Comparison of putative active sites of MGS enzymes crystallized in this study and their structural homologues revealed by structure similarity searches. Sticks are shown for bound substrates and residues predicted to participate in catalytic reactions and/or interact with substrates. The catalytic triads for the esterases in the top row are labelled. The substrate-binding canals for MGS-M2 and HsaD are also shown in solvent-accessible surface representations, coloured by electrostatic potential, highlighting the disparate charge features, which are shown under their respective cartoon images. The non-NADH/NADPH substrate-binding residues for 2,5-DKGR and LDH, plus the equivalent residues in MGS-M4 and MGS-5, are labelled.

Table S1. General features of reported enzymes isolated from deep-sea regions. The data are based on bibliographic records that are specifically cited.

Table S2. Hydrochemistry of selected deep-sea regions, deep-sea libraries and esterase screening statistics.

Table S3. General features and residues potentially involved in catalysis, substrate recognition and thermal and halophilic adaptations in the proteins investigated.

A. General features of esterase-like proteins.

B. Percentage of identity between esterases with the α/β hydrolase fold as determined by the Matcher (EMBOSS package). Matches/alignment lengths (% identity) are specifically indicated.

C. General features of proteins characterized from the *Medee* basin other than esterases.

D. Thermophilic adaptations for proteins with determined crystal structures.

E. Halophilic adaptations for proteins with determined crystal structures.

F. High-pressure adaptations for proteins with determined crystal structures.

Table S4. Compositional similarities between the DNA fragments containing the genes of interest and bacterial genomes as shown by GOHTAM and TBLASTX analyses.

Table S5. List of primers used in the study.

Table S6. X-ray diffraction statistics.

Appendix S1. Supplementary Results and Discussion.

CAPÍTULO 5
ESTERASAS DE *Rimicaris exoculata*

RESUMEN DEL CONTENIDO

En la línea de trabajo de identificar enzimas de interés en ambientes poco explorados, en este cuarto capítulo se presenta la identificación y caracterización de 3 esterasas de la super-familia de las α/β-hidrolasas de una comunidad bacteria de las branquias de una especie de gamba, *Rimicaris exoculata*, que vive a 2,320 m de profundidad en zonas cercanas a una chimenea hidrotermal en la cuenca de la Dorsal Mesoatlántica. *R. exoculata* vive en zonas cercanas a las chimeneas en el gradiente entre los fluidos hidrotermales y el agua marina, que tiene una temperatura entre 3 y 25ºC (media de 8.7ºC), y una salinidad de aproximadamente 23 g/l de sal (inferior a la del agua de mar). Hasta la fecha, no se han identificado y caracterizado enzimas de bacterias que viven asociadas a la superficie (epibiontes) de macro-organismos que habitan ambientes marinos profundos.

Desde un punto de vista metodológico se han empleado todas las técnicas descritas en el resumen del Capítulo 2, junto con estudios de dicroísmo circular para calcular la temperatura de desnaturalización de las proteínas y estudios de la actividad específica a altas presiones (empleados también en el Capítulo 3). Tras rastrear con un sustrato modelo (acetato de naftilo) un total de 27,200 clones (816 Mpb) se obtuvieron 10 clones positivos. Esto supone una incidencia de 1 clone positivo por cada 2,720 clones. Tras la secuenciación del DNA contenido en los mismos se identificaron 3 genes que codificaban presuntas esterasas de la super-familia de las α/β-hidrolasas. La masa molecular y el punto isoeléctrico de las enzimas estudiadas oscila entre 31,792 y 84,278 Da, y 5.55 y 6.21, respectivamente. A nivel de secuencia las enzimas identificadas y analizadas presentan homologías a nivel de identidad entre el 41% y el 72% con otras descritas en las bases de datos y entre el 11.9 y 63.7% entre sí. El pH óptimo varía entre 8.0 y 8.5, la temperatura óptima de actividad oscila entre los 30 y los 50ºC y las concentraciones de sal (NaCl) para actividad óptima varían entre 0 y 3.6 M.

Entre los resultados más relevantes en este trabajo hemos demostrado que el estudio de actividades enzimáticas en ambientes no estudiados con anterioridad puede proporcionar no solo nuevos conocimientos a nivel básico sino también aplicado. A nivel básico los estudios presentados en este capítulo revelan por primera vez las características de enzimas de un grupo taxonómico, concretamente proteobacterias relacionadas con el género *Thiothrix*/*Leucothrix*, que pertenece a una región del árbol filogenético poco caracterizado a nivel bioquímico. Así se ha demostrado que las esterasas de estas bacterias muestran actividad enzimática a mayor temperatura, salinidad y presión que otras enzimas de otras bacterias aisladas del mismo ambiente. Desde un punto de vista biotecnológico, y tras el estudio comparativo con una biblioteca de 131 ésteres y 31 esterasas y lipasas descritas en la bibliografía, comerciales o descritas en los capítulos anteriores de esta Tesis Doctoral, se ha profundizado en dos preguntas: i) si la especificidad de sustrato de las enzimas viene marcada por las condiciones geoquímicas de los ambientes de procedencia; y ii) si *R. exoculata* contiene enzimas con potencial biotecnológico. Los datos experimentales presentados sugieren ambos hechos, que son inéditos en la bibliografía.

El CD adjunto a esta Memoria contiene Material Suplementario relacionado con este Capítulo.

AEM
Journals.ASM.org

Identification and Characterization of Carboxyl Esterases of Gill Chamber-Associated Microbiota in the Deep-Sea Shrimp *Rimicaris exoculata* by Using Functional Metagenomics

María Alcaide,[a] Anatoli Tchigvintsev,[b] Mónica Martínez-Martínez,[a] Ana Popovic,[b] Oleg N. Reva,[c] Álvaro Lafraya,[a*] Rafael Bargiela,[a] Taras Y. Nechitaylo,[d] Ruth Matesanz,[e] Marie-Anne Cambon-Bonavita,[f] Mohamed Jebbar,[g] Michail M. Yakimov,[h] Alexei Savchenko,[b] Olga V. Golyshina,[i] Alexander F. Yakunin,[b] Peter N. Golyshin,[i] Manuel Ferrer,[a] for The MAMBA Consortium

Consejo Superior de Investigaciones Científicas (CSIC), Institute of Catalysis, Madrid, Spain[a]; Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada[b]; Department of Biochemistry, University of Pretoria, Pretoria, South Africa[c]; Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology, Jena, Germany[d]; Centro de Investigaciones Biológicas, CSIC, Madrid, Spain[e]; Ifremer, Centre de Brest, Laboratoire de Microbiologie des Environnements Extrêmes, REM/DEEP/LM2E, UMR 6197 (Ifremer-CNRS-UBO), ZI de la Pointe du Diable, Plouzané, France[f]; Université de Bretagne Occidentale, Laboratoire de Microbiologie des Environnements Extrêmes-UMR 6197 (CNRS-Ifremer-UBO), Plouzané, France[g]; Institute for Coastal Marine Environment, CNR, Messina, Italy[h]; School of Biological Sciences, Bangor University, Gwynedd, United Kingdom[i]

The shrimp *Rimicaris exoculata* dominates the fauna in deep-sea hydrothermal vent sites along the Mid-Atlantic Ridge (depth, 2,320 m). Here, we identified and biochemically characterized three carboxyl esterases from microbial communities inhabiting the *R. exoculata* gill that were isolated by naive screens of a gill chamber metagenomic library. These proteins exhibit low to moderate identity to known esterase sequences (≤52%) and to each other (11.9 to 63.7%) and appear to have originated from unknown species or from genera of *Proteobacteria* related to *Thiothrix*/*Leucothrix* (MGS-RG1/RG2) and to the *Rhodobacteraceae* group (MGS-RG3). A library of 131 esters and 31 additional esterase/lipase preparations was used to evaluate the activity profiles of these enzymes. All 3 of these enzymes had greater esterase than lipase activity and exhibited specific activities with ester substrates (≤356 U mg^{-1}) in the range of similar enzymes. MGS-RG3 was inhibited by salts and pressure and had a low optimal temperature (30°C), and its substrate profile clustered within a group of low-activity and substrate-restricted marine enzymes. In contrast, MGS-RG1 and MGS-RG2 were most active at 45 to 50°C and were salt activated and barotolerant. They also exhibited wider substrate profiles that were close to those of highly active promiscuous enzymes from a marine hydrothermal vent (MGS-RG2) and from a cold brackish lake (MGS-RG1). The data presented are discussed in the context of promoting the examination of enzyme activities of taxa found in habitats that have been neglected for enzyme prospecting; the enzymes found in these taxa may reflect distinct habitat-specific adaptations and may constitute new sources of rare reaction specificities.

Metagenomics provides a means for the discovery of entirely new enzymes in microorganisms and their communities without the need to culture these microorganisms as individual species, which is technically very difficult (1–5). Although the discovery of new enzyme activities has progressed considerably (6), it has not managed to go beyond the effective identification of enzymatic activities at a rather limited number of environmental sites. While an extensive search in the specialized literature and public databases indicated that microbial communities from approximately 1,800 different sites worldwide have been examined for their genomic content, only in approximately 200 of those locations (11% of the total) have new active clones or enzymes been identified and/or partially characterized. Thus, only a tiny fraction of the Earth's biosphere has been explored for the purpose of enzyme discovery, despite the fact that natural microbial diversity has been recognized to be the major source of new microbes (7). Such diversity also contains novel genes and functions (8) whose activities remain mostly unknown but whose potential to contribute to an innovation-based economy is recognized (9, 10).

One of the habitats less explored to date in terms of examining enzyme repertoires is the deep-sea realm, where microbial communities are shaped by salinity, pressure, and temperature (11). The examination of a set of enzymes recently collected from free-living bacteria from a number of deep-sea, salt-saturated marine habitats (3) reveals that deep-sea marine habitats harbor broad biochemical diversity reflective of the unique environmental conditions under which they evolved. An investigation of these enzymes also demonstrates that enzymes may constitute good model

Received 17 October 2014 **Accepted** 5 January 2015

Accepted manuscript posted online 16 January 2015

Citation Alcaide M, Tchigvintsev A, Martínez-Martínez M, Popovic A, Reva ON, Lafraya Á, Bargiela R, Nechitaylo TY, Matesanz R, Cambon-Bonavita M-A, Jebbar M, Yakimov MM, Savchenko A, Golyshina OV, Yakunin AF, Golyshin PN, Ferrer M, MAMBA Consortium. 2015. Identification and characterization of carboxyl esterases of gill chamber-associated microbiota in the deep-sea shrimp *Rimicaris exoculata* by using functional metagenomics. Appl Environ Microbiol 81:000–000. doi:10.1128/AEM.03387-14.

Editor: P. Schloss

Address correspondence to Manuel Ferrer, mferrer@icp.csic.es, or Peter N. Golyshin, p.golyshin@bangor.ac.uk.

M.A. and A.T. contributed equally to this work.

* Present address: Álvaro Lafraya, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP), Institut National des Sciences Appliquées, Toulouse, France.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03387-14.

doi:10.1128/AEM.03387-14

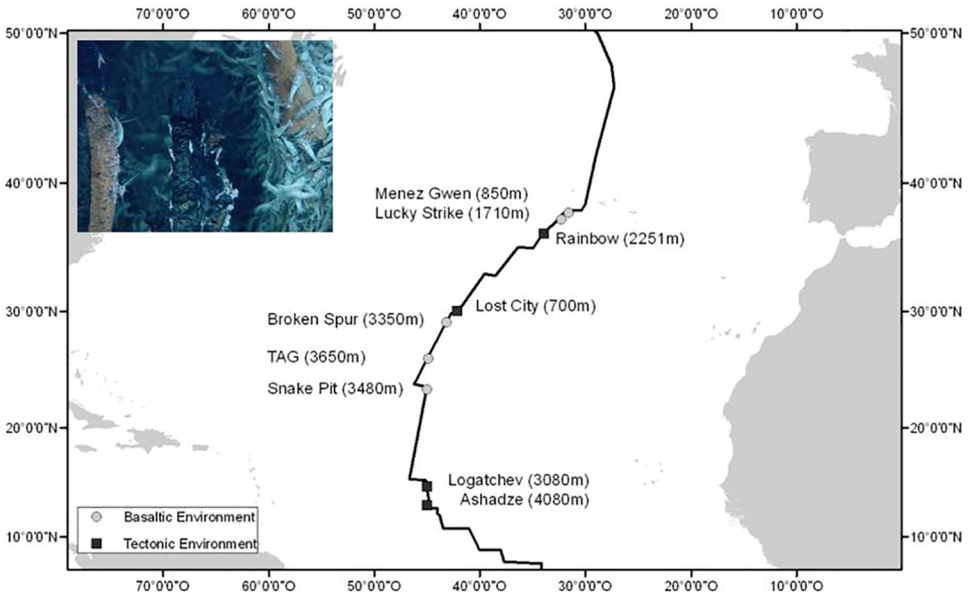


FIG 1 Global map displaying the locations of the deep-sea sites at which specimens of *R. exoculata* were sampled. Hydrochemistry data for the Mid-Atlantic Ridge (19, 20) are the following: density, 1.000 g liter^{-1}; depth, 2,320 m; total salinity, 23.03 g kg^{-1} (Na^+, 12.71 g kg^{-1}; K^+, 0.79 g kg^{-1}; Ca^{2+}, 2.67 g kg^{-1}; Cl^-, 12.75 g kg^{-1}; Mg^{2+}, 0 g kg^{-1}; SO_4^{2-}, 0 g kg^{-1}); temperature (average), 8.7°C; pH, 7.0 to 8.0. (Map reproduced with permission from Ifremer; photograph reproduced with permission from Ifremer-Victor/Exomar 2005.)

systems for understanding protein functionality and adaptation under multiple extremes. Although the deep sea is also home to macroorganisms and their associated microbial communities (12, 13), to our knowledge, no previous investigation of the characteristics and habitat-specific adaptations of enzymes from microbial communities associated with deep-sea macroorganisms has been reported. Moreover, only a small number of enzymes from marine sponge- and marine mollusc-associated bacterial communities (14–18) have been characterized, none of which were isolated from marine habitats below a depth of 200 m. Most of these enzymes, for which limited biochemical data are available, show significant homology (54 to 100%) to known proteins, are halotolerant, and exhibit an activity-stability trade-off observed in cold-adapted enzymes (16); however, limited biochemical information is known with respect to their substrate profiles. Thus, our understanding of the metabolic significance and protein functionality of bacterial enzymes associated with macroorganisms living within extreme deep-sea habitats is limited.

In this study, we analyzed enzymes from the shrimp *Rimicaris exoculata* (Williams and Rona, 1986), which dominates the fauna at many hydrothermal vent sites along the Mid-Atlantic Ridge (<2,320 m depth). *R. exoculata* thrives along deep chimney walls in the gradient between hydrothermal fluids and cold, oxygenated, and slightly saline ambient waters (19). Based on functional gene surveys, it was found that shrimp specimens, including adult *R. exoculata*, host a dense bacterial community within their gill chambers (12). Such bacteria are known to provide energy and nutrients for the eukaryotic host by cometabolizing inorganic and organic compounds (e.g., H_2 and CH_4) directly discharged by the hydrothermal system (20–24) or obtained from the water column (25, 26). It remains to be determined, however, how the multiple collective pressures characterizing the habitat in which *R. exoculata* thrives impact the properties and functional characteristics of microbial enzymes such as those from the gill chamber environment. Following on from this, we aimed to understand the influence of marine conditions, particularly the combined effects of salt, pressure, and temperature, on the properties and functional characteristics of enzymes, as well as to increase the number of characterized enzymes from deep-sea environments. We sequenced and analyzed the biochemical characteristics of three bacterial carboxyl esterases with an α/β hydrolase fold that were identified by naive screens of deep-sea microbiota associated with the gill chambers of *R. exoculata* shrimp inhabiting hydrothermal fluids along the Mid-Atlantic Ridge (Fig. 1). These enzymes first were selected as model enzymes for study because they have been reported to be suitable enzymes to investigate protein functionality and adaptation to polyextremes (3), particularly as a function of temperature, salt, and pressure in deep-sea habitats. In this context, our recent investigation examining the characteristics of a set of esterases has provided the first clear evidence suggesting that, in salt-saturated deep-sea habitats, the adaptation to high pressure is linked to high thermal resistance by unknown structural adaptation mechanisms (3). Therefore, understanding the characteristics of other deep-sea enzymes, particularly esterases for which multiple biochemical data are available compared to other deep-sea enzymes (3), might help in understanding the phenomenon of protein versatility and adaptability to different polyextremes. Second, esterases were selected because they are ubiquitous enzymes that are widespread in nature (at least one per genome), multiple commercial preparations that are used at industrial scale are available, and biochemical data of such enzymes from a number of marine habitats (3, 27, 28) have been reported recently, which collectively may assist to perform an extensive comparative analysis. Finally, they are of great interest as biocatalysts for chemical synthesis (29). A comparative analysis using a set of 31 different enzymes with esterase and lipase activity, including enzymes from marine (deep-sea, hydrothermal vent, and superficial seawater) and brackish water habitats and commercial

preparations, allowed us to infer habitat-specific features of the gill-associated bacterial enzymes as well as to evaluate their specificities and reactivities. Note that two of the enzymes reported in this study originate from an unknown species or possibly from genera related to *Thiothrix* and/or *Leucothrix*, a group of organisms largely neglected with respect to enzyme discovery; only a small number of isolates from these genera are available, and no enzymes from these genera have been characterized.

MATERIALS AND METHODS

Metagenomic library, metagenome screening, and sequence analysis. Gill chamber samples were collected from *R. exoculata* specimens found at locations in the western flank of the Rainbow Ridge in the Atlantic Ocean (36°14′N, 33°54′W; 2,320 m depth; MAR site), where the average seawater temperature is 8.7°C (ranging from 3 to 25°C) (Fig. 1) (19). Total DNA was extracted from the gill chambers of the collected specimens as previously described (19); from this DNA, a large-insert pCCFOS1 fosmid library was generated using the *Escherichia coli* EPI300-T1^R strain (Epicentre Biotechnologies; Madison, WI, USA), and the library was scored for the ability to hydrolyze α-naphthyl acetate and tributyrin (27, 30). Positive clones were selected, and their DNA inserts were sequenced using a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Life Sequencing SL (Valencia, Spain) or were completely Sanger sequenced using universal primers and subsequent primer walking. Upon completion of sequencing, the reads were assembled to generate nonredundant metasequences using Newbler GS De Novo Assembler v.2.3 software (Roche, Branford, CT, USA). GeneMark software (31) was employed to predict potential protein-coding regions (open reading frames [ORFs] with ≥20 amino acids) from the sequences of each assembled contig, and deduced proteins were screened using BLASTP and PSI-BLAST searches (32). Multiple protein alignments were performed using the ClustalW program built into BioEdit software, version 7.0.9.0 (33). Domains with a significant propensity to form transmembrane helices were identified with TMpred software (ExPASy, Swiss Institute of Bioinformatics) (34).

Protein expression and purification. The cloning, expression, and purification of selected proteins using the p15TV-Lic vector and *E. coli* BL21(DE3) Codon Plus-RIL (for MGS-RG3) and Ek/LIC 46 and *E. coli* BL22 (for MGS-RG1 and MGS-RG2) were performed as described previously (27) using the following primer pairs: MGS-RG1Fwd (5′-GAC GAC GAC AAG ATG ACT GAT TTG TTA CC-3′), MGS-RG1Rev (5′-GAG GAG AAG CCC GGT CAA ATC AAA AC-3′); MGS-RG2Fwd (5′-ATG ATG AAA AAT ATG TCA GAA TTA CC-3′), MGS-RG2Rev (5′-TTA TTT TAA TAT TTT TTG TAG CCA TGC C-3′); and MGS-RG3Fwd (5′-TTG TAT TTC CAG GGC ATG CAG GAA CTT CCC GAT GC-3′), MGS-RG3Rev (5′-CAA GCT TCG TCA TCA GGT CCG GAA CCG CGC CTT G-3′). Purity greater than 98% was obtained after a single His_6 tag purification step (see Fig. S1 in the supplemental material). After purification, all enzymes were maintained at a concentration of 2 mg ml^{-1} in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, pH 7.0; the protein stock solution was stored at −20°C until it was used in assays. All chemicals used for enzymatic tests were of the purest grade available and were purchased from either Fluka-Aldrich-Sigma Chemical Co. (St. Louis, MO, USA) or Apin Chemicals (Oxon, United Kingdom).

Hydrolase assays. Carboxyl esterase activity was assayed using *p*-nitrophenyl (*p*NP) esters (read at 410 nm) and structurally diverse esters other than *p*NP esters (read at 540 nm) in 96-well plates as previously described (27). Unless stated otherwise, standard assay reactions were conducted by adding 2 μl of the 2 mg ml^{-1} protein stock solution to an assay mixture containing 2 μl of ester stock solution (100 mM in acetone [for *p*NP esters] or acetonitrile [for other esters]) in 196 μl of 50 mM Tris-HCl buffer, pH 8.0 (for *p*NP esters), or 5 mM *N*-(2-hydroxyethyl) piperazine-*N*′-(3-propanesulfonic acid) (EPPS) buffer, pH 8.0 (for other esters). The final volume of the assay was 200 μl, and the final protein and substrate concentrations were 10 μg ml^{-1} and 1 mM, respectively. All assays were conducted at pH 8.0 at the optimal temperature for each enzyme (MGS-RG1, 45°C; MGS-RG2, 50°C; MGS-RG3, 30°C). In all cases, absorbance was determined using a microplate reader every 1 min for a total time of 15 min (Synergy HT multi-mode microplate reader; BioTek). All reactions were performed in triplicate. One unit of enzyme activity was defined as the amount of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions. All values were corrected for nonenzymatic hydrolysis (background rate). If not otherwise stated, 1 mM *p*NP-propionate was used as the standard assay substrate for the determination of the conditions under which each enzyme displayed activity; pH values between 4.5 and 9.0 (at the optimal temperature), temperatures between 4 and 80°C (using 50 mM Tris-HCl buffer, pH 8.0), and NaCl, KCl, and $MgCl_2$ concentrations of up to 4 M (using 50 mM Tris-HCl buffer, pH 8.0, and optimal temperatures) were tested. The buffers used to determine the optimal pH for each enzyme have been described previously (3, 27).

Pressure perturbation studies. The effect of hydrostatic pressure on enzymatic activity was analyzed by placing the reaction mixtures in a high-pressure incubating system consisting of a 2-ml high-pressure cell connected to a pressure generator (High Pressure Equipment, Erie, PA, USA) capable of generating pressures of up to 10,000 lb/in^2. Two pressure values were examined: 23,000 and 35,000 kPa. A total of 20 μg of pure protein was added to 20 ml of freshly prepared 50 mM Tris-HCl buffer, pH 8.0, containing 1.0 mM *p*NP-propionate. A total of 2 ml of the reaction mixture was immediately transferred to the high-pressure reactor, and a separate 2-ml aliquot of the mixture was kept at atmospheric pressure. The experiments were performed at 25°C. In all cases, reactions without protein were used as negative controls. After incubation for 5 min, the samples were immediately depressurized, and the extent of the reaction was monitored by spectrophotometry at 410 nm (as described above). All experiments were performed in triplicate. Note that reactions were performed in (i) the absence of salt, (ii) the presence of 23.03 g kg^{-1} NaCl, which corresponds to the habitat salinity (19), and (iii) the presence of NaCl concentrations required for optimal enzyme activity. In all three cases, similar effects by pressure on enzyme activity were found, and only data in the absence of salt are presented.

CD. Circular dichroism (CD) spectra were acquired between 190 and 255 nm using a Jasco J-720 spectropolarimeter equipped with a Peltier temperature controller (4 to 95°C); the spectra were obtained using a 0.1-cm-path cell and were conducted at 25°C. The protein concentration was determined spectrophotometrically at 280 nm in accordance with the amino acid sequence of the protein (www.expasy.org/tools/protparam.html). The spectra were analyzed, and denaturation temperatures were determined in 50 mM HEPES buffer, pH 7.0, at 220 nm as previously reported (35, 36).

Taxonomic binning. Protein-coding genes in fosmid sequences were predicted using GeneMark software (31) and the antiSMASH server (37) and then were checked manually using the Artemis browser (38). The predicted protein sequences were aligned against the National Center for Biotechnology Information nonredundant (NCBI nr) database using a BLASTP search. Taxonomic binning of the orphan sequences in fosmids was performed by summarizing the top significant BLASTP hits with E values of ≤0.00001. Additionally, genes encoding esterases and their closest homologs identified by BLASTP were compared by patterns of codon usage biases predicted using a Web-based codon usage calculator (39). Composition-based binning of fosmids was performed using the GOHTAM web server (40). The dissimilarity in synonymous codon usage between two genes was calculated as the Pearson correlation between vectors of fraction values for 59 codons (41). Compositional similarity between fosmid insert sequences was estimated using the SeqWord Genome Browser (42).

Sequence and substrate profile distributions. Multiple protein alignment was conducted using the MUSCLE application in BioEdit software (33) with default settings (43). The resulting alignment then was used for

TABLE 1 General features of the enzymes and amino acid residues potentially involved in catalysis

Enzyme	Molecular mass (Da)	pI	No. of amino acids	Catalytic triad	GXSXG-lipase motif[a]	Oxyanion hole motif[a]	Classified family
MGS-RG1	24,983.6	5.07	222	S118, D/E, H?	G116-F117-S118-Q119-G120-G121	H24-G25-L26-G27-A28	Possibly VI
MGS-RG2	25,363.2	5.31	225	S111, D175, H206	G109-T110-S111-V112-E113-K114	H27-G28-L29-G30-A31	Possibly VI
MGS-RG3	30,611.2	5.51	277	S125, E218, H248	G123-D124-S125-A126-G127-G128	H55-G56-G57-G58-Y59	Possibly IV

[a] Consensus GXSXG-lipase motifs are given for families (46–48). Family IV, GDSAGG; family VI, NGGPG.

phylogenetic reconstruction. A neighbor-joining tree was constructed in MEGA v.6.06 (44) using the settings for the Poisson model and homogenous patterning between lineages. Bootstrap analysis was done with 1,000 sample trees. The scale bar in the figure reflects the number of substitutions per position. Clustering of the substrate spectral profile based on a binomial distribution of the presence or absence of activity for particular ester substrates was performed, and Pearson's correlation coefficient was used to calculate the distances.

Nucleotide sequence accession numbers. The DNA sequences of the carboxyl esterase-positive metagenomic DNA fragments and the polypeptide sequences of the enzymes were deposited in GenBank under the accession numbers KF831416 (for MGS-RG1), KF831417 (for MGS-RG2), and KC986402 (for MGS-RG3).

RESULTS

Metagenome library construction and screening for carboxyl esterase activity. A subset of the 27,200 clones from the *R. exoculata* gill chamber microbiome library generated in this study, which included nearly 816 Mbp of community genomes, was scored for the ability of individual clones to hydrolyze α-naphthyl acetate, as reported previously (30). The subset of clones also was screened for carboxyl esterase activity using tributyrin plates (30). Hydrolysis of both substrates was indicative of esterase/lipase activity. A total of 10 unique positive clones, corresponding to a hit rate of 1:2,720, were identified as active; 3 of these were selected based on activity phenotype (they were among the most active, as judged by their halos/color formation). The inserts were sequenced, analyzed, and compared to the sequences available in the public NCBI nr database (32). Three predicted metagenome sequence (MGS)-encoding carboxyl esterases from *Rimicaris* gill chambers and with an α/β hydrolase fold were identified and successfully expressed as soluble proteins in *E. coli* (see Fig. S1 in the supplemental material), and their properties were investigated. These three proteins were named MGS-RG1, MGS-RG2, and MGS-RG3 (Table 1).

Sequence analysis of carboxyl esterases from the *Rimicaris* gill chamber microbiome. Based on BLAST searches of the NCBI nr database, the 3 studied protein sequences belong to the α/β hydrolase superfamily and had up to 58% sequence identity with homologous proteins in this database. The MGS-RG1 and MGS-RG2 proteins were most similar to presumptive carboxyl esterases from bacteria of the *Thiotrichales* order (41 to 58% similarity; WP_022950445 and Protein Data Bank [PDB] code 4f21 [45]), whereas MGS-RG3 was most similar to α/β hydrolases from *Rhodobacteraceae* (53% similarity; WP_005862880) and esterases/lipases from uncultured bacteria (41% similarity; PDB code 3V46). The deduced molecular masses and estimated pI values of these proteins ranged from 24.9 to 30.69 kDa and from 5.07 to 5.51, respectively (Table 1). The pairwise amino acid sequence identity ranged from 11.94% to 63.7%; MGS-RG1 and MGS-RG2 were the most similar to each other (63.7% sequence identity), whereas MGS-RG1 and MGS-RG3 (17.3%) and MGS-RG2 and MGS-RG3 (11.9%) were the most divergent at the sequence level. The selected α/β hydrolases contain a classical Ser-Asp-His catalytic triad, but the catalytic elbow and oxyanion hole [i.e., the GXSXGG and H/N-GGG(A)/P-X motifs] often diverged from the consensus, as revealed by sequence analysis (Table 1). Nonetheless, the degree of sequence conservation among these catalytic motifs and in the overall enzyme sequence was sufficient to categorize the enzymes into the following two esterase/lipase subfamilies (46–48): family VI (MGS-RG1 and MGS-RG2) and family IV (MGS-RG3) (see Fig. S2 in the supplemental material). No significant rigorously confirmed transmembrane regions were identified in any of the sequences with TMpred (34); thus, the proteins most likely are intracellular proteins.

A search against the GOHTAM database (40) and TBLASTX analysis revealed compositional similarities between the DNA fragment (6,756 bp) containing the gene for MGS-RG3 with genomes of the alphaproteobacterial genus *Rhodobacteraceae* (best hit score, 248), whereas the DNA fragments encoding MGS-RG1 (24,228 bp) and MGS-RG2 (29,267 bp) exhibited similarity to chromosomal sequences of *Gammaproteobacteria*. A BLASTP search demonstrated that the MGS-RG1- and MGS-RG2-coding genes were most similar to those found in the *Thiothrix/Leucothrix* genomes (49–51). DNA fragments bearing MGS-RG1, MGS-RG2, and MGS-RG3 do not include any insertion sequence elements, integrases, or any other phage- or conjugative plasmid-associated genes. The tetranucleotide usage patterns of MGS-RG1 and MGS-RG2 were similar to each other and homogenous for 5,000-bp sliding windows stepping by 500 bp. MGS-RG3 was too short for the above-described analysis of oligonucleotide word usage patterns to be performed using the SeqWord Genome Browser (42). The esterases encoded by MGS-RG1 and MGS-RG2 exhibited very similar synonymous codon usage, with Pearson correlations from 0.81 to 0.84, but they were distinguishable from patterns of synonymous codon usage biases in the homologous genes from *Thiothrix* and *Leucothrix*; for these, the Pearson correlation coefficient was in the range of 0.42 to 0.52. It may be concluded that MGS-RG1 and MGS-RG2 originated from chromosomes of organisms closely related to the *Thiothrix/Leucothrix* lineage, although these two proteins were not members of these genera. This agrees with the close phylogenetic positioning of the sequences encoding such esterases (see Fig. S2 in the supplemental material).

Determination of the requirements for enzyme activity. Using *p*NP-propionate as a model substrate, the purified proteins

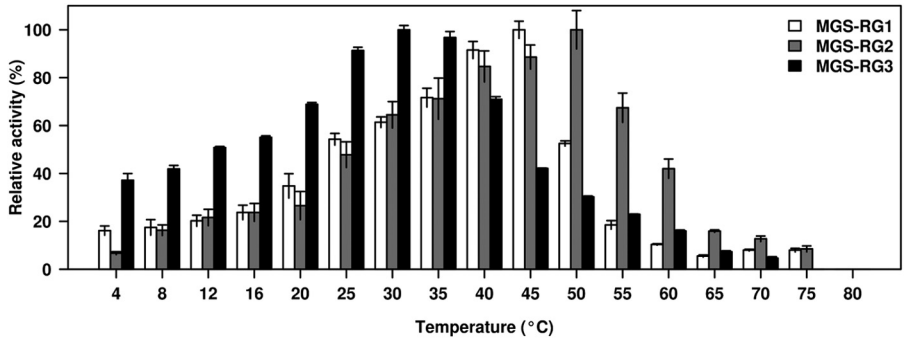


FIG 2 Temperature profiles of the carboxyl esterases from microorganisms inhabiting the gill chamber of *R. exoculata*. The data represent the relative percentages of specific activity (U mg^{-1}) compared with the maximum activity (100%; MGS-RG1, 0.756 U mg^{-1}; MGS-RG2, 1.754 U mg^{-1}; MGS-RG3, 159.2 U mg^{-1}). The specific activities were calculated using 2 μg of protein and 1 mM *p*NP-propionate as the assay substrate at pH 8.0 in 50 mM Tris-HCl, as described in Materials and Methods. Standard deviations (SDs) of the results of assays conducted in triplicate are shown.

exhibited optimal activities at pH values ranging from 8.0 to 8.5 (see Fig. S3 in the supplemental material). At the optimal pH for each enzyme, MGS-RG1 and MGS-RG2 were most active at 45 and 50°C, respectively, and retained ≤16% activity at ≥65°C and 4°C (Fig. 2). In contrast, MGS-RG3 was most active at lower temperatures (its maximal activity occurred at 30°C) and retained approximately 40% activity at 4°C, whereas it was strongly inhibited at ≥45°C. The temperatures found to be optimal for the activity of each enzyme are in agreement with the corresponding protein-denaturing temperatures, as determined by circular dichroism (MGS-RG1, 47.6°C; MGS-RG2, 53.3°C; MGS-RG3, 42°C) (Table 2).

The enzymatic activities of purified MGS-RG1/MGS-RG2 proteins, tested at optimal pH and temperature, were further shown to be stimulated up to 2.4-fold by the addition of NaCl, KCl, and $MgCl_2$ to the reaction mixture (Fig. 3). Maximal activity for MGS-RG1 and MGS-RG2 was achieved at 3.6 and 3.2 M for NaCl, 3.0 and 2.4 M for KCl, and 0.8 and 0.4 M for $MgCl_2$, respectively. In contrast, MGS-RG3 was strongly inhibited by all three salts (Fig. 3). Note that at 2 M NaCl, commercial esterase/lipase preparations, such as CalA, CalB, Novozym 388 L, and porcine liver esterase (PLE), retained ≤9.1% of the activity they display in the absence of salt (Table 2).

Because pressure is one of the most representative environmental parameters of deep-sea habitats, including hydrothermal vents, the influence of hydrostatic pressure (23,000 and 35,000 kPa) on enzyme performance was further evaluated in high-pressure 2-ml reactors at pH 8.0 and 25°C as described in Materials and Methods. The data presented in Table 2 show that MGS-RG1 and MGS-RG2 are barophilic enzymes; at 23,000 kPa, corre-

TABLE 2 Denaturing temperatures and pressure resistance for activity of the esterases reported here and for other esterases from marine habitats

Enzyme	Origin	T_d[a] (°C)	Relative activity (%) at:		
			23,000 kPa[b]	35,000 kPa[c]	2.0 M NaCl[d]
MGS-RG1	*R. exoculata* enzymes	47.6	152.0 ± 7.4	97.9 ± 3.5	173 ± 2.7
MGS-RG2		53.3	144.0 ± 7.5	97.7 ± 4.5	136 ± 7.1
MGS-RG3		42.0	77.9 ± 6.7	57.2 ± 2.9	9.3 ± 3.2
CalA[e]	Commercial preparations	ND/NA	2.0 ± 0.1	0.2 ± 0.01	7.0 ± 1.5
CalB[e]		53.0	8.0 ± 0.4	3.1 ± 0.02	9.1 ± 4.4
Novozym 388 L[e]		ND	5.7 ± 0.2	0.4 ± 0.08	6.8 ± 4.0
PLE[e]		ND/NA	1.8 ± 0.1	0.2 ± 0.05	1.5 ± 0.7
MGS-M1[f]	Deep-sea enzymes	65.2	ND/NA	45.3 ± 3.5	289 ± 5.6
MGS-M2[f]		71.4	ND/NA	110 ± 14	1645 ± 32.3
MGS-B1[f]		52.4	ND/NA	115 ± 8	84 ± 3.4
MGS-K1[f]		40.3	ND/NA	28 ± 7	148 ± 7.1
MGS-MT1[f]		55.7	ND/NA	150 ± 15	170 ± 9.4
ABO_1197, ABO_1251, MGS0010, MGS0105, and MGS0109[g]	Superficial seawater enzymes	45.7–48.1	ND/NA	ND/NA	~125–40

[a] Protein denaturation temperature (T_d) as determined by circular dichroism. T_d reported by Qian and Lutz (52). For other commercial preparations, T_d are not shown either because the values, to the best of our knowledge, have not been not been reported in the specialized literature or the commercial preparations are not pure enough to perform CD experiments in the present study (they are available as a complex mixture of proteins and additives). ND/NA, data not determined or not available.

[b] Relative activity at the sampling site (12) refers to that at atmospheric pressure.

[c] Pressure selected for comparative purposes, as previously reported deep-sea hydrolases have been tested under this pressure (3). The relative activity refers to that shown at atmospheric pressure.

[d] The relative activity in the presence of 2 M NaCl refers to that shown in the absence of NaCl (data for MGS-RG1 to MGS-RG3 from Fig. 3). Activity measured at 30°C and 50 mM Tris-HCl buffer, pH 8.0, 10 μg ml^{-1} total protein, and 1 mM *p*NP-propionate.

[e] Commercial esterase/lipase preparations from Novozymes A/S (Bagsværd, Denmark) (CalA, CalB, and Novozyme 388 L) and Sigma Chemical Co. (St. Louis, MO, USA) (PLE).

[f] Esterases from deep-sea basins of the Mediterranean Sea. Values were reported in reference 3.

[g] Esterases from superficial seawater at the Mediterranean Sea. Values were reported in reference 28.

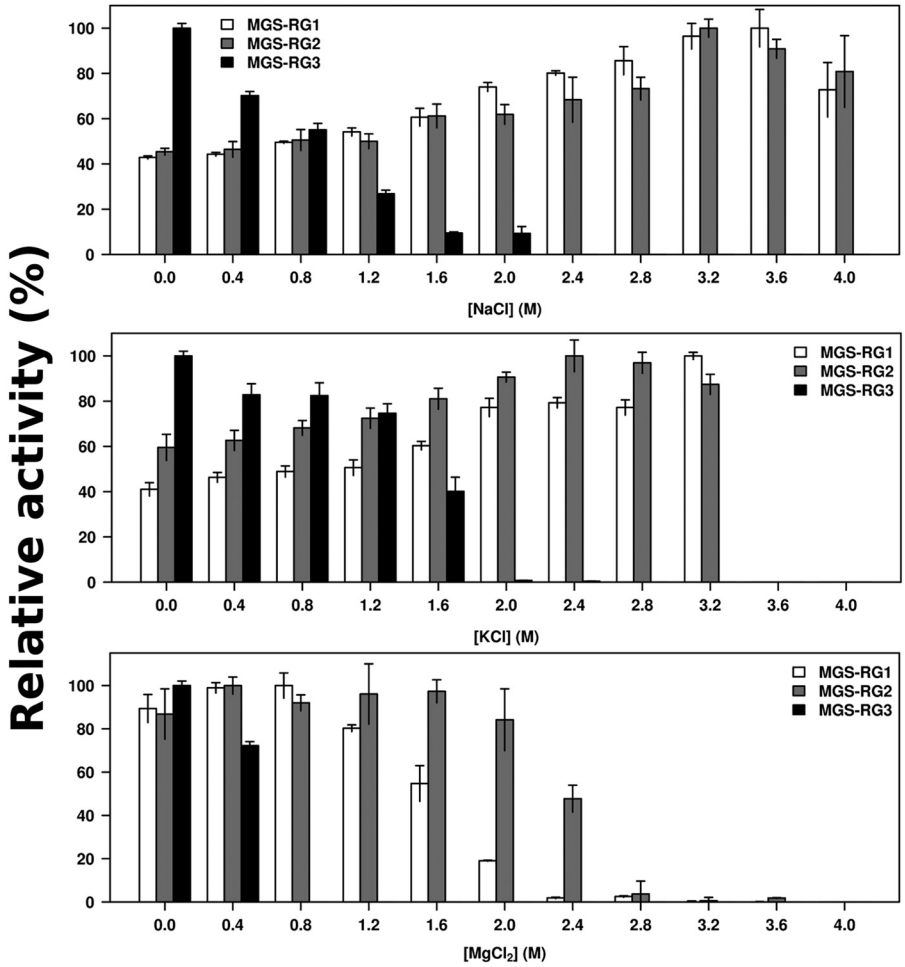


FIG 3 Activity profiles displaying the activities of carboxyl esterases at various concentrations of NaCl, KCl, and $MgCl_2$. The data represent the relative percentages of specific activity (U mg^{-1}) compared with the maximum activity (100%). The specific activities were calculated using 2 μg of protein and 1 mM *p*NP-propionate as assay substrate at pH 8.0 (50 mM Tris-HCl), and the assays were conducted at the optimal temperature for each enzyme (MGS-RG1, 45°C; MGS-RG2, 50°C; MGS-RG3, 30°C) as described in Materials and Methods. The SDs of the results of assays conducted in triplicate are shown. The activity of each enzyme in the absence of salt is given in the legend to Fig. 2.

sponding to the site pressure at 2,320 m (19), both enzymes were activated (≥1.4-fold), and at 35,000 kPa, they retain ≥97.7% of the activity they display at atmospheric pressure. In contrast, under our assay conditions, MGS-RG3 was inhibited by high pressure (77.9% and 57.2% activity at 23,000 and 35,000 kPa, respectively). Under similar conditions of pH and temperature for assays, commercial esterase/lipase preparations, such as CalA, CalB, Novozym 388 L, and PLE, retained at pressure above 23,000 kPa ≤8% of the activity they display at atmospheric pressure.

Taken together, we found that the two barotolerant and moderately halophilic proteins (MGS-RG1 and MGS-RG2) exhibited the highest optimal temperatures (45 to 50°C), whereas the nonbarophilic and nonhalophilic MGS-RG3 protein displayed the lowest optimal temperature (30°C).

Substrate fingerprints: activity against rare, chemically distinct esters. A total of 131 ester-like chemicals were used to evaluate the substrate ranges and specific activities (U mg^{-1}) of the three enzymes at pH 8.0 and at optimal temperature. These chemicals included 11 model esters (7 *p*NP esters and 4 triacylglycerols) (Fig. 4), as well as a battery of 120 structurally different esters (3, 6, 27, 28, 37, 53) (Fig. 5). Substrate fingerprints revealed that MGS-RG3 (17 positive substrates) exhibited a relatively narrow substrate range compared to those of MGS-RG2 and MGS-RG1, which were active against a larger number of substrates (32 and 37 positive substrates, respectively).

All three ester hydrolases from the α/β-hydrolase family preferred short-to-medium-chain-length *p*NP-esters and triacylglycerols (Fig. 4) and alkyl, alkenyl, and/or aryl esters but with different orders of preference (Fig. 5). Based on determinations of the specific activity for each enzyme, methyl benzoate (for MGS-RG1; 25.23 ± 0.95 U mg^{-1}), methyl 2-bromopropionate (for MGS-RG2; 40.7 ± 0.3 U mg^{-1}), and *p*NP-butyrate (for MGS-RG3; 356.3 ± 1.2 U mg^{-1}) were the preferred substrates for these enzymes. The activity ratio between the most and least active enzymes was approximately 1:14.

As shown in Fig. 5, among esters other than common *p*NP-esters and triacylglycerols, only phenyl acetate was hydrolyzed by all three enzymes, suggesting that these enzymes exhibited a high substrate profile divergence. MGS-RG1 and MGS-RG2 were the most similar, in that they were able to hydrolyze 14 common sub-

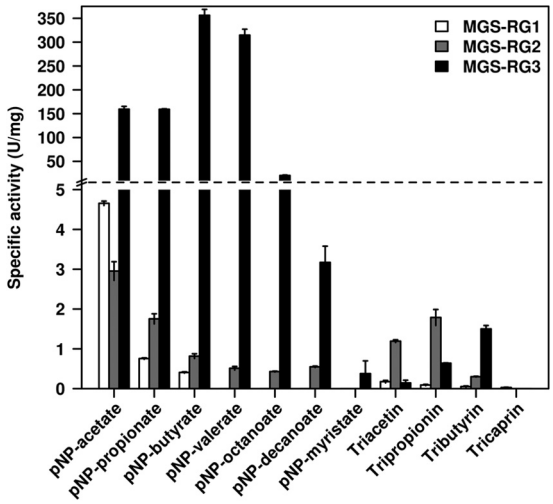


FIG 4 Substrate profiles of the activities of carboxyl esterases from microorganisms inhabiting the gill chamber of *R. exoculata* against model *p*NP-esters and triacylglycerols. The specific activity of each enzyme (in U mg^{-1}) against a set of structurally diverse substrates was measured using 2 μg protein at pH 8.0 in 50 mM Tris-HCl buffer for *p*NP-esters and 5 mM EPPS buffer for triacylglycerols, and the assays were conducted at the optimal temperature for each enzyme (MGS-RG1, 45°C; MGS-RG2, 50°C; MGS-RG3, 30°C) as described in Materials and Methods. The SDs of the results of assays conducted in triplicate are shown.

strates, albeit to different extents; this finding is consistent with their high sequence homology (63.7%) and similar taxonomic origins. However, 13 of the tested esters, including 8 halogenated esters (ethyl chloroacetate, ethyl fluoroacetate, ethyl-2-bromopropionate, methyl-2-chloropropionate, methyl-2-chloro-3-hydroxypropionate, ethyl-4-bromobutyrate, ethyl-3-bromopropionate, and ethyl-α-isobromobutyrate), methyl glycolate, methyl butyrate, ethyl-*trans*-cinnamate, butyl acetate, and methyl benzoate, were hydrolyzed only by MGS-RG1. Nine esters were hydrolyzed only by MGS-RG2; they included three halogenated esters [methyl 2-bromopropionate, methyl bromoacetate, and methyl (±)-α-bromophenylacetate], caproic acid methyl ester, methyl-(*R*)-lactate, methyl-(*S*)-lactate, γ-butyrolactone, α-D-glucose pentaacetate, and tri-*O*-acetyl-(D)-glucal. Thus, despite the high sequence homology of these enzymes, distinct substrate preferences were observed; these preferences may be related to distinct architectures and/or to the accessibility of various substrates to the active sites of the enzymes. Finally, MGS-RG3, which shared only 5 common substrates with MGS-RG1 and 3 common substrates with MGS-RG2, was the only enzyme of the 3 able to hydrolyze 4 esters, including vinyl benzoate, (*R*)-menthyl acetate, ethyl 4-bromobutyrate, and methyl chloroacetate.

The ability to hydrolyze halogenated alkyl and aryl esters (including those containing bromide, chloride, fluoride, and iodide) was demonstrated for MGS-RG1 (16 esters), MGS-RG2 (9 esters), and MGS-RG3 (4 esters). Based on specific activity determinations (U mg^{-1}) and using optimal substrates for each enzyme (Fig. 5), MGS-RG2 was found to be the most active carboxyl esterase for haloesters (40.7 ± 0.3 U mg^{-1} for methyl 2-bromopropionate), followed by MGS-RG1 (7.55 ± 0.06 U mg^{-1} for ethyl-α-isobromobutyrate) and, to a far lesser extent, MGS-RG3 (1.66 ± 0.01 U mg^{-1} for ethyl iodoacetate). The activity ratio between the most active and least active enzymes for the hydrolysis of haloesters was approximately 1:25. Notably, whereas MGS-RG2 only degraded bromide-containing esters, MGS-RG1 and MGS-RG3 acted on iodine, chloride, and bromide esters (Fig. 5). Finally, it is notable that methyl-2-bromo-2-butenoate, which was used as a model alkenyl haloester, was the only such substrate for MGS-RG2 (9.76 ± 1.08 U mg^{-1}), followed to a far lesser extent by MGS-RG1 (0.162 ± 0.03 U mg^{-1}).

MGS-RG2 utilized tri-O-acetyl-glucal (0.286 ± 0.04 U mg^{-1}) and the carbohydrate ester α-D-glucose pentaacetate (0.373 ± 0.011 U mg^{-1}) as substrates, while the hydroxycinnamic-like ester ethyl-*trans*-cinnamate was a substrate for MGS-RG1 (4.57 ± 0.06 U mg^{-1}) (Fig. 5). The hydrolysis of such compounds is commonly associated with polysaccharide (including mucus) degradation (54–56).

Finally, under our assay conditions, all three carboxyl esterases

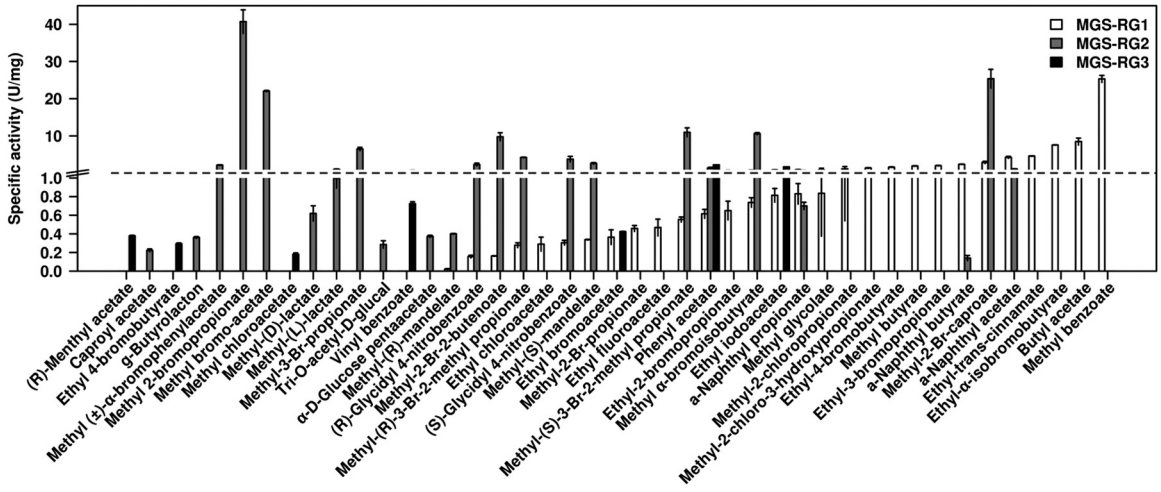


FIG 5 Substrate profiles of the activities of carboxyl esterases from microorganisms inhabiting the gill chamber of *R. exoculata* compared with those of a set of structurally distinct esters. The specific activities (U mg^{-1}) of each enzyme were measured against a set of structurally diverse substrates using 2 μg of protein at pH 8.0 in 5 mM EPPS buffer, and the assays were conducted at the optimal temperature for each enzyme (MGS-RG1, 45°C; MGS-RG2, 50°C; MGS-RG3, 30°C) as described in Materials and Methods. The SDs of the results of assays conducted in triplicate are shown.

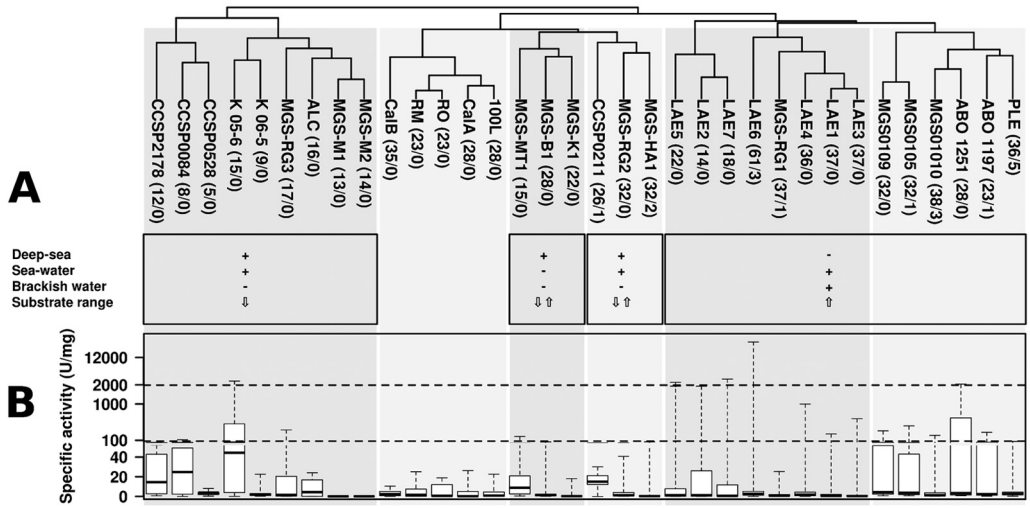


FIG 6 Clustering and specific activities (U mg^{-1}) of a set of 131 ester substrates of carboxyl esterases isolated from the gill chamber-associated microbiota from the deep-sea shrimp *R. exoculata* and from commercial and previously reported preparations. The source of chemicals are as reported previously (3, 6, 27, 28, 37, 53). Pearson's correlation was used to calculate the distances. (A) Hierarchical clustering was based on a binomial distribution of the presence or absence of activity for a portfolio of 131 substrates. The type of habitat (deep sea, superficial seawater, or brackish water) and the estimation of enzyme versatility, as estimated by the number of substrates (substrate range, low [⇓], medium [⇓⇑], or high [⇑]) upon which a given enzyme acts are shown for the enzymes in each cluster. (B) Box plots of the specific activity (U mg^{-1}) for the set of enzymes and esters shown in panel A. (A complete list of the esters tested is given in the materials and methods section of the supplemental material.) The activity values for enzymes other than those investigated in this study are reported elsewhere (3, 6, 27, 28, 37, 53). The values for MGS-RG1, MGS-RG2, and MGS-RG3 are shown in Fig. 4 and 5.

also were found to be enantioselective to different degrees for 5 chiral esters, which were found to serve as substrates for these enzymes (Fig. 5). Whereas methyl-(±)-mandelate, (±)-glycidyl-4-nitrobenzoate and (±)-methyl-3-bromo-2-methyl propionate were hydrolyzed by MGS-RG1 and MGS-RG2, methyl-(±)-lactate was a substrate for MGS-RG2 only, and menthyl-(±)-acetate was a substrate for MGS-RG3 only. The specific activities of the three enzymes for chiral esters ranged from 0.021 ± 0.006 to 10.97 ± 1.18 U mg^{-1}, and methyl-(*S*)-3-bromo-2-methyl propionate was the preferred chiral substrate. Based on a calculation of the apparent enantiomeric ratios for separate enantiomers (6), the enantiomeric ratios and substrate preferences are tentatively the following: (i) 158.4 ± 11.0 [preference for methyl-(*S*)-mandelate], 19.6 ± 0.8 [for (*S*)-glycidyl-4-nitrobenzoate], and 20.0 ± 0.7 [for (*S*)-methyl-3-bromo-2-methyl propionate] for MGS-RG1; (ii) 16.1 ± 1.3 [for methyl-(*S*)-lactate], 16.3 ± 1.1 [for (*S*)-glycidyl-4-nitrobenzoate], 26.1 ± 2.2 [for (*S*)-methyl-3-bromo-2-methyl propionate], and 65.4 ± 7.9 [for methyl-(*S*)-mandelate] for MGS-RG2; and (iii) 300 ± 9.3 for MGS-RG3 and menthyl-(*R*)-acetate.

Substrate fingerprints: comparison with known enzymes. The substrate profiles of the enzymes reported in this study against a set of 131 ester substrates (see the section on materials and methods in the supplemental material) were compared to a set of well-characterized commercial preparations with high substrate versatility (including CalA, CalB, Novozym 388 L, and PLE) and a set of carboxyl esterases obtained via genomic mining of marine bacteria (ABO_1197, ABO_1251, CCSP0084, CCSP0211, CCSP0528, and CCSP2178) or via metagenomic approaches of communities from marine (27, 28) and brackish sediment samples (6).

A cluster analysis (Fig. 6A), which was generated from a binomial distribution based on the presence or absence of activity at 30°C and at pH 8.0 against the set of 131 different esters used in this study and reported previously (3, 6, 27, 28, 37, 53), revealed that the three characterized carboxyl esterases clustered separately, consistent with the finding that they possess different substrate spectra (Fig. 4 and 5). Notably, the positioning of the enzymes based on their substrate profiles did not correlate with the scattered positioning based on sequence analysis (see Fig. S2 in the supplemental material), suggesting that the phylogenetic positions of these enzymes are independent of their biochemical profiles.

As shown in Fig. 6A, a number of observations can be made from the data. First, MGS-RG1 clustered more closely with a group of enzymes from marine (28) and brackish lake (6) samples that have been reported to be cold adapted and that show a broad substrate range. Interestingly, MGS-RG2 was functionally closer and possesses a substrate spectrum similar to that of a group of enzymes with a substrate portfolio of medium size; this group includes a recently reported carboxyl esterase isolated from a 100-m-depth seawater sample obtained from a hydrothermal vent at Saint Paul Island (3). Finally, MGS-RG3 is located within a group of enzymes with a highly restricted substrate range; this group includes two enzymes from deep-sea saline lakes in the Mediterranean Sea (3) and enzymes from oil-contaminated seawater samples (27). Of the enzymes tested, MGS-RG1 was the only enzyme able to hydrolyze methyl-2-Br-propionate. In addition, MGS-RG1 and MGS-RG3 showed various capacities to hydrolyze (*R*,*S*)-glycidyl 4-nitrobenzoate and methyl-2-bromo-2-butenoate, esters that otherwise were hydrolyzed only by MGS-HA1, an enzyme isolated from a hydrothermal vent (3).

DISCUSSION

Hydrothermal vents in the Mid-Atlantic Ridge hosting the deep-sea hydrothermal vent shrimp *R. exoculata* (19) lie 2,320 m below sea level, exhibit a salinity slightly lower than that of seawater

(~23.03 versus ~35.5 g kg^{-1}), and are characterized by high pressure (~23,000 kPa) and by a moderately low-to-warm temperature gradient that consistently ranges from 3 to 25°C (19). Using a metagenomic approach, the DNA of microbial communities inhabiting the gill chamber of *R. exoculata* was harvested and cloned to establish the fosmid library, which was used to screen for enzymatic activities of interest. This screening led to the discovery of three new carboxyl esterases, which were biochemically characterized after expression in a surrogate microbial host (*E. coli*). The results of this study demonstrate that microbes associated with the gill chamber of *R. exoculata* contain biochemically heterogeneous carboxyl esterases belonging to the α/β-hydrolase family; the properties of the newly discovered esterases correlate with their differing bacterial origins and most likely with different adaptations of their hosts to the prevailing environmental conditions.

The taxonomic distribution of top protein hits, as well as genome linguistics analysis, suggested that the metagenomic fragments containing these enzymes belong to *Alphaproteobacteria* (*Rhodobacterales*) and *Gammaproteobacteria* (unknown species or a genus related to *Thiothrix* and/or *Leucothrix*) from the *R. exoculata* gill chamber. This finding is consistent with the fact that *Alphaproteobacteria* and *Thiothrix/Leucothrix* (the closest cultured relative of which is the sulfur-oxidizing bacterium *L. mucor* [57]) accounted for approximately 4% and 7.5%, respectively, of the total *R. exoculata* gill chamber microbial population, as reported earlier, and that they were absent from the habitat seawater (19). Notably, *Thiothrix/Leucothrix*-related bacteria are of interest not only because of their location in a presently genomically uncharacterized region of the tree of life (51) but also because no functional data with respect to their metabolic capacities or their enzyme arsenal have been reported in the specialized literature.

The activity levels of the characterized enzymes (maximum for best substrates of 25.2 to 356 U mg^{-1}, depending on the carboxyl esterase) are in the range of other reported enzymes with esterase and lipase activity from macroorganism-associated microbiota. The only such enzymes previously reported are associated with marine sponges and have maximum activities of 2.8 to 2,700 U mg^{-1} (15, 16). The activity levels of the esterases reported here against ester substrates under conditions similar to those used in this study were similar to or somewhat lower than those of commercial preparations, as well as to those of previously reported enzymes isolated from marine seawater and from brackish water (average of up to 1,227 U mg^{-1}) (Fig. 6B). Their activities also were comparable to the activities of previously characterized carboxyl esterases from Mediterranean deep-sea locations (average, 54.5 U mg^{-1} [3]) and other marine deep-sea sediment (~1.70 to 560 U mg^{-1} [58–60]). The specific activities of these enzymes also were in the range of that observed for most similar proteins (34 to 39% sequence identity) that have been characterized (150 to 2,500 U mg^{-1} [61, 62]).

Further, the biochemical properties of the novel carboxyl esterases reported in this study revealed that although all three enzymes showed an activity-stability trade-off characteristic of cold-adapted enzymes, MGS-RG1 and MGS-RG2 exhibit higher temperature windows and greater thermostability than MGS-RG3. This finding is in agreement with the stability trade-off found via determination of the aggregation/denaturation temperatures (T_d) of these enzymes. Thus, the T_ds for MGS-RG1 (47.6°C) and MGS-RG2 (53.3°C) are higher than the T_d for MGS-RG3 (42.0°C). The T_d for the first subgroup of enzymes is comparable to that of the OLEI01171 carboxyl esterase from the cold-adapted marine bacterium *Oleispira antarctica* RB-8 (45.7°C), as well as to those of previously described esterases from superficial seawater (45.7 to 48.1°C [28]) and deep-sea (40.3 to 71.4°C [3]) (Table 2) and mesophilic bacteria (51 to 60°C [63]), whereas that for MGS-RG3 is slightly lower. The higher (for MGS-RG2) and lower (for MGS-RG3) T_ds are consistent with the fact that MGS-RG2 was the most active of the three enzymes at 55°C (70% of maximal activity), and that MGS-RG3 was the most active at 4°C (40% of maximal activity). The optimal temperatures for the activities of MGS-RG1 (45°C) and MGS-RG2 (50°C) are within the range of optimal temperatures of most similar characterized enzymes, including the top hit, EstB, from *Pseudomonas fluorescens* (34 to 37% amino acid sequence identity; UniProtKB/Swiss-Prot code Q51758.1), which has an optimal temperature for activity of 45°C (61). A similar observation was made for MGS-RG3; in that case, the most similar previously characterized hydrolase, MLH (39% amino acid sequence identity), showed an optimal temperature of 30°C and similar residual activity at low temperature (50% of the maximum activity at 10°C; UniProtKB/Swiss-Prot code Q9EX73.1 [62]). Differences in optimal temperature of at least 10 to 20°C between the two groups of enzymes investigated in this study also have been reported for several carboxyl esterases from the marine bacteria *Alcanivorax borkumensis* SK2 (28) and *Oleispira antarctica* RB-8 (64), as well as from uncultured bacteria from seawater samples (28), suggesting that such heterogeneous profiles are common in cold or moderate-temperature marine habitats.

Similarly, while the MGS-RG1 and MGS-RG2 enzymes exhibited similar hyperactivation at high salt concentrations and similar resistance to high pressure, MGS-RG3 exhibited low remaining activity in the presence of salt and pressure. Because salt and temperature dependence have not been reported for most similar enzymes, a direct comparison could not be done. However, examination of the effect of salt concentration on the activities of metagenomic esterases (3, 28) revealed that heterogeneous effects also are common among enzymes found in marine habitats, including deep-sea habitats (Table 2). Notably, the level of activation of both MGS-RG1 and MGS-RG2, which was as high as 2.4-fold at 3.2 to 3.6 M salt, and the significant pressure tolerance of these enzymes (>97% at 35,000 kPa) are similar to the salt and pressure dependence reported for other deep-sea enzymes (3), with similar optimal temperatures for activity. This is an important factor, because it was recently found that the development of enzymes (i.e., MGS-M2 [Table 2]) with higher optimal temperatures for activity by organisms in deep-sea environments can be linked to pressure and salt resistance, implicating yet-unknown mechanisms in this type of adaptation (3). In agreement with this, MGS-RG3, which exhibits low resistance to salt and pressure, was cold adapted, whereas the halophilic and barotolerant MGS-RG1 and MGS-RG2 enzymes were more active at moderate temperatures.

The different sensitivities of the three enzymes to environmental constraints, namely, temperature, pressure, and salinity, also were reflected at the level of reaction specificities and reactivities, with MGS-RG1 and MGS-RG2 being at least 14-fold less active than MGS-RG3 yet capable of accepting a broader range of ester substrates. Based on these findings, we speculate that the existence of two distinct subgroups of these enzymes based on their biochemical properties is a direct consequence of the different bacterial origins of the enzymes, as well as of the different adaptation

capacities of the corresponding bacterial hosts to the prevailing environmental conditions (pressure, salinity, and temperature). In addition, we infer that such differences also become marked at the highest level of functional hierarchy, that is, in relation to the substrate spectra of the enzymes, differences that are likely to support the different abilities of the corresponding bacterial hosts to metabolize nutrients. In agreement with this, we found that, in contrast to MGS-RG3, MGS-RG1 and MGS-RG2 are able to hydrolyze triacylglycerols; (non)halogenated alkyl, alkenyl, and aryl esters; lactones; chiral epoxides; and/or cinnamoyl and carbohydrate esters to similar extents. Such broad promiscuity is rare among enzymes with esterase and lipase activity (6, 28). In addition, both MGS-RG1 and MGS-RG2 can support polysaccharide degradation, an enzymatic activity that is critical for the ecological success of microorganisms in deep-sea environments, as well as for the utilization by these microorganisms of nutrients produced or liberated by the host in the gill chamber (50, 56). The polysaccharide degradation capability of these enzymes was demonstrated by their ability to degrade tri-O-acetyl-glucal, α-D-glucose pentaacetate, and ethyl-*trans*-cinnamate.

Finally, it should be emphasized that *R. exoculata* has been reported to move from one vent to another; thus, the microbes inhabiting the gill chamber of this species might contain versatile enzymes capable of adaptation to environmental and nutrient-availability gradients. This hypothesis is supported by the results presented here, which demonstrate the heterogeneous salt, pressure, and temperature dependence and stability tradeoff of the three enzymes investigated. As shown in Fig. 2 and 3 and Table 2, all three enzymes, albeit to different extents, showed enzymatic activity under the relevant conditions in the hydrothermal vent (3 to 25°C; 23.03 g kg^{-1} salt [or 0.42 M NaCl]; pressure, 230 atm at depth of 2,320 m). In addition, the heterogeneous substrate profiles of the three enzymes also agree with previous considerations. Further experimental evidence on this issue with a larger set of enzymes will be required to determine whether heterogeneity is common in other marine habitats. Related to this point, Fig. 6A of our work clearly shows that 3 of 4 enzymes from the polyaromatic-degrading marine bacterium *Cycloclasticus* sp. strain ME7 (CCSP codes), 2 from seawater at Kolguev Island (K codes), 2 from deep-sea Medee Lake (MGS-M codes), 3 from other deep-sea habitats (MGS-B/K/MT codes), 7 from a karst lake (LAE codes), and a set of 5 from oil-contaminated marine seawater samples and bacteria isolated from those samples (ABO codes and MGS0105, MGS0109, and MGS1010) formed separated homogeneous clusters in their reactivity profiles. Accordingly, the fact that other enzymes from different locations were ranked functionally closer to each other compared to enzymes from other locations while those from *R. exoculata* were scattered within the reactivity-based tree may support the existence of a wide functional heterogeneity in the gill chamber of *R. exoculata*. Note that the substrate profiles of the enzymes reported in this study also differ significantly from those of the most similar (34 to 39% sequence identity) characterized enzymes (61, 62); thus, structural factors may account for such differences.

Using a library of structurally diverse esters, we further demonstrated that the enzymes described in this study are characterized by high hydrolytic rates and broad substrate spectra together with enantioselectivity, all of which are desired features for biocatalysis. As an example, an enantiomeric ratio greater than 20 typically is associated with hydrolases with high enantioselectivity (6). Notably, the high and broad activity of these enzymes toward halogenated esters also may be of use in the optical enrichment of starting haloesters (65). We also found that while MGS-RG1 accepts methyl benzoate as the substrate, MGS-RG3 accepts only vinyl benzoate; therefore, we speculate that MGS-RG1 and MGS-RG3 can be used for selective transesterification procedures that rely upon irreversible benzoylation with methyl or vinyl esters (66).

The enzyme characteristics and versatile reactivity of the enzymes reported in this study, from a biochemically unexploited marine habitat and from a neglected taxonomic group of bacteria associated with macroorganisms inhabiting deep-sea hydrothermal vents, must be further evaluated and exploited on a larger scale.

Finally, one question that might rise after examining the phylogenetic positioning of sequences (see Fig. S2 in the supplemental material) and the biochemical clustering (Fig. 6) of all deep-sea enzymes, including the ones reported here, is whether the evolution of these model enzymes suggests any particular constraint in the deep-sea environment. The limited number of esterases (about 8) from deep-sea environments identified and characterized to date (3) does not allow us to reach a clear-cut conclusion. However, the distinct placement of some of the deep-sea enzymes at the level of substrate profiles (i.e., MGS-MT1, B1, K1, M1, and M2) (Fig. 6) and at the level of sequence (see Fig. S2) may point at the drivers of the evolution of enzymes in the deep sea. However, before drawing any conclusions, a further experimental analysis with a larger set of deep-sea enzymes is needed.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support provided by the European Community project MAMBA (FP7-KBBE-2008-226977). This work was further funded by grant BIO2011-25012 from the Spanish Ministry of the Economy and Competitiveness (formerly MICINN). P.N.G. and O.V.G. were supported by EU FP7 project MICROB3 (FP7-OCEAN.2011 287589). This work received support from the Government of Canada through Genome Canada and the Ontario Genomics Institute (grant 2009-OGI-ABC-1405 to A.F.Y. and A.S.) and from the U.S. National Institutes of Health (grants GM074942 and GM094585 to A.S. through the Midwest Center for Structural Genomics).

We thank José M. Andreu and María A. Olivia (CIB-CSIC) for their excellent support in relation to the circular dichroism experiments and analyses.

We have no competing interests to declare.

REFERENCES

1. **Singh BK.** 2010. Exploring microbial diversity for biotechnology: the way forward. Trends Biotechnol **28:**111–116. http://dx.doi.org/10.1016/j.tibtech.2009.11.006.
2. **Chistoserdova L.** 2014. Is metagenomics resolving identification of functions in microbial communities? Microb Biotechnol **7:**1–4. http://dx.doi.org/10.1111/1751-7915.12077.
3. **Alcaide M, Stogios PJ, Lafraya A, Tchigvintsev A, Flick R, Bargiela R, Chernikova TN, Reva ON, Hai T, Leggewie CC, Katzke N, La Cono V, Matesanz R, Jebbar M, Jaeger KE, Yakimov MM, Yakunin AF, Golyshin PN, Golyshina OV, Savchenko A, Ferrer M.** 20 October 2014. Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. Environ Microbiol http://dx.doi.org/10.1111/1462-2920.12660.
4. **Guazzaroni ME, Silva-Rocha R, Ward RJ.** 13 August 2014. Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. Microb Biotechnol http://dx.doi.org/10.1111/1751-7915.12146.
5. **Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH.** 2012. Are uncultivated bacteria really uncultivable? Microbes Environ **27:**356–366. http://dx.doi.org/10.1264/jsme2.ME12092.

6. **Martínez-Martínez M, Alcaide M, Tchigvintsev A, Reva O, Polaina J, Bargiela R, Guazzaroni ME, Chicote A, Canet A, Valero F, Rico Eguizabal E, Guerrero Mdel C, Yakunin AF, Ferrer M.** 2013. Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. Appl Environ Microbiol **79:**3553–3562. http://dx.doi.org/10.1128/AEM.00240-13.
7. **Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R.** 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol **12:**635–645. http://dx.doi.org/10.1038/nrmicro3330.
8. **Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J, Colwell RR, Danchin A, Dawyndt P, Dedeurwaerdere T, DeLong EF, Detter JC, De Vos P, Donohue TJ, Dong XZ, Ehrlich DS, Fraser C, Gibbs R, Gilbert J, Gilna P, Glöckner FO, Jansson JK, Keasling JD, Knight R, Labeda D, Lapidus A, Lee JS, Li WJ, Ma J, Markowitz V, Moore ER, Morrison M, Meyer F, Nelson KE, Ohkuma M, Ouzounis CA, Pace N, Parkhill J, Qin N, Rossello-Mora R, Sikorski J, Smith D, Sogin M, Stevens R, Stingl U, Suzuki K, et al.** 2014. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol **12:**e1001920. http://dx.doi.org/10.1371/journal.pbio.1001920.
9. **Jemli S, Ayadi-Zouari D, Hlima HB, Bejar S.** 2014. Biocatalysts: application and engineering for industrial purposes. Crit Rev Biotechnol **6:**1–13.
10. **Timmis K, de Lorenzo V, Verstraete W, Garcia JL, Ramos JL, Santos H, Economidis I, Nogales B, Timmis JK, Fonseca C, Pruzzo C, Karagouni A, Panopoulos N, Dixon B.** 2014. Pipelines for new chemicals: a strategy to create new value chains and stimulate innovation-based economic revival in southern European countries. Environ Microbiol **16:**9–18. http://dx.doi.org/10.1111/1462-2920.12337.
11. **Harrison JP, Gheeraert N, Tsigelnitskiy D, Cockell CS.** 2013. The limits for life under multiple extremes. Trends Microbiol **21:**204–212. http://dx.doi.org/10.1016/j.tim.2013.01.006.
12. **Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM.** 2008. Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. Ecology **89:**S91–S107. http://dx.doi.org/10.1890/07-1162.1.
13. **Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, Pelletier E, Fink D, Borowski C, Bach W, Dubilier N.** 2011. Hydrogen is an energy source for hydrothermal vent symbioses. Nature **476:**176–180. http://dx.doi.org/10.1038/nature10325.
14. **Karpushova A, Brümmer F, Barth S, Lange S, Schmid RD.** 2005. Cloning, recombinant expression and biochemical characterisation of novel esterases from *Bacillus* sp. associated with the marine sponge *Aplysina aerophoba*. Appl Microbiol Biotechnol **67:**59–69. http://dx.doi.org/10.1007/s00253-004-1780-6.
15. **Okamura Y, Kimura T, Yokouchi H, Meneses-Osorio M, Katoh M, Matsunaga T, Takeyama H.** 2010. Isolation and characterization of a GDSL esterase from the metagenome of a marine sponge-associated bacteria. Mar Biotechnol **12:**395–402. http://dx.doi.org/10.1007/s10126-009-9226-x.
16. **Selvin J, Kennedy J, Lejon DP, Kiran GS, Dobson AD.** 2012. Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge *Haliclona simulans*. Microb Cell Fact **11:**72. http://dx.doi.org/10.1186/1475-2859-11-72.
17. **Kim D, Kim SN, Baik KS, Park SC, Lim CH, Kim JO, Shin TS, Oh MJ, Seong CN.** 2011. Screening and characterization of a cellulase gene from the gut microflora of abalone using metagenomic library. J Microbiol **49:**141–145. http://dx.doi.org/10.1007/s12275-011-0205-3.
18. **Sim SJ, Baik KS, Park SC, Choe HN, Seong CN, Shin TS, Woo HC, Cho JY, Kim D.** 2012. Characterization of alginate lyase gene using a metagenomic library constructed from the gut microflora of abalone. J Ind Microbiol Biotechnol **39:**585–593. http://dx.doi.org/10.1007/s10295-011-1054-0.
19. **Jan C, Petersen JM, Werner J, Teeling H, Huang S, Glöckner FO, Golyshina OV, Dubilier N, Golyshin PN, Jebbar M, Cambon-Bonavita MA.** 2014. The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of Zetaproteobacteria. Environ Microbiol **16:**2723–2738. http://dx.doi.org/10.1111/1462-2920.12406.
20. **Charlaou JL, Donwal JP, Fouquet Y, Jean-Baptiste P, Jean-Luc C, Jean-Pierre D, Yves F, Jean-Baptiste P, Holm N.** 2002. Geochemistry of high H_2 and CH_4 vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14′N, MAR). Chem Geol **191:**345–359. http://dx.doi.org/10.1016/S0009-2541(02)00134-1.
21. **Zbinden M, Shillito B, Le Bris N, de Villardi de Montlaur C, Roussel E, Guyot F, Gaill F, Cambon-Bonavita M-A.** 2008. New insights on the metabolic diversity among the epibiotic microbial community of the hydrothermal shrimp *Rimicaris exoculata*. J Exp Mar Biol Ecol **359:**131–140. http://dx.doi.org/10.1016/j.jembe.2008.03.009.
22. **Hügler M, Petersen JM, Dubilier N, Imhoff JF, Sievert SM.** 2011. Pathways of carbon and energy metabolism of the epibiotic community associated with the deep-sea hydrothermal vent shrimp *Rimicaris exoculata*. PLoS One **6:**e16018. http://dx.doi.org/10.1371/journal.pone.0016018.
23. **Guri M, Durand L, Cueff-Gauchard V, Zbinden M, Crassous P, Shillito B, Cambon-Bonavita MA.** 2012. Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. ISME J **6:**597–609. http://dx.doi.org/10.1038/ismej.2011.133.
24. **Ponsard J, Cambon-Bonavita MA, Zbinden M, Lepoint G, Joassin A, Corbari L, Shillito B, Durand L, Cueff-Gauchard V, Compere P.** 2013. Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. ISME J **7:**96–109. http://dx.doi.org/10.1038/ismej.2012.87.
25. **Konn C, Testemale D, Querellou J, Holm NG, Charlou JL.** 2011. New insight into the contributions of thermogenic processes and biogenic sources to the generation of organic compounds in hydrothermal fluids. Geobiology **9:**79–93. http://dx.doi.org/10.1111/j.1472-4669.2010.00260.x.
26. **Wahl M, Goecke F, Labes A, Dobretsov S, Weinberger F.** 2012. The second skin: ecological role of epibiotic biofilms on marine organisms. Front Microbiol **3:**292. http://dx.doi.org/10.3389/fmicb.2012.00292.
27. **Alcaide M, Tornés J, Stogios PJ, Xu X, Gertler C, Di Leo R, Bargiela R, Lafraya A, Guazzaroni ME, López-Cortés N, Chernikova TN, Golyshina OV, Nechitaylo TY, Plumeier I, Pieper DH, Yakimov MM, Savchenko A, Golyshin PN, Ferrer M.** 2013. Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. Biochem J **454:**157–166. http://dx.doi.org/10.1042/BJ20130552.
28. **Tchigvintsev A, Tran H, Popovic A, Kovacic F, Brown G, Flick R, Hajighasemi M, Egorova O, Somody JC, Tchigvintsev D, Khusnutdinova A, Chernikova TN, Golyshina OV, Yakimov MM, Savchenko A, Golyshin PN, Jaeger KE, Yakunin AF.** 7 September 2014. The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. Appl Microbiol Biotechnol http://dx.doi.org/10.1007/s00253-014-6038-3.
29. **Turner NJ, Truppo MD.** 2013. Biocatalysis enters a new era. Curr Opin Chem Biol **17:**212–214. http://dx.doi.org/10.1016/j.cbpa.2013.02.026.
30. **Reyes-Duarte D, Ferrer M, García-Arellano H.** 2012. Functional-based screening methods for lipases esterases and phospholipases in metagenomic libraries. Methods Mol Biol **861:**101–113. http://dx.doi.org/10.1007/978-1-61779-600-5_6.
31. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26:**1107–1115. http://dx.doi.org/10.1093/nar/26.4.1107.
32. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
33. **Hall TA.** 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95–98.
34. **Hofmann K, Stoffel W.** 1993. TMbase—a database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler **374:**166.
35. **Pace N, Scholtz JM.** 1997. Measuring the conformational stability of a protein, p 299–322. *In* Creighton TE (ed), Protein structure: a practical approach. IRL Press at Oxford University Press, Oxford, United Kingdom.
36. **Schmid FZ.** 1997. Optical spectroscopy to characterize protein conformation, p 261–298. *In* Creighton TE (ed), Protein structure: a practical approach. IRL Press at Oxford University Press, Oxford, United Kingdom.
37. **Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T.** 2013. AntiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res **1:**W204–W212.

38. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
39. **Stothard P.** 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques **28:**1102–1104.
40. **Ménigaud S, Mallet L, Picord G, Churlaud C, Borrel A, Deschavanne P.** 2012. GOHTAM: a website for genomic origin of horizontal transfers, alignment and metagenomics. Bioinformatics **28:**1270–1271. http://dx.doi.org/10.1093/bioinformatics/bts118.
41. **Suzuki H, Saito R, Tomita M.** 2009. Measure of synonymous codon usage diversity among genes in bacteria. BMC Bioinformatics **10:**167. http://dx.doi.org/10.1186/1471-2105-10-167.
42. **Ganesan H, Rakitianskaia AS, Davenport CF, Tümmler B, Reva ON.** 2008. The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. BMC Bioinformatics **9:**333. http://dx.doi.org/10.1186/1471-2105-9-333.
43. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113. http://dx.doi.org/10.1186/1471-2105-5-113.
44. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725–2729. http://dx.doi.org/10.1093/molbev/mst197.
45. **Filippova EV, Weston LA, Kuhn ML, Geissler B, Gehring AM, Armoush N, Adkins CT, Minasov G, Dubrovska I, Shuvalova L, Winsor JR, Lavis LD, Satchell KJ, Becker DP, Anderson WF, Johnson RJ.** 2013. Large scale structural rearrangement of a serine hydrolase from *Francisella tularensis* facilitates catalysis. J Biol Chem **288:**10522–10535. http://dx.doi.org/10.1074/jbc.M112.446625.
46. **Arpigny JL, Jaeger KE.** 1999. Bacterial lipolytic enzymes: classification and properties. Biochem J **343:**177–183. http://dx.doi.org/10.1042/0264-6021:3430177.
47. **Hausmann S, Jaeger K-E.** 2010. Lipolytic enzymes from bacteria, p 1099–1126. *In* Timmis KN (ed), Handbook of hydrocarbon and lipid microbiology. Springer, Berlin, Germany.
48. **Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Böttcher D, Joosten HJ, Bornscheuer UT.** 2010. The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. Chembiochem **11:**1635–1643. http://dx.doi.org/10.1002/cbic.201000213.
49. **Howarth R, Unz RF, Seviour EM, Seviour RJ, Blackall LL, Pickup RW, Jones JG, Yaguchi J, Head IM.** 1999. Phylogenetic relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii* sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp. nov. Int J Syst Bacteriol **4:**1817–1827.
50. **Brock TD.** 2006. The genus *Leucothrix*. Prokaryotes **6:**931–938.
51. **Lapidus A, Nolan M, Lucas S, Glavina Del Rio T, Tice H, Cheng JF, Tapia R, Han C, Goodwin L, Pitluck S, Liolios K, Pagani I, Ivanova N, Huntemann M, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Brambilla EM, Rohde M, Abt B, Verbarg S, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Woyke T.** 2011. Genome sequence of the filamentous, gliding *Thiothrix nivea* neotype strain ($JP2^T$). Stand Genomic Sci **5:**398–406. http://dx.doi.org/10.4056/sigs.2344929.
52. **Qian Z, Lutz S.** 2005. Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. J Am Chem Soc **127:**13466–13467. http://dx.doi.org/10.1021/ja053932h.
53. **Martínez-Martínez M, Lores I, Peña-García C, Bargiela R, Reyes-Duarte D, Guazzaroni ME, Peláez AI, Sánchez J, Ferrer M.** 2014. Biochemical studies on a versatile esterase that is most catalytically active with polyaromatic esters. Microb Biotechnol **7:**184–191. http://dx.doi.org/10.1111/1751-7915.12107.
54. **Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni ME, Polaina J, Strittmatter AW, Reva O, Waliczek A, Yakimov MM, Golyshina OV, Ferrer M, Golyshin PN.** 2010. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl Environ Microbiol **76:**5934–5946. http://dx.doi.org/10.1128/AEM.00902-10.
55. **Vieites JM, Ghazi A, Beloqui A, Polaina J, Andreu JM, Golyshina OV, Nechitaylo TY, Waliczek A, Yakimov MM, Golyshin PN, Ferrer M.** 2010. Inter-conversion of catalytic abilities in a bifunctional carboxyl/feruloyl-esterase from earthworm gut metagenome. Microb Biotechnol **3:**48–58. http://dx.doi.org/10.1111/j.1751-7915.2009.00135.x.
56. **Werner J, Ferrer M, Michel G, Mann AJ, Huang S, Juarez S, Ciordia S, Albar JP, Alcaide M, La Cono V, Yakimov MM, Antunes A, Taborda M, da Costa MS, Hai T, Glöckner FO, Golyshina OV, Golyshin PN, Teeling H, Consortium MAMBA.** 2014. *Halorhabdus tiamatea*: proteogenomics and glycosidase activity measurements identify the first cultivated euryarchaeon from a deep-sea anoxic brine lake as potential polysaccharide degrader. Environ Microbiol **16:**2525–2537. http://dx.doi.org/10.1111/1462-2920.12393.
57. **Grabovich MY, Muntyan MS, Lebedeva VY, Ustiyan VS, Dubinina GA.** 1999. Lithoheterotrophic growth and electron transfer chain components of the filamentous gliding bacterium *Leucothrix mucor* DSM 2157 during oxidation of sulfur compounds. FEMS Microbiol Rev **178:**155–161. http://dx.doi.org/10.1111/j.1574-6968.1999.tb13772.x.
58. **Hårdeman F, Sjöling S.** 2007. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. FEMS Microbiol Ecol **59:**524–534. http://dx.doi.org/10.1111/j.1574-6941.2006.00206.x.
59. **Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ, Lee JH.** 2009. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. Appl Microbiol Biotechnol **81:**865–874. http://dx.doi.org/10.1007/s00253-008-1656-2.
60. **Jeon JH, Kim JT, Lee HS, Kim SJ, Kang SG, Choi SH, Lee JH.** 2011. Novel lipolytic enzymes identified from metagenomic library of deep-sea sediment. Evid Based Complement Alternat Med **2011:**271419. http://dx.doi.org/10.1155/2011/271419.
61. **Hong KH, Jang WH, Choi KD, Yoo OJ.** 1991. Characterization of *Pseudomonas fluorescens* carboxylesterase: cloning and expression of the esterase gene in *Escherichia coli*. Agric Biol Chem **55:**2839–2845. http://dx.doi.org/10.1271/bbb1961.55.2839.
62. **van der Vlugt-Bergmans CJ, van der Werf MJ.** 2001. Genetic and biochemical characterization of a novel monoterpene epsilon-lactone hydrolase from *Rhodococcus erythropolis* DCL14. Appl Environ Microbiol **67:**733–741. http://dx.doi.org/10.1128/AEM.67.2.733-741.2001.
63. **Lemak S, Tchigvintsev A, Petit P, Flick R, Singer AU, Brown G, Evdokimova E, Egorova O, Gonzalez CF, Chernikova TN, Yakimov MM, Kube M, Reinhardt R, Golyshin PN, Savchenko A, Yakunin AF.** 2012. Structure and activity of the cold-active and anion-activated carboxyl esterase OLEI01171 from the oil-degrading marine bacterium *Oleispira antarctica*. Biochem J **445:**193–203. http://dx.doi.org/10.1042/BJ20112113.
64. **Kube M, Chernikova TN, Al-Ramahi Y, Beloqui A, Lopez-Cortez N, Guazzaroni ME, Heipieper HJ, Klages S, Kotsyurbenko OR, Langer I, Nechitaylo TY, Lünsdorf H, Fernández M, Juárez S, Ciordia S, Singer A, Kagan O, Egorova O, Petit PA, Stogios P, Kim Y, Tchigvintsev A, Flick R, Denaro R, Genovese M, Albar JP, Reva ON, Martínez-Gomariz M, Tran H, Ferrer M, Savchenko A, Yakunin AF, Yakimov MM, Golyshina OV, Reinhardt R, Golyshin PN.** 2013. Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. Nat Commun **4:**2156. http://dx.doi.org/10.1038/ncomms3156.
65. **Azim A, Sharma SK, Olsen CE, Parmar VS.** 2001. Lipase catalysed synthesis of optically enriched α-haloamides. Bioorg Med Chem **9:**1345–1348. http://dx.doi.org/10.1016/S0968-0896(01)00006-2.
66. **Ciffreda P, Alessandrini L, Terraneo G, Santaniello E.** 2003. Lipase-catalyzed selective benzoylation of 1,2-diols with vinyl benzoate in organic solvents. Tetrahedron Asymmetr **14:**3197–3201. http://dx.doi.org/10.1016/j.tetasy.2003.07.003.

CAPÍTULO 6
DISCUSIÓN GENERAL

CAPÍTULO 6
DISCUSIÓN GENERAL

En el **Capítulo 6** se pretende aportar una discusión global de los trabajos y resultados más relevantes correspondientes a los Capítulos 2, 3, 4 y 5. Se demostrará como todos los capítulos han tenido un hilo conductor similar y que en su conjunto han aportado resultados novedosos a nivel individual y colectivo.

6.1 Introducción al conjunto de enzimas recogidas en esta Memoria

Hasta la fecha, los procesos de investigación para la generación de nuevas enzimas se han basado, principalmente, en: (i) la exploración de la capacidad de las enzimas ya conocidas para la resolución de nuevas reacciones o de moléculas conocidas bajo condiciones *a-la-carta*, que incluyen estudios de ingeniería del biocatalizador; (ii) la exploración de actividades enzimáticas mejoradas o nuevas en entornos naturales gracias a la aplicación de técnicas de alto rendimiento como la mutagénesis dirigida (que requiere una capacidad de cálculo notable) y el *high throughput screening; y* (iii) la identificación de nuevas actividades enzimáticas a partir del aislamiento de la flora microbiana existente en ambientes y entornos naturales biológicos distintos, incluyendo las áreas fuertemente contaminadas o extremos como entornos potencialmente generadores de nuevas formas de vida y por lo tanto de nuevas actividades enzimáticas. Las estrategias seguidas hasta la fecha en estas tres grandes líneas de generación de conocimiento, presentan ventajas e inconvenientes importantes, por lo que el uso conjunto de estas tres estrategias es deseable. Pese al desarrollo en estas tres áreas, no se ha conseguido ir más allá de la identificación eficaz de actividades enzimáticas ya existentes, que en la mayor parte de los casos suponen como mucho la identificación de pocas variantes enzimáticas que catalizan (con mejor o peor fortuna) reacciones previamente resueltas, y con estructuras conocidas.

En la presente Tesis Doctoral, se han empleado herramientas de metagenómica funcional, es decir, la generación de librerías de clones obtenidas a partir de DNA de comunidades microbianas que posteriormente son testeadas por actividades de interés, para la identificación de nuevas enzimas. La presente Tesis Doctoral no se centra en desarrollos metodológicos en lo que a la aplicación de técnicas metagenómicas se refiere, sino en la utilización de herramientas existentes en generar una amplia colección de enzimas.

La **Tabla 2** resume el listado de las enzimas estudiadas durante esta Tesis Doctoral y cuyas características y/ o estructuras han quedado recogidas en los Capítulos anteriores. En ella queda reflejado el lugar de procedencia de las mismas, así como su actividad enzimática. Como se muestra en la tabla, en particular, en la presente Tesis Doctoral, se han identificado y caracterizado 25 nuevas enzimas (22 esterasas/lipasas, 1 beta-glucosidasa, 1 aldo-ceto reductasa, 1 (L)-lactato dehidrogenasa) de metagenomas creados a partir de DNA de comunidades microbianas procedentes de 8 hábitat diferentes y un genoma de una bacteria cultivable. En particular de: i) cuatro fosas marinas (*Medee*, *Kryos*, *Bannock* y *Matapan-Vavilov*) del Este del Mar Mediterráneo (3 de ellas hipersalinas); ii) un lago cárstico (Lago Arreo); iii) el microbioma de agallas de una gamba (*Rimicaris exoculata*) que viene a 2,320 m de profundidad en la zona cercana a un fuente hidrotermal en la Dorsal Mesoatlántica; iv) agua marina superficial contaminada con crudo cercana a la Isla de Kolguev en el Mar de Barents; v) una fuente hidrotermal no profunda en la Isla de San Paul (Alaska); y vi) de la bacteria marina hidrocarbonoclástica *Cycloclasticus* sp. ME7.

La temperatura, salinidad y profundidad de los hábitats estudiados oscila entre 4-16.5ºC, entre 1.1 y 348 g/kg y entre 0 y 4,908 m, respectivamente. Las características geoquímicas de estos hábitats los convierten en ejemplos de ambientes poco explorados a nivel enzimático y la amplia diversidades de factores ambientales, en particular, salinidad, temperatura y presión, los conviertes en hábitats adecuados para estudios estructura-función, adaptaciones a medios extremos y análisis de promiscuidad catalítica.

Enzima	Localización	ACTIVIDAD	AA	MW	*p*I	Identidad (%)	ESTRUCTURA
LAE1	Lago Arreo	EST	272	29.214	6.41	31	NO
LAE2		EST	324	34.002	10.04	31	NO
LAE3		EST	300	31.614	4.78	36	NO
LAE4		EST	309	33.499	8.07	25	NO
LAE5		EST	217	24.328	6.95	54	NO
LAE6		EST	315	33.936	4.85	41	SI
LAE7		EST	310	32.981	9.43	32	NO
K_05-6	Isla Kolguev	EST:MCP-H	319	35.487	5.04	50	NO
K_06-5		EST:MCP-H	316	34.301	8.44	70	NO
CCSP0084	*Cycloclasticus* ME7	EST:MCP-H	282	31.547	5.36	38	SI
CCSP0211		EST:MCP-H	223	24.190	4.66	41	NO
CCSP0528		EST	287	32.498	5.78	56	NO
CCSP2178		EST:MCP-H	282	31.354	5.37	62	NO
MGS-M1	*Medee*	EST	239	26.801	5.77	49	SI
MGS-M2		EST	276	31.602	5.39	48	SI
MGS-M3		GLY	734	84.278	5.55	41	NO
MGS-M4		ACR	274	31.792	5.95	54	SI
MGS-M5		LDH	314	34.792	6.21	72	SI
MGS-B1	*Bannock*	EST	316	34.551	4.38	51	NO
MGS-K1	*Kryos*	EST	514	56.234	4.89	44	NO
MGS-MT1	*Matapan-Vavilov*	EST	323	35.651	7.29	62	SI
MGS-HA1	San Pablo	EST	261	28.862	5.99	99	NO
MGS-RG1	*R. exoculata*	EST	222	24.983	5.07	50	NO
MGS-RG2		EST	225	25.363	5.31	52	NO
MGS-RG3		EST	277	30.611	5.51	47	NO

Tabla 2 | Listado de enzimas identificadas y caracterizadas en esta Tesis Doctoral. Se recoge el nombre de la enzima, el lugar de procedencia, la actividad, el número de aminoácidos, el *p*I, la masa molecular, y el porcentaje de identidad frente a secuencias depositadas en las bases de datos. Abreviaturas: EST, esterasa; MCP-H, *meta-cleavage product hydrolase*; ACR, aldo-ceto reductase; LDH, lactato dehidrogenasa. La tabla resume también si la estructrura de la enzima ha sido resuelta y recogida en la presente Memoria.

6.2 Diferencias en la eficacia en los rastreos y homología de las enzimas

Una de las primeras cuestiones a comparar entre todas las librerías estudiadas es si el porcentaje de clones positivos es similar o no. La **Tabla 3** muestra los valores de índices positivos, es decir, cuantos clones tengo que rastrear para encontrar un clon positivo. En particular, clones con actividad esterasa/lipasa, usando sustratos modelo como acetato de naftilo y tributirina. Ambos sustratos se emplean de forma rutinaria en el rastreo de actividad hidrolasa (Reyes-Duarte *et al*., 2012). El tamaño medio de las librerías sometidas a rastreos funcionales oscila entre 120 y 816 millones de pares de bases (Mpb), y el número de positivos para las actividades de interés por genoteca oscila entre 1 clon positivo por cada 667 clones (1:667) a 1:15,000. Observando los valores mostrados en la **Tabla 3** podemos decir que el Lago Arreo y la fosa de *Matapan-Vavilov*, que son los sitios con menor salinidad, son los que muestran índices más favorables (1 cada 1,152 para el Lago Arreo y 1 cada 667 para *Matapan-Vavilov*), frente a las fosas marinas hipersalinas (desde 1 cada 2,624 para *Medee* a 1 cada 15,000 para *Bannock*). Esto sugiere que las características geoquímicas de los ambientes producen un sesgo importante en el mayor o menor éxito en los rastreos usando condiciones estándar que se emplean comúnmente en los ensayos enzimáticos (temperatura ambiente, baja salinidad (0.15 M NaCl), pH cercano a la neutralidad, y no presión. Esto concuerda con el hecho de que la incidencia de clones positivos es muy baja en el caso de la librería procedente de la fuente hidrotermal de la Isla de San Pablo, probablemente debido a que muchas de las enzimas de los microorganismos

presentes en esta muestra son termo-activas y muchas no mostraran actividad a la temperatura a la que se hacen los rastreos (temperatura ambiente); por este motivo, habrá que rastrear muchos clones para encontrar uno positivo (1 cada 11,000). En el caso del valor para la librería de la Isla de *Kolguev*, los valores no son del todo comparativos porque en este caso la genoteca es de fagos y no de fósmidos como para las otras muestras, y los valores de incidencia no son comparativos.

Fuente ambiental	Índice de positivos	Mpb rastreados	Salinidad (g/kg)	Temperatura (ºC)	Profundidad (m)
Lago Arreo	1:1,152	342	1,10	6.90	24
Isla de *Kolguev*	1:7,000	4,800	35.50	4.00	0
Medee	1:2,624	472	347.80	14.97	3,010
Bannock	1:15,000	450	263.20	14.50	3,342
Kryos	1:5,280	158.4	313.92	16.50	3,340
Matapan-Vavilov	1:667	120	38.60	14.29	4,908
Isla de San Pablo	1:11,000	330	35.50	65.00	100
R. exoculata	1:2,720	816	23.03	8.70	2,320

Tabla 3 | Eficacia en los rastreos de actividad esterasa/lipasa en las librerías de clones analizadas en la presente Memoria. Como valor de eficacia se indica la incidencia de clones positivos en las librerías de clones analizadas en la presente Tesis Doctoral; este valor corresponde al valor del número de clones totales a rastrear para encontrar un clon con actividad esterasa/lipasa, usando sustratos modelo como acetato de naftilo y tributirina.

Una segunda pregunta que surge de las enzimas identificadas es: ¿Cuál es grado de novedad a nivel de secuencia? Este es un punto importante ya que de forma recurrente se piensa que a través de las técnicas metagenómicas se puede acceder a enzimas novedosas (Fernández-Arrojo *et al*., 2010).

La masa molecular y el punto isoeléctrico (*p*I) de las enzimas estudiadas oscila entre 24,190 y 84,278 Da, y 4.66 y 10.04, respectivamente (**Tabla 2**). Tales rangos son comunes en las enzimas similares descritas en las bases de datos. Destaca como los *p*I suelen ser semejantes entre las diferentes enzimas de un mismo hábitat, excepto para las enzimas procedentes del Lago Arreo, cuyos valores difieren notablemente de unas a otras (oscilan entre 4.8-10.0). También ocurre lo mismo si comparamos el origen microbiano de las 25 enzimas estudiadas. Así, un análisis por frecuencia de tetra-nucleótidos y homología (Ménigaud *et al*., 2012) sugiere que las enzimas del Lago Arreo son las que presentan mayor heterogeneidad en lo que se refiere a la diversidad de bacterias productoras de dichas enzimas (**Tabla 4**). Esto ocurre a menor nivel en los otros hábitats analizados, donde se observa que enzimas procedentes de un mismo hábitat tienen un origen microbiano similar. Ambos factores (*p*I y origen microbiano) podrían estar correlacionados con el hecho de que las características bioquímicas de las enzimas del Lago Arreo difieren notablemente entre si, a pesar de proceder de un mismo hábitat (ver Capítulo 2).

A nivel de secuencia las enzimas identificadas y analizadas presentan valores de identidad, respecto a enzimas homólogas descritas en las bases de datos, entre el 25% y el 72% (**Tabla 2**). Solo una de las enzimas (MGS-HA1) presentaba una homología del 99%; ésta corresponde a la esterasa de un organismo cultivable (*Geobacter* sp.) aislado de una fuente hidrotermal no profunda en la Isla de San Pablo (Alaska). Esto sugiere que la mayoría de las enzimas identificadas y estudiadas a lo largo de esta Tesis Doctoral difieren considerablemente a nivel de secuencia de otros enzimas homólogas descritas en las bases de datos, aunque en algún caso, como es la enzima MGS-HA, se observa un grado de homología cercano al 99%.

La **Tabla 2** también nos muestra que los niveles más bajos de homología son los que muestran las esterasas del Lago Arreo (de 31 al 54%), comparándolos con las del Mar de Barents y *Cycloclasticus* (del 41 al 70%), las de las fosas marinas (del 41 al 72%), la de la Isla de San Pablo (99%) y las de *R. exoculata* (del 47 al 52%). Estos

datos nos demuestran que explorar ambientes extremos, raros o difíciles de muestrear, no asegura encontrar enzimas más diferentes a nivel de secuencia respecto a las que ya se conocen o respecto a ambientes menos extremos como es el Lago Arreo. De hecho, los datos descritos en el Capítulo 2, demuestran que las enzimas del Lago Arreo, que puede considerarse como un ambiente de fácil muestreo y no extremo (temperatura media que fluctúa entre los 4.7 y 19.8 ºC, y una salinidad de 1.10 g/kg), son las que presentan menos homología frente a las descritas en las bases de datos. Dicho esto, es importante mencionar que los menores o mayores niveles de homología no implican una bioquímica similar o diferente, ya que diferencias estructurales y pequeños cambios en el centro activo pueden ocasionar grandes cambios en la actividad y especificidad frente a distintos sustratos, independientemente de la homología que las enzimas muestren en sus secuencias. Este hecho ha quedado claramente demostrado en la presente Tesis Doctoral, como se discutirá más adelante.

Enzima	Origen microbiano
LAE1 y LAE2	Betaproteobacteria (*Cupriavidus/Ralstonia*)
LAE3	Alphaproteobacteria (*Methylobacterium*)
LAE4	Alphaproteobacteria (*Rhizobium*)
LAE5	Gamma/Alphaproteobacteria (*Azotobacter*)
LAE6	Alphaproteobacteria (*Sphingomonas*)
LAE7	Betaproteobacteria (*Burkholderia*)
K_05-6	Gammaproteobacteria
K_06-5	Betaproteobacteria
CCSP0084, CCSP0211, CCSP0528 y CCSP2178	Gammaproteobacteria (*Cycloclasticus*)
MGS-M1, MGS-M2, MGS-M3, MGS-M4, MGS-M5	Mollicutes (*Haloplasma*)
MGS-B1	Betaproteobacteria (*Variovorax*)
MGS-K1	Desconocido
MGS-MT1	Gammaproteobacteria (*Alteromonas*)
MGS-HA1	Bacilli (*Geobacillus*)
MGS-RG1, MGS-RG2	Gammaproteobacteria (*Thiothrix/Leucothrix*)
MGS-RG3	Alphaproteobacteria (*Rhodobacteraceae*)

Tabla 4 | Posible origen microbiano de las enzimas recogidas en la Memoria. La asignación taxonómica se basa en el empleo de herramientas de análisis basadas en la frecuencia de tetranucleótidos en las secuencias de los fósmidos que contienen la secuencia que codifica la enzima de interés y la homología con secuencias conocidas y depositadas en las bases de datos. Como se aprecia, la mayoría de las enzimas procede de bacterias del filo Proteobacteria (18), seguido de Tenericutes (5) y Firmicutes (1); en uno de los casos, no fue posible sugerir la posible bacteria de origen.

Dado que de entre todas las enzimas estudiadas las más representadas son las hidrolasas con actividad esterasa y/o lipasa (22 de las 25 totales), es posible también analizar los valores de homología entre ellas. En la **Tabla 5** se puede observar que el valor mínimo es del 0.5% y el máximo de 63.7%. Esta diferencia tan marcada sugiere que las enzimas estudiadas difieren considerablemente entre si a nivel secuencia y que, por tanto, durante esta Tesis Doctoral se han caracterizado enzimas con una amplia diversidad, hecho éste que puede tener implicación a nivel de las reactividades mostradas por las diferentes enzimas, como se discutirá más adelante.

6.3 Características físico-químicas de las enzimas estudiadas

La **Figura 17** resume de forma comparativa el efecto del pH, la temperatura, y la concentración de sal (NaCl) para todas las enzimas identificadas y caracterizadas. Se aprecian similitudes y diferencias significativas en lo que al efecto de estos tres factores en la actividad enzimática se refiere. El pH es un factor que afecta de forma homogénea a la actividad con valores máximos entre un pH de 7.0 y 8.5, independientemente del origen de la enzima. Estos valores son muy comunes en las enzimas similares descritas en la bibliografía y correlacionan con el pH del hábitat de procedencia (cercano a la neutralidad).

	LAE1	LAE2	LAE3	LAE4	LAE5	LAE6	LAE7	K_05-6	K_06-5	CCSP0084	CCSP0211	CCSP0528	CCSP2178	MGS-M1	MGS-M2	MGS-B1	MGS-K1	MGS-MT1	MGS-HA1	MGS-RG1	MGS-RG2	MGS-RG3
LAE1		45.6	5.7	16.2	2.0	14.0	18.4	12.2	14.9	15.9	15.1	11.6	16.0	12.4	11.3	17.2	11.9	18.9	17.1	17.9	17.2	17.6
LAE2	45.6		12.2	8.6	8.3	13.2	30.0	7.4	8.3	8.5	7.6	6.4	8.4	6.4	6.2	10.1	8.4	10.3	8.7	8.8	8.2	9.2
LAE3	5.7	12.2		21.3	14.7	25.9	14.1	15.2	18.6	17.1	19.0	17.6	17.6	19.1	15.4	23.5	13.9	23.4	15.8	14.3	14.9	33.1
LAE4	16.2	8.6	21.3		12.2	18.6	8.9	18.1	4.6	15.7	15.4	17.0	17.4	18.2	17.7	19.1	10.3	17.5	15.8	15.8	13.2	14.9
LAE5	2.0	8.3	14.7	12.2		10.3	16.0	2.0	3.4	14.5	8.9	18.2	17.1	3.6	16.1	9.7	6.6	10.4	14.6	8.5	8.5	12.2
LAE6	15.3	13.2	25.9	18.6	10.3		18.1	15.3	19.9	2.7	9.8	15.1	7.5	14.9	4.9	37.6	13.2	19.3	17.6	7.9	8.2	27.2
LAE7	18.4	30.0	14.0	8.9	16.0	18.1		6.3	12.1	7.8	14.6	12.4	11.5	3.1	0.5	8.9	12.2	21.4	10.1	12.0	3.7	10.7
K_05-6	12.2	7.4	15.2	18.1	2.0	15.3	6.3		23.7	20.4	14.9	19.0	20.0	13.3	18.3	13.7	8.9	17.5	14.6	8.8	11.3	15.5
K_06-5	14.9	8.3	18.6	4.6	3.6	19.9	12.1	23.7		18.9	15.1	22.2	20.1	14.4	16.6	16.9	11.3	14.2	16.7	18.2	10.9	8.2
CCSP0084	15.9	8.5	17.1	15.7	14.5	2.7	7.8	20.4	19.4		13.7	34.9	32.2	17.3	23.3	15.3	14.2	9.7	18.6	14.7	11.6	18.7
CCSP0211	15.1	7.6	19.2	15.4	8.9	9.8	14.6	14.9	15.1	13.7		20.2	14.2	14.6	14.1	11.7	8.7	16.2	15.9	43.9	43.6	17.2
CCSP0528	11.6	6.4	17.6	17.0	18.3	15.1	12.4	1.0	22.2	34.9	20.2		33.0	15.6	20.9	15.0	11.5	7.3	17.0	16.2	15.0	16.8
CCSP2178	16.0	8.4	17.6	17.4	17.1	7.5	11.5	20.0	20.1	32.2	14.2	33.0		18.6	21.2	11.6	10.1	15.6	13.6	18.0	16.5	21.1
MGS-M1	12.4	6.4	19.1	18.2	3.4	14.9	3.1	13.4	14.4	17.3	14.6	16.2	18.6		12.8	17.1	8.3	15.6	16.7	16.4	17.4	17.6
MGS-M2	11.3	6.2	15.4	16.8	16.1	4.9	0.5	18.3	16.6	23.3	14.1	20.9	21.2	12.8		14.4	11.2	13.4	17.2	16.7	14.5	2.2
MGS-B1	17.2	10.1	23.5	19.1	9.7	37.6	8.9	13.7	16.9	15.3	11.7	15.0	11.6	17.1	14.4		18.3	18.5	13.7	15.0	15.2	22.2
MGS-K1	11.9	8.4	13.9	10.3	6.6	13.2	12.2	8.9	11.3	14.2	8.7	11.5	8.4	8.3	11.2	18.3		13.4	4.6	8.1	7.9	12.3
MGS-MT1	18.9	10.3	23.4	17.5	10.4	19.3	20.9	17.5	14.2	9.7	16.2	7.3	15.6	15.6	13.6	18.5	13.4		4.8	12.7	11.2	22.8
MGS-HA1	17.1	8.7	15.8	15.8	14.6	17.6	10.1	14.6	16.7	18.6	15.9	17.0	13.6	16.7	17.2	13.7	4.6	4.8		22.1	19.0	11.3
MGS-RG1	17.9	8.8	14.3	15.8	8.5	7.9	12.0	8.8	18.2	14.7	43.9	16.2	18.0	16.4	16.7	15.0	8.1	12.7	22.1		63.7	17.3
MGS-RG2	17.2	8.2	14.9	13.2	8.5	8.2	3.7	11.3	10.9	11.6	43.6	15.0	16.5	17.4	14.5	15.2	7.9	11.2	19.0	63.7		11.9
MGS-RG3	17.6	9.2	33.1	14.9	12.2	27.2	10.7	15.5	8.2	18.7	17.3	16.8	21.1	18.2	2.2	22.2	12.3	22.8	11.3	17.3	11.9	

Tabla 5 | Homología a nivel de identidad entre las diferentes secuencia de aminoácidos que codifican esterasas/lipasas identificadas en la presente Tesis Doctoral.

La temperatura óptima de todas las enzimas oscila entre los 12 y los 75ºC y en la mayoría de los casos se observa que los valores son superiores a los de la temperatura del hábitat de procedencia (**Tabla 3**; entre 4 y 65ºC). Más concretamente las enzimas del Lago Arreo presentan valores de actividad máximos entre los 12 y 40ºC, si bien la temperatura media del agua es de 8.7ºC. Las aisladas de ambientes marinos superficiales (Mar de Barents) o bacterias de dichos ambientes (*Cycloclasticus* sp. ME7) entre los 25 y 50ºC, si bien la temperatura de dicho hábitat oscila entre los 4 y 23ºC. Aquellas de las fosas marinas entre los 16 y 70ºC, si bien su temperatura nunca supera los 16.5ºC o es inferior a los 14ºC. Solo en el caso de la enzima MGS-HA, procedente de una bacteria termófila aislada de una fuente hidrotermal tiene una temperatura óptima (75ºC) que se aproxima a la temperatura de crecimiento (65ºC) y al origen termófilo de la misma. Estos resultados concuerdan con estudios recientes que han demostrado que una alto porcentaje de enzimas microbianas presentan una temperatura óptima superior a la temperatura de crecimiento del microorganismo productor (Kube *et al*., 2013).

Pese a ello, los resultados sugieren que las enzimas de microorganismos de origen marino independientemente de su salinidad o de ambientes acuáticos ligeramente salinos presentan en su mayoría actividades que sugieren la presencia de bacterias psicrófilas y/o mesófilas. Sin embargo, uno de los resultados más relevantes de esta Tesis Doctoral es que los ambientes marinos hipersalinos, pese a su baja temperatura (≤16.5ºC) pueden contener enzimas con características similares a aquellas de organismos termófilos o hipertermófilos. Sirva de ejemplo la enzima MGS-M2 (ver Capítulo 3). Esto no significa la presencia de organismos termófilos en estos ambientes, en particular la fosa hipersalina *Medee*, sino que la acción combinada del efecto de la alta concentración de sal y la presión (≥ 350 bares) pueden favorecer, mediante mecanismos distintos a los descritos hasta la fecha, las actividades y estabilidades de las proteínas a temperaturas muy altas. Este hecho ha quedado demostrado en esta Tesis Doctoral mediante la resolución estructural de 5 de las proteínas identificadas (ver Capítulo 3). Resaltar, que hasta la fecha no se había descrito ninguna enzima de un ambiente no termófilo, capaz de mantener la integridad estructural y la actividad a una temperatura de 70-75ºC.

Las concentraciones de sal (NaCl) para la actividad óptima varían entre 0 y 4.0 M (**Figura 17**). De las 25 enzimas estudiadas, solo 3 mostraron mayor actividad en ausencia de sal (LAE3, MGS-M5 y MGS-RG3), lo que sugieren que las enzimas marinas o de medios acuáticos ligeramente salinos son activadas en presencia de dicha sal. Pese a ello, es de destacar es que las enzimas procedentes de ambientes aquellos

ambientes caracterizados por menor salinidad (Lago Arreo y Mar de Barents, salinidad ≤ 35 g/L) presentan una fuerte bajada de la actividad a concentraciones de sal superiores a 1.2 M (70 g/kg). El hecho de que tres enzimas de ambientes marinos profundos caracterizados por bajos niveles de salinidad (≤ 35 g/L) presenten alta actividad a 2.8-3.6 M de NaCl (163-210 g/kg), sugiere que la presión pueda ser un factor adicional que favorezca la estabilidad a altas concentraciones de sal. Dicho esto, las enzimas de fosas marinas hipersalinas son las que presentan mayores niveles de actividad a concentraciones muy elevadas de NaCl (hasta 4 M).

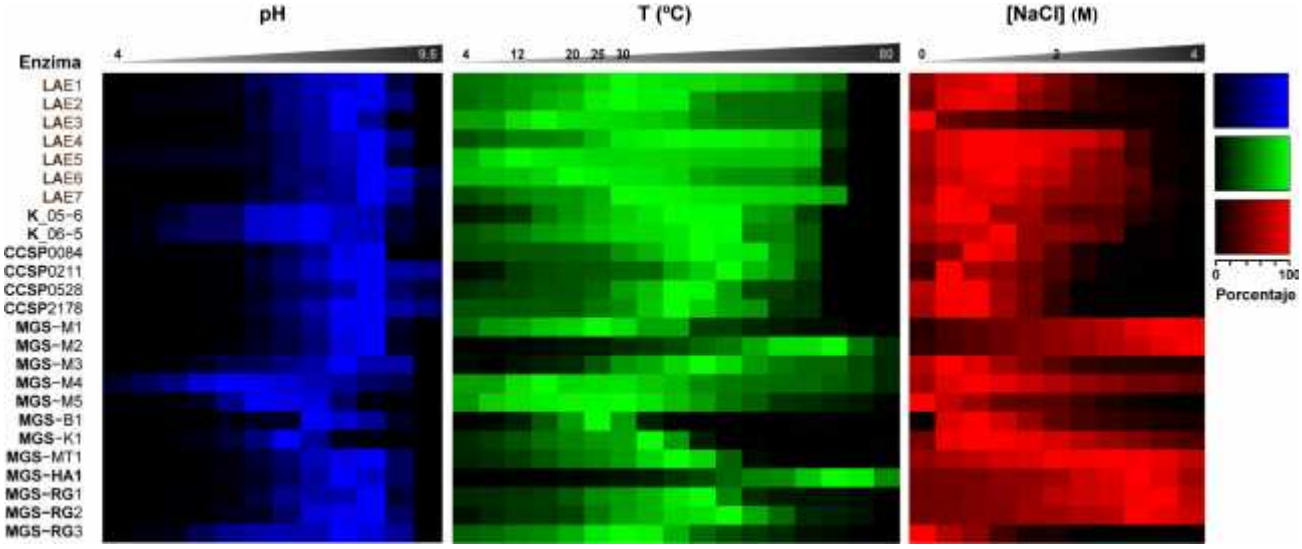


Figura 17 | Mapa de color indicando los parámetros óptimos de actividad de las enzimas caracterizadas en la presente Tesis Doctoral. Los parámetros, en orden de aparición de izquierda a derecha, incluyen, pH, temperatura y concentración de sal (NaCl).

Las concentraciones de sal (NaCl) para la actividad óptima varían entre 0 y 4.0 M (**Figura 17**). De las 25 enzimas estudiadas, solo 3 mostraron mayor actividad en ausencia de sal (LAE3, MGS-M5 y MGS-RG3), lo que sugieren que las enzimas marinas o de medios acuáticos ligeramente salinos son activadas en presencia de dicha sal. Pese a ello, es de destacar que las enzimas procedentes de ambientes caracterizados por menor salinidad (Lago Arreo y Mar de Barents, salinidad ≤ 35 g/L) presentan una fuerte bajada de la actividad a concentraciones de sal superiores a 1.2 M (70 g/kg). El hecho de que dos enzimas de ambientes marinos profundos (MGS-M5 y MGS-RG3) caracterizados por bajos niveles de salinidad (≤ 35 g/L) presenten alta actividad a 2.8-3.6 M de NaCl (163-210 g/kg), sugiere que la presión pueda ser un factor adicional que favorezca la estabilidad a altas concentraciones de sal. Dicho esto, las enzimas de fosas marinas hipersalinas son las que presentan mayores niveles de actividad a concentraciones muy elevadas de NaCl (hasta 4 M) (**Figura 17**).

Pese a que se ha demostrado en la bibliografía una relación entre el *p*I de una proteína y la actividad frente a alta concentración de sal (Ferrer *et al*., 2011), no se ha encontrado en la presente Tesis Doctoral, una correlación entre ambos factores (ver **Tabla 2** y **Figura 17**). Esto sugiere que factores estructurales, aún por definir, pueden estar implicados en estas diferencias.

6.4 Amplia colección de enzimas: diversidad enzimática y promiscuidad

Actualmente la introducción a nivel general en la industria química de síntesis de las enzimas está siendo mucho más lenta de lo que inicialmente se suponía. Una de las razones principales es que cuando las enzimas disponibles no son capaces de promover la reacción química deseada, el proceso requerido para finalmente disponer de la enzima que de novo pueda catalizar la reacción de interés, es un proceso lento y costoso. Este proceso es más fácil cuando se parte de enzimas disponibles comercialmente, y mucho más complejo cuando ello no es posible. Por esta razón enfocar la investigación en la resolución sistemática y

ordenada de las reacciones químicas de interés, y encontrar enzimas apropiadas para ello, es de especial importancia. Las presente Tesis Doctoral ha pretendido abordar este tema, creando un amplio Catálogo de enzimas verdaderamente promiscuas. Este Catálogo también permitiría no solo acceder a nuevas reactividades sino también a entender el fenómeno de promiscuidad enzimática.

Para ello, y a diferencia de otros estudios anteriores durante la presente Tesis Doctoral, planteamos la caracterización bioquímica usando un set muy amplio de sustratos, con un total de 210. Más concretamente se emplearon 137 ésteres para la caracterización de actividad esterasa y/o lipasa, 31 derivados de azúcares para la caracterización de actividad glicosidasa, 41 aldehídos y cetonas para la caracterización de actividad aldo-ceto reductasa y 1 para la caracterización de actividad lactato dehidrogenasa. Se ha pretendido con ello, responder a una pregunta: ¿Las enzimas identificadas constituyen nuevas variantes enzimáticas que poseen capacidades catalíticas similares o no a las descritas en las bases de datos?

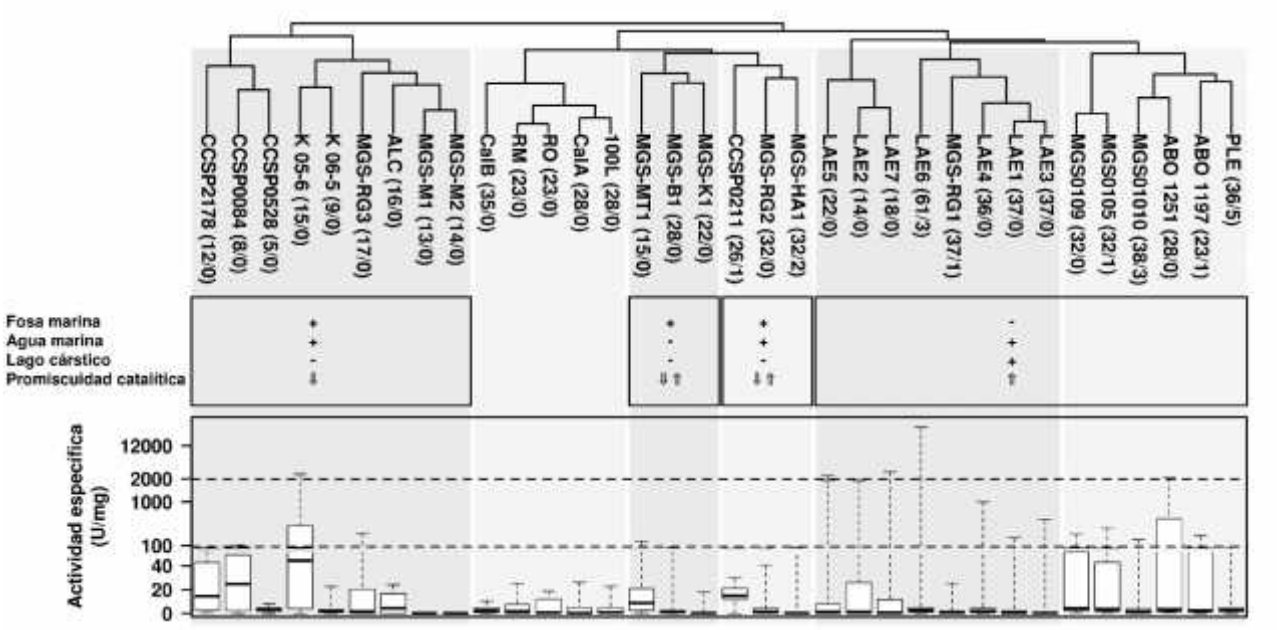


Figura 18 | Agrupación gráfica de la similitud y/o diferencias en la reactividad (A) y actividad específica (units mg^{-1}) (B) frente a un set de 137 ésteres diferentes de las diferentes esterasas caracterizadas en la presente Tesis Doctoral, así como otras enzimas similares comerciales (CalA, CalB, 100L, RM, RO, y PLE) y otras descritas de ambientes similares en la bibliografía (ABO 1197, ABO 1251, MGS01010, MGS 0105, y MGS0109). El panel A muestra la agrupación jerárquica en base a una distribución binomial de la presencia o ausencia de actividad para un stock de 137 sustratos, para cada una de las enzimas analizadas. Abreviaturas: CalA, lipasa A de *Candida antarctica* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); CalB, lipasa B de *Candida antarctica* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca);100L, lipasa de *Thermomyces lanuginosa* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); RM, lipasa de *Rhizomocur miehei* (suministrada por Novozymes A/S, Bagsvaerd, Dinamarca); RO, lipasa de *Alcaligenes* sp. (suministrada por Meito Sangyo Co. (Japón); PLE, esterasa de hígado de cerdo.

La **Figura 18** resume gráficamente las similitudes y diferencias en la reactividad de las 23 hidrolasas con actividad esterasa caracterizadas en la presente Tesis Doctoral, así como otras enzimas descritas en la bibliografía. La **Figura 18A** muestra la agrupación jerárquica en base a una distribución binomial de la presencia o ausencia de actividad para un stock de 137 sustratos. Para cada grupo de enzimas estudiadas, se muestra también el tipo de hábitat (fosa marina, agua de mar y lago cárstico) y la estimación de la versatilidad de la enzima según la estimación del número de sustratos (rango de sustrato: bajo (⇓), medio (⇓⇑) o alto (⇑)) sobre los que actúa una determinada enzima. La **Figura 18B** resume la actividad específica (unidades mg^{-1}) media y el rango de actividad, para todos los ésteres capaces de ser hidrolizados, para cada una de las enzimas estudiadas. Como se ha comentado en el Capítulo 5, a través de los resultados mostrados en la **Figura 18** se concluye que: *i*) las enzimas descritas en esta Tesis Doctoral presentan valores de actividad similares a los de otras enzimas homólogas descritas en la bibliografía, incluidas

enzimas comerciales (**Figura 18B**); y *ii*) que las enzimas aisladas de un mismo ambientes presentan un perfil de sustratos similar y diferenciado a las de otras enzimas de otros hábitats (**Figura 18A**), independientemente de las diferencias a nivel de posición filogenética basadas en la secuencia de aminoácidos de las mismas (**Figura 19**). Esto podría ser debido al hecho de que ambientes con características geoquímicas similares pueden albergar una biodiversidad muy similar independientemente de la localización geográfica; por lo tanto esta biodiversidad similar podría contener enzimas con reactividades adaptadas a ese nicho ecológico. Pese a ello, también se observan diferencias a nivel de actividad específica (**Figura 18B**) y el número de sustratos susceptibles de ser hidrolizados entre las enzimas estudiadas en un mismo hábitat (**Figura 18A**).

De los ésteres testeados, las hidrolasas caracterizadas fueron capaces de hidrolizar de 5 a 61 ésteres. Estas diferencias en especificidad de sustrato, se correlacionan con la amplia diferencia a nivel de secuencia entre las diferentes enzimas (**Tabla 5**). Sin embargo, es de destacar que enzimas que presentan muy bajos niveles de identidad, por ejemplo LAE1 y LAE3 (5.7%) del Lago Arreo son bioquímicamente muy similares entre sí en lo que a especificidad de sustrato se refiere (**Figura 18A**). Sin embargo no lo son a nivel de temperatura óptima (30ºC y 16ºC, respectivamente) y concentración de NaCl óptima (1.2 vs 0 M, respectivamente) (**Figura 17**). Este hecho también es extensible para otras enzimas de otros hábitats como por ejemplo las enzimas MGS-M1 y MGS-M2 de la fosa *Medee*. Ambas enzimas, que comparten solo un 12.8% de identidad de secuencia son bioquímicamente muy similares entre sí en lo que a especificidad de sustrato se refiere (**Figura 18A**), si bien presentan parámetros físico-químicos para su actividad óptima muy diferenciados. Sirva de ejemplo las diferentas en temperatura óptima: 25ºC para MGS-M1 y 70ºC para MGS-M2. Por lo tanto, los datos sugieren que mientras que la especificidad de sustrato de las enzimas de un mismo hábitat es más similar entre sí independientemente de la homología, las diferencias en los parámetros óptimos de actividad son más acusadas e influyen de manera mucho más heterogénea.

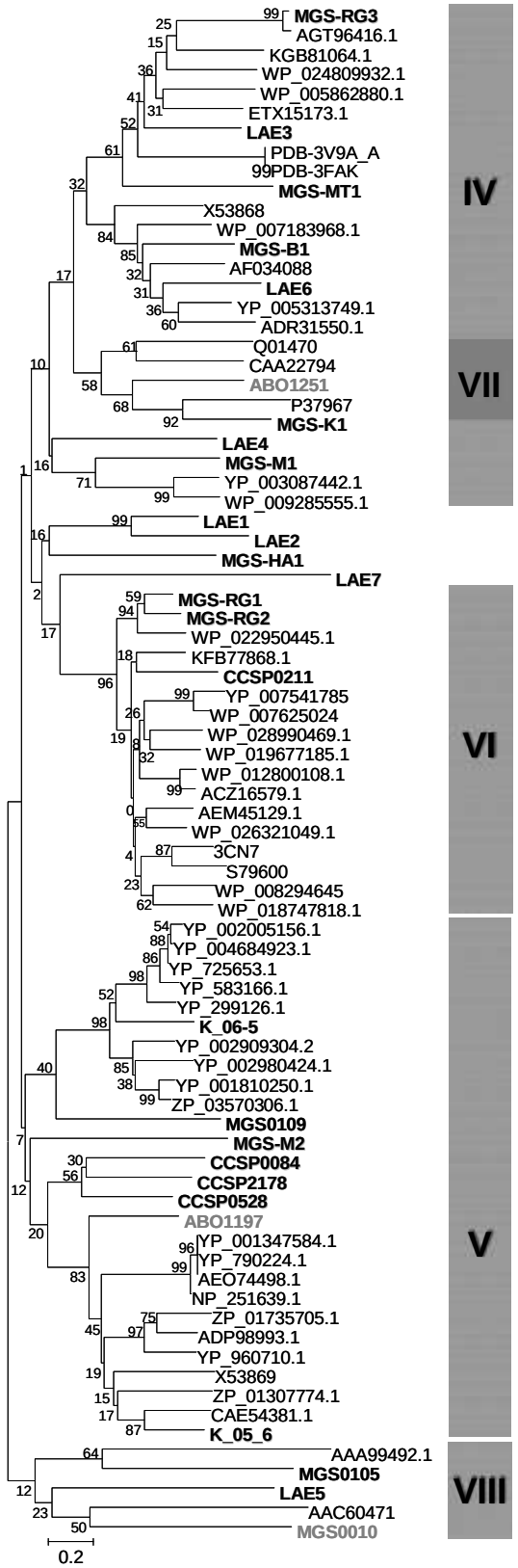


Figura 19 | Relación filogenética de las esterasas identificadas en la presente Tesis Doctoral (negrita y sombra). Ver detalles en Alcaide *et al.* (2015). La figura muestra las sub-familias de la familia de esterasas/lipasas a las que pertenecen las diferentes enzimas (Arpigny y Jaeger, 1999).

Centrándonos en la especificidad de sustrato, los resultados presentados han mostrado que un amplio número de las enzimas caracterizadas muestran una alta promiscuidad de sustrato. Tal información se ha conseguido a través del estudio de una amplia diversidad enzimática y una amplia diversidad química de los sustratos empleados, que se distinguen claramente de otros trabajos o estudios similares descritos en la bibliografía. Estos niveles de promoscuidad son especialmente visibles en las enzimas LAE6 (Lago Arreo), que también es la enzima más activa de todas las analizadas en esta Tesis Doctoral (**Figura 20**), MGS-HA1 (Isla de San Pablo), MGS-B1 (fosa *Bannock*), MGS-K1 (fosa *Kryos*), y MGS-RG1 y MGS-RG2 (*R. exoculata*). Estas enzimas son capaces de hidrolizar una amplia diversidad de ésteres con estructuras muy diferencias que en su conjunto no se ha demostrado hasta la fecha que sean sustratos de una sola esterasa/lipasa descrita. Estos sustratos incluyen, triglicéridos, ésteres de alkilo, alkenilo o arilo halogenados y no halogenados, derivados de ésteres de ácido cinámico, ésteres de carbohidratos, lactonas, epóxidos. Además, 10 de las 23 (o el 43% del total) esterasas/lipasas estudiadas mostraron altos niveles de enantioselectividad aparente (rango de 20 a 816; Janes **et al**., 1997; Baumann *et al*., 2001) para ésteres quirales como (*S*)-mandelato de metilo (CCSP2178, CCSP0211, LAE4, RG1), (*S*)-4-nitrobenzoato de glicidilo (RG1), (*S*)-3-bromo-2-metil-propionato de metilo (MGS-K1, RG1, RG2), (*S*)-lactato de metilo (CCSP0211, RG2), (*R*)-mandelato de metilo (LAE1, LAE3, RG3), y (*R*)-butirato de glicidilo (LAE6).

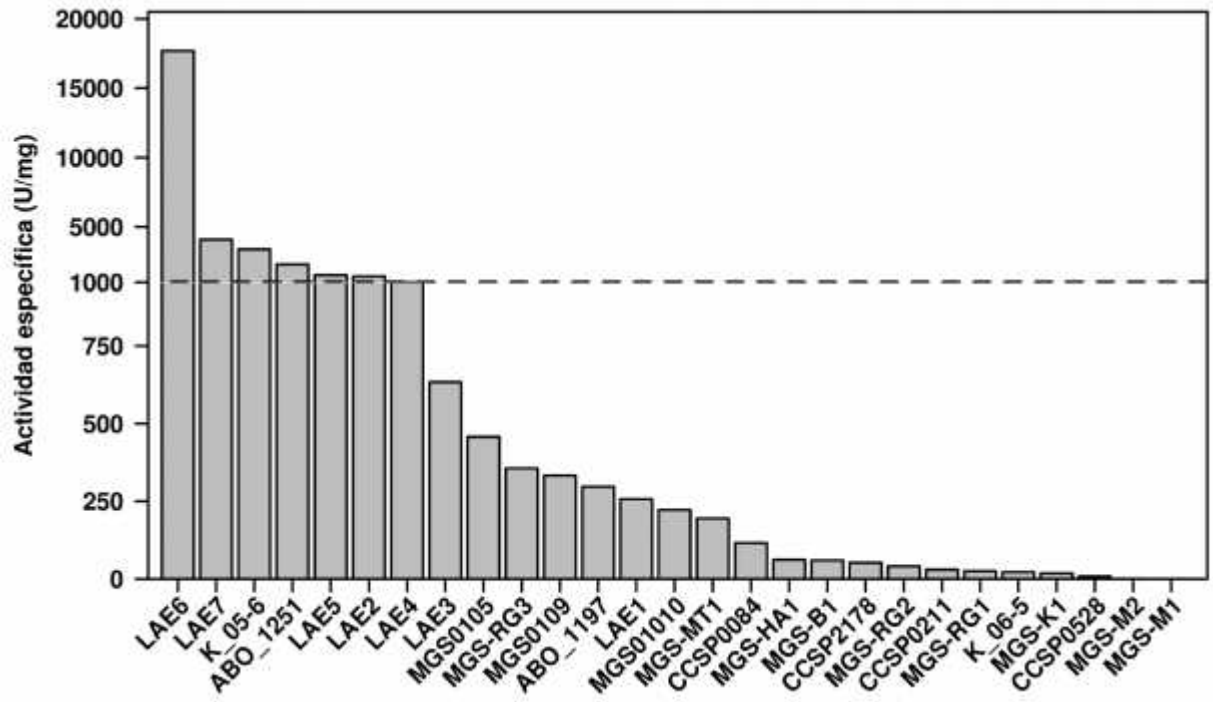


Figura 20 | Actividad específica máxima (para el mejor sustrato) de las diferentes esterasas/lipasas caracterizadas en la presente Tesis Doctoral, así como otras enzimas de ambientes similares descritas en la bibliografía (ABO 1197, ABO 1251, MGS01010, MGS 0105, y MGS0109).

Pese a que se ha demostrado el alto potential de las esterasas descritas en la presente Tesis Doctoral en reacciones de hidrólisis, su uso en biotransformaciones está aún por determinar. En este sentido hay que recordar que factores tales como la actividad y estabilidad frente a disolventes orgánicos y el efecto de la inmovilización, son claves, entre otros factores, en la incorporación de tales enzimas en el mercado (Plou *et al*., 2002, 2003; Bomarius y Paye, 2013).

Otro ejemplo a resaltar en lo que se refiere a como la metagenómica puede proporcionar nuevas enzimas con actividades enzimáticas diferentes a las ya conocidas, lo constituyen las enzimas K_05-6 y K_06-5 aisladas de comunidades microbianas del Mar de Barents y CCSP0084, CCSP0528 y CCSP2178 de la bacteria *Cycloclasticus* sp. ME7. Estas enzimas, que presentan homología tanto con esterasas como con hidrolasas de enlaces carbono-carbono. Si

bien, las enzimas con actividad esterasas tienen un alto potencial biotecnológico (ver Introducción), las MCP hidrolasas no lo tienen debido a que todas las enzimas de este tipo descritas en la bibliografía presentan una especificidad de sustrato muy restringida; solo son capaces de hidrolizar HOHD y HOPHD, moléculas que se forman durante la degradación de compuestos aromáticos (p.e. catecol y tolueno) (Alcaide *et al*., 2013). De hecho, estas enzimas paticipan en rutas de degradación de aromáticos y no son capaces, en su mayoría, de hidrolizar enlaces ésteres. Los datos presentados en esta Tesis han demostrado por primera vez que una misma enzima puede poseer ambas actividades, esterasa y MCP hidrolasa, con un grado de promiscuidad de sustrato hasta ahora desconocido. Las 5 enzimas analizadas han mostrado actividad no solo frente a HOHD y HOPHD sino también frente a un total de más de 12 ésteres diferentes, incluídos ésteres quirales. Por lo tanto, la presente Tesis Doctoral abre por primera vez la posibilidad de que enzimas con actividad dual esterasa:MCP hidrolasa sean más abundante en la Naturaleza de lo que inicialmente se pensaba. La caracterización de otras 9 enzimas similares aisladas de ambientes crónicamente contaminados que se está realizando actualmente en el laboratorio, confirman este hecho. Dado que estas enzimas han sido aisladas de ambientes contaminados o bacterias degradadoras de aromáticos, es posible que la presión producida por eventos de contaminación crónica pueda abrir nuevos escenarios evolutivos que permitan la adquisición de tales actividades a fin de que los microorganismos productores puedan acceder a dichas moléculas.

La presencia de esta actividad dual en ambientes no contaminados no se ha demostrado hasta la fecha en la bibliografía. Esto, junto con el hecho de que enzimas homólogas (Alcaide *et al*., 2013; Capítulo 3) similares a K_05-6, K_06-5, CCSP0084, CCSP0528 y CCSP2178, se encuentran en genomas de organismos con capacidades biodegradativas, sugiere que dicha actividad dual puede estar más asociada a comunidades microbianas de ambientes contaminados.

6.5 Entendimiento de los mecanismos subyacentes a la promiscuidad de las enzimas

Durante la presente Tesis Doctoral algunas enzimas representativas con características interesantes (promiscuidad y adaptación a parámetros físico-químicos extremos) fueron seleccionadas y sometidas a fermentaciones con alta densidad celular y técnicas de cristalización. En paralelo se aplicaron técnicas de mutagénesis dirigida. Se pretendía con ello proporcionar un conocimiento estructural detallado de las enzimas más representativas y entender el fenómeno y relevancia de la promiscuidad enzimática y la transferencia a otras clases de enzimas, así como entender los mecanismos de adaptación a condiciones poli-extremas de temperatura, salinidad y presión.

Para ello, se seleccionaron y cristalizaron enzimas de tres fuentes distintas: i) LAE6 del Lago Arreo, ejemplo de esterasa ampliamente promiscua; ii) CCSP0084 de *Cycloclasticus* sp. ME7, ejemplo de enzima con actividad dual esterasa:MCP hidrolasa; iii) 5 enzimas (3 esterasas, 1 aldo-ceto reductasa y 1 lactato dehidrogenasa) de ambientes marinos profundos, en particular la fosa marina hipersalina *Medee*, y la fosa marina *Matapan-Vavilov* de baja salinidad. La resolución de la estructura se logró con éxito en todos los casos, habiendo reportado hasta la fecha solo la estructura de todas ellas (**Figura 21**), menos de LAE6, que está pendiente de publicación. EL análisis de la estructura de la enzima CCSP0084, en combinación con experimentos de mutagénesis dirigida, ha revelado por primera vez que en la transformación de una esterasa a una MCP hidrolasa están implicados un bajo número de aminoácidos cercanos al centro activo (**Figura 22**). En particular, se han identificado 3 residuos en la proteína CCSP0084, cuya mutación afecta a la capacidad de esta enzima de hidrolizar enlaces éster y enlaces C-C, y que por lo tanto estos residuos están involucrados en el reconocimiento de los distintos sustratos y en el consiguiente grado de promiscuidad. El hecho de que solo un bajo número de aminoácidos esté involucrado en la modulación de la promiscuidad sugiere que factores ambientales, como la contaminación crónica por la actividad antropogénica, puedan ser suficientes para la evolución de diferentes tipos de actividad hidrolasa, hecho éste que se ha demostrado para la aparición de atividad dehalogenasa (Beloqui *et al*., 2010).

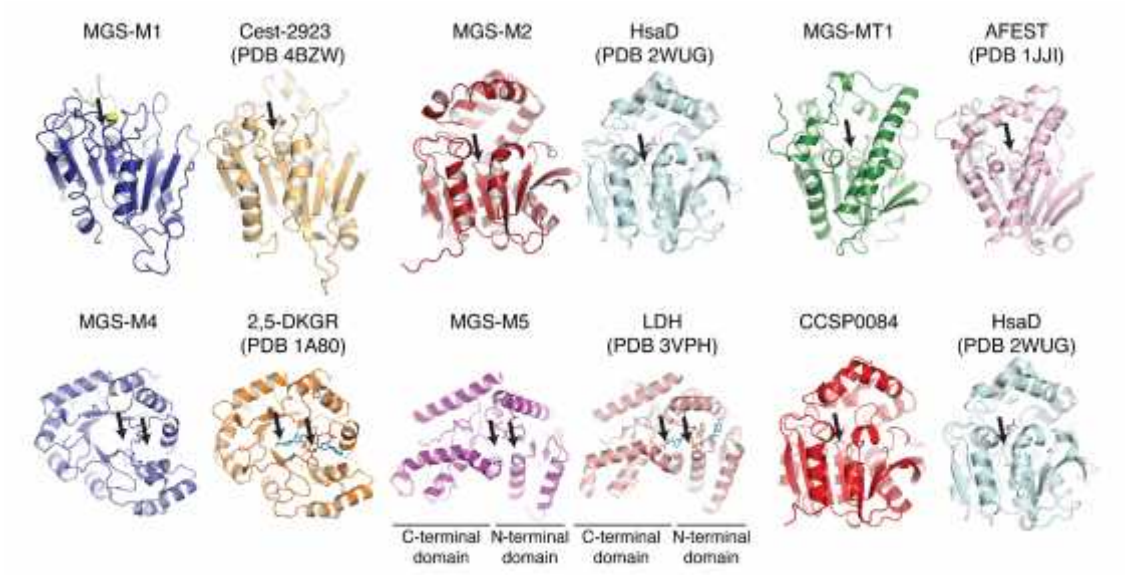


Figura 21 | Comparación de las estructuras de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Las enzimas se representan en viñetas donde se aprecia la localización de la serina catalítica (para estersas en la parte superior, y la enzima CCSP0084 en la parte inferior) y de los sitios de unión a NADPH/NADH y sustrato (para las enzimas MGS-M4 y MGS-M5 en la parte inferior).

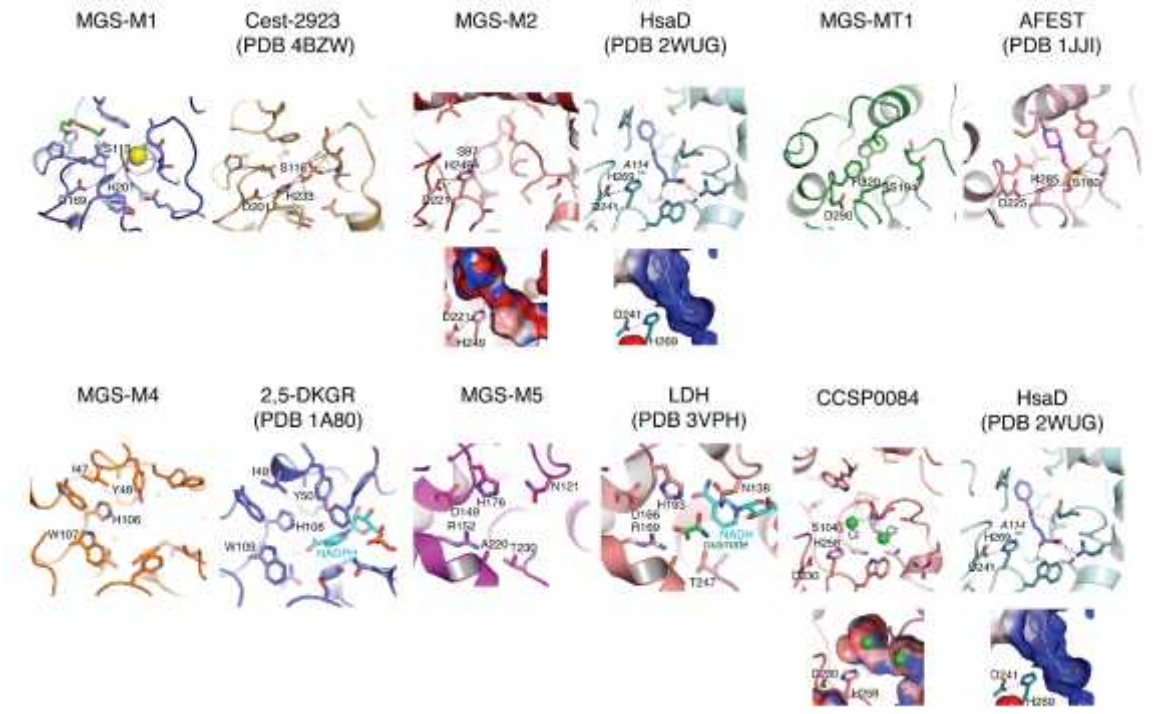


Figura 22 |Comparación de los centros activos putativos de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Para cada enzima se muestran los sitios de unión a sustrato y aquellos que participan en la catálisis. En el caso de enzimas con actividad esterasa (parte superior y enzima CCSP084 en la parte inferior) se representa la triada catalítica. El canal de unión a sustrato de la enzima MGS-M2 y su homólogo estructural HsaD se representan como representación de superficie accesible a solvente; el potencial electroestático se representa con colores, donde se aprecian las diferencias en residuos cargados. Se muestran también los residuos de unión a NADH/NADPH y de sustrato para la aldo-ceto reductasa MGS-M4 y la lactato dehidrogenasa MGS-M5, y sus homólogos estructurales.

El análisis de la estructura de las enzimas MGS-M1, M2, M4, M5 y MT1, de ambientes marinos profundos, es de especial interés por dos razones. Primero porque constituye la mayor colección de estructuras de enzimas de ambientes marinos profundos. En segundo lugar porque ha permitido obtener información que revela mecanismos moleculares desconocidos involucrados en la adaptación de las proteínas a condiciones poli-extremas (**Figura 23**). En tercer lugar porque ha permitido entender las diferencias a nivel de centros activos que explican las distintas especificidades de sustratos de las enzimas entre sí y en comparación con otras descritas en las bases de datos.

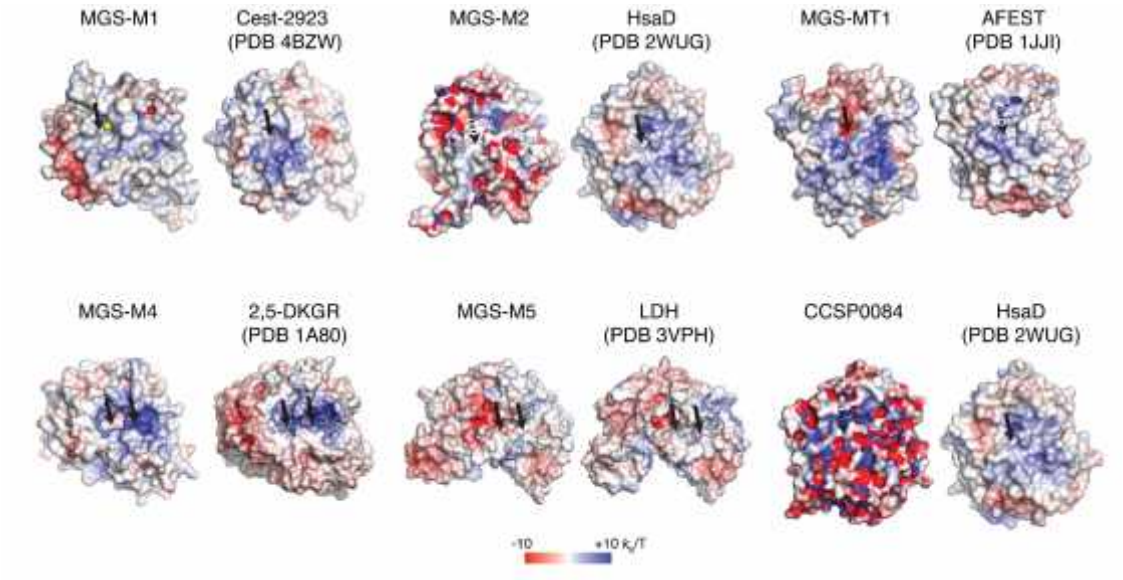


Figura 23 | Comparación de la superfie de las enzimas cristalizadas y presentadas en esta Memoria, y sus homólogos estructurales en base a búsquedas por similitud. Para cada enzima se muestran los sitios de unión a sustrato y aquellos que participan en la catálisis. La estructura aparece como representación de superficie accesible a solvente; el potencial electroestático (en unidades kb/T = constante de Boltzmann frente a la temperatura) se representa con colores, donde se aprecian las diferencias en residuos cargados. Las flechas indican la localización de la serina catalítica (para estersas en la parte superior, y la enzima CCSP0084 en la parte inferior) y de los sitios de unión a NADPH/NADH y sustrato (para las enzimas MGS-M4 y MGS-M5 en la parte inferior).

Como se ha comentado en el Capítulo 4, el análisis de las estructuras de las enzimas de fosas marinas revela no solo una confirmación y una distribución de cargas (**Figura 23**) diferente a la de otras enzimas similares descritas en la bibliografía. Ese hecho se ha observado también para la enzima CCSP0084 de *Cycloclasticus* sp. ME7 (**Figura 23**). Esto concuerda con el hecho de que las propiedades bioquímicas y especificidad de sustrato de las enzimas presentadas en esta Memoria son muy diferentes a las de enzimas similares descritas hasta la fecha.

Un estudio comparativo entre todas las estructuras presentadas en la Memoria arroja diferencias significativas. Por ejemplo, MGS-M1 no presenta una tapadera (sub-dominio α-hélice) que si presentan MGS-M2 y CCSP0084 (**Figura 24**). Es posible que dicha tapadera pudiera estar implicada no solo en las diferencias en la especificidad de sustrato, sino también en la estabilidad de dichas proteínas. Así, mientras MGS-M1 muestra actividad frente a sustratos largos, por ejemplo, *p*-nitrofenil dodecanoato (ver **Figura S2**, Capítulo 3), MGS-M2 y CCSP084 (más activos con *p*-nitrofenil butirato) lo son solo frente a sustratos más cortos (ver **Figura S1** en Capítulo 2). Además, mientras que MGS-M1 es más activa a 25ºC, MGS-M2 y CCSP0084 lo son a mayor temperatura (70 y 45ºC, respectivamente) (**Figura 17**). Finalmente, MGS-M2 y CCSP0084 comparten un mayor grado de similitud a nivel de cargas, incluídas en la zona de la tapadera (**Figura 23**). Dicho esto, una validación experimental es necesaria para determinar la implicación real de dicha tapadera en ambos factores.

La similitud a nivel de cargas y a nivel de la presencia de un sub-dominio α-hélice en MGS-M2 y CCSP0084 contrasta con el hecho de que

solo 7 de los 18 posibles aminoácidos involucrados en catálisis o sitios de unión a sustrato se compartan entre ambas enzimas (**Figura 25A**). Esto, junto con la presencia de dos ligandos desconocidos en la estructura de MGS-M2, uno de ellos unido covalentemente a la Ser97 (**Figura 25A**), puede se responsable, sin embargo, de la diferente especificidad de sustrato entre ambas enzimas, si bien, ambas se encuentran entre las enzimas presentadas en la Memoria que actúan sobre menor número de sustratos: 9 en el caso de CCSP0084 y 13 en el caso de MGS-M2, de un total de más de 130 sustratos testeados. Más concretamente, entre los sustratos diferenciadores, CCSP0084 actúa sobre enlaces C-C en HODH y HOPDH, acetato de geranilo y acetato de butilo, mientras MGS-M2 actúa sobre triglicéridos cortos, lactato de metilo, bromo acetato dimetilo y triacetato de glucal.

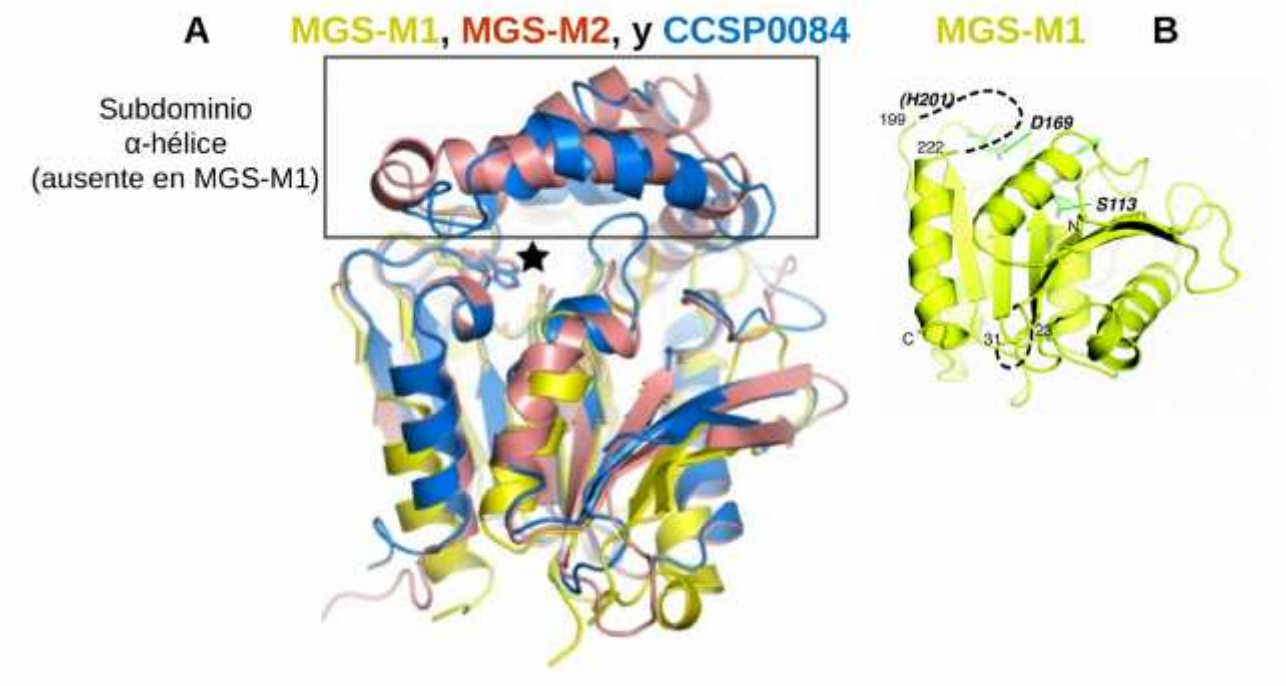


Figura 24 | A. Vista detallada de las estructuras de MGS-M1, MGS-M2 y CCSP084 superpuestas. Como se aprecia las tres enzimas presentan un plegamiento α/β hidrolasas, si bien la esterasa MGS-M1 no presenta un sub-domino α-hélice, que si presentan MGS-M2 y CCSP0084. Estrella negra: localización del centro activo. El panel (B) representa la estructura detallada de MGS-M1 donde se aprecian los elementos estructurales ausentes en comparación con las otras dos proteínas.

6.6 Consideraciones finales

La presente Tesis Doctoral no solo ha incrementado sustancialmente el número de enzimas caracterizadas, que suponen un aumento de al menos el 25% respecto a las enzimas caracterizadas en la bibliografía por técnicas metagenómicas (Martínez-Martínez *et al.*, 2013), sino abre la posibilidad para futuros trabajos relacionados con la aplicación de tales enzimas en biotransformaciones de interés. Si bien, los resultados indican un alto potencial en reacciones de hidrólisis, hay que tener en cuenta que en un proceso biocatalítico, como en cualquier reacción química en general, el solvente o medio de reacción tiene a menudo tanta importancia como la estructura del sustrato y el modo de acción enzimático, de manera que esta variable añade un factor importante de complejidad a la hora de diseñar biotransformaciones con las enzimas presentadas en esta Memoria.

Es bien conocido que el proceso de investigación y desarrollo de nuevos procesos enzimáticos, más fácil cuando es razonable partir de enzimas disponibles comercialmente y mucho más complejo cuando ello no es posible, es a menudo costoso y lento. Excepto para aquellas grandes corporaciones industriales con capacidad técnica suficiente y grandes volúmenes de producción y venta de compuestos químicos específicos, para la mayor parte de las empresas

incluso de tamaño medio, y para los laboratorios ello supone un inconveniente suficientemente significativo como para que renuncien a la posibilidad de la aplicación de técnicas metagenómicas en sus procesos industriales. Ello es especialmente notorio cuando los volúmenes de venta del compuesto a producir son bajos y los márgenes comerciales, forzados por la competencia, muy estrechos, o bien porque la introducción de cualquier cambio en el proceso productivo supone una dificultad a menudo insalvable, en términos del tiempo y dinero.

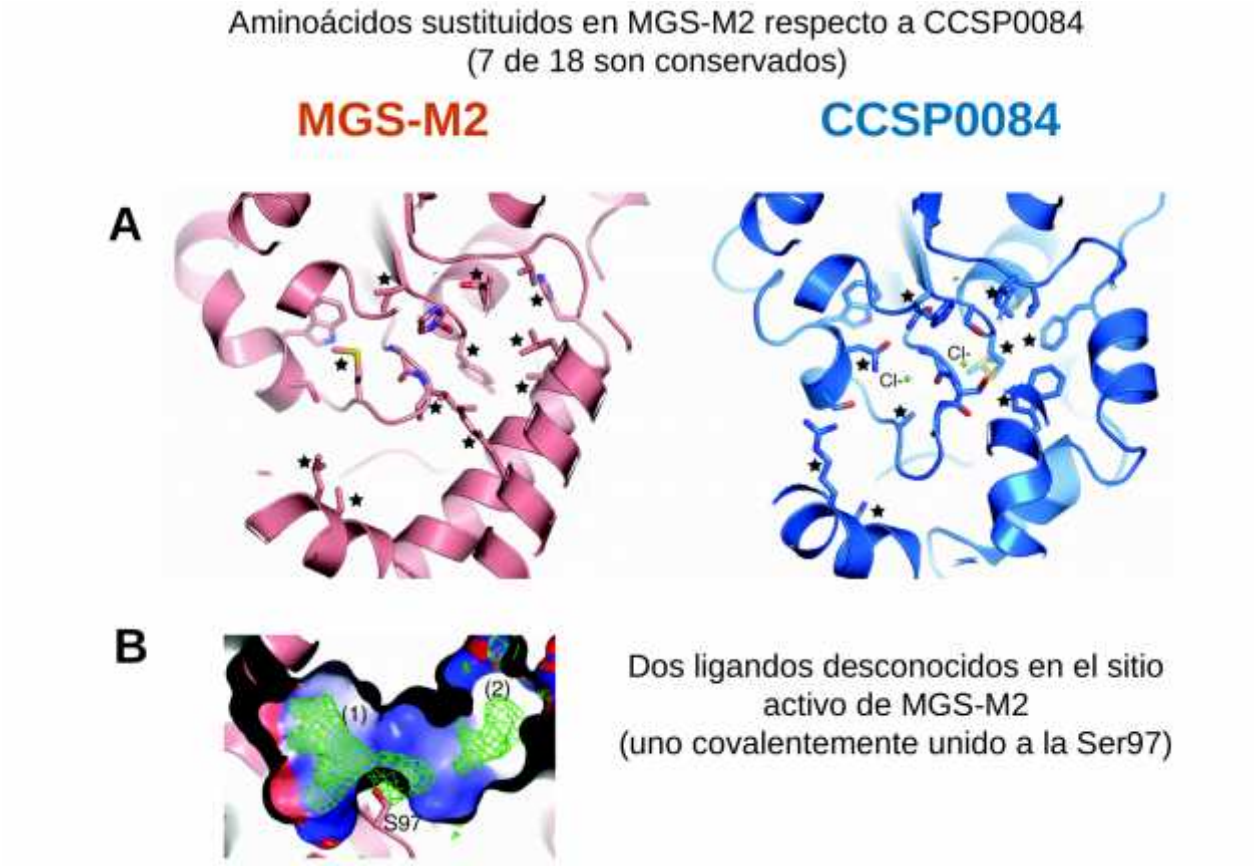


Figura 25 | Vista detallada de las estructuras de MGS-M2 y CCSP084 donde se detallan los residuos cercanos al centro activo. Las estrellas negras representan los residuos potenciales que participan en la catálisis o unión a sustrato. B. Detalle del canal de la estructura de MGS-M2 donde se aprecia la presencia de dos ligandos desconocidos, uno de ellos unido a la Ser97.

Es por ello, que en el futuro, habría que enfocar la investigación básica en biocatálisis a la resolución sistemática y ordenada, mediante el empleo de las enzimas presentadas en esta Memoria (y otras), de las reacciones químicas de interés para la industria, poniendo especial énfasis en aquellas de gran potencial de aplicación, pero que se enmarcan en un tipo de reacción química poco desarrollado hasta la fecha. En este punto, seria de especial interés la edición de un catálogo accesible a empresas y grupos de la academia donde se resumen las propiedades generales de tales enzimas, a fin de atraer financiación o iniciar colaboraciones en el marco de desarrollos biocatalíticos. En segundo lugar habría que realizar ensayos de las capacidades enzimáticas en presencia de medios orgánicos y no acuosos como medio de incrementar las capacidades catalíticas de las enzimas estudiadas en la presente Tesis Doctoral. Por otro lado, habría que realizar un estudio detallado de los efectos que sobre la actividad enzimática de las diferentes enzimas presentadas en esta Memoria tienen los procesos de inmovilización, con independencia de la mejora evidente que suponen en relación a la posibilidad que ofrecen de ser recuperados al final del proceso y estar en disposición de ser utilizados de nuevo, y a la mejora de su resistencia a los efectos ambientales y de los disolventes. Solo a través de un estudio conjunto de éstas, y otras

variables, se podrá llegar a conocer el potencial de las nuevas enzimas descritas en esta Tesis Doctoral en futuras biotransformaciones de interés básico e industrial.

Finalmente me gustaría mencionar que en el estudio de los procesos biocatalíticos, el conocimiento profundo del mecanismo de la reacción química y el entender de qué manera la enzima modifica los estados energéticos de transición por los que pasa el proceso, de manera que se reduzca la cantidad de energía implicada en el proceso y la reacción química se vea favorecida, es clave. Este estudio, que requiere de la caracterización bioquímica y la cristalización de la enzima para determinar su estructura- especialmente de la zona del centro activo- y de una notable capacidad de cálculo, debe ser realizado de manera específica para cada enzima. Las estructuras aportadas en la presente Memoria pueden abrir futuros trabajos en estas líneas en los que la aplicación de técnicas de dinámica molecular, mecánica cuántica y otros sistemas de análisis (Martín-Garcia, *et al*., 2013; Mendieta-Moreno *et al*., 2014) pueden ser de interés, junto con herramientas de muetagénesis dirigida y aleatoria.

CAPÍTULO 7

CONCLUSIONES

1. Las características geoquímicas de los ambientes producen un sesgo importante en el índice de incidencia de clones activos.
2. Se ha demostrado que el empleo de herramientas de metagenómica funcional permite la identificación de biocatalizadores novedosos y auténticamente versátiles, que incluyen esterasas, lipasas, glicosidasas, aldo-ceto reductasas y lactato dehidrogenasas.
3. Al menos un 45% de las 25 enzimas identificadas y caracterizadas son enzimas ampliamente promiscuas, ya que poseen especificidades de reacción relajadas, actividad frente a un amplio número de sustratos, y alta estereoselectividad y estabilidad bajo amplias condiciones.
4. Se ha demostrado la capacidad de hidrolizar eficazmente enlaces ésteres y carbono-carbono en una misma hidrolasa. Así mismo, se ha sugerido que dicho grado de promiscuidad pueda deberse a la presión selectiva producida por la introducción de nuevas moléculas, en particular contaminantes, en el hábitat de origen, y a procesos evolutivos en los que están involucrados un número muy limitado de aminoácidos.
5. Se sugiere que el potencial biotecnológico de las hidrolasas de enlaces C-C es superior al pensado hasta la fecha en la bibliografía.
6. Se ha demostrado la existencia de enzimas termo-activas y termo-resistentes a 70ºC en ambientes marinos profundos hipersalinos donde la temperatura no es superior a 16.5ºC. Existe una relación directa entre la resistencia a la presión y el aumento en la temperatura óptima y de desnaturalización de las enzimas en ambientes marinos profundos hiper-salinos.
7. Mecanismos moleculares desconocidos están involucrados en la adaptación de las enzimas procedentes de microorganismos de ambientes marinos profundos a condiciones poli-extremas, en particular en ambientes caracterizados por una alta salinidad y presión y una temperatura moderada.
8. Sugerimos que las enzimas aisladas de un mismo ambiente presentan, por lo general, un perfil de sustratos similar, independientemente de las diferencias a nivel filogenético y de homología. Esto podría ser debido a que ambientes con características geoquímicas similares pueden albergar una biodiversidad muy similar independientemente de la localización geográfica. Por lo tanto esta biodiversidad similar podría contener enzimas con reactividades adaptadas a cada nicho ecológico.

CAPÍTULO 8

REFERENCIAS

Adrio JL, Demain AL. (2003). Fungal biotechnology. *Int. Microbiol*.6:191-199.

Akeroyd M, Olsthoorn M, Gerritsma J, Gutker-Vermaas D, Ekkelkamp L, van Rij T, Klaassen P, Plugge W, Smit E, Strupat K, Wenzel T, van Tilborg M, van der Hoeven R. (2013). Searching for microbial protein over-expression in a complex matrix using automated high throughput MS-based proteomics tools. *J. Biotechnol*. 164:112-120.

Akondi KB, Lakshmi VV. (2013). Emerging trends in genomic approaches for microbial bioprospecting. *OMICS* 17:61-70.

Alcaide M, Stogios PJ, Lafraya A, Tchigvintsev A, Flick R, Bargiela R, Chernikova TN, Reva ON, Hai T, Leggewie CC, Katzke N, La Cono V, Matesanz R, Jebbar M, Jaeger KE, Yakimov MM, Yakunin AF, Golyshin PN, Golyshina OV, Savchenko A, Ferrer M. (2014). Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol.* en prensa, doi: 10.1111/1462-2920.12660.

Alcaide M, Tornés J, Stogios PJ, Xu X, Gertler C, Di Leo R, Bargiela R, Lafraya A, Guazzaroni ME, López-Cortés N, Chernikova TN, Golyshina OV, Nechitaylo TY, Plumeier I, Pieper DH, Yakimov MM, Savchenko A, Golyshin PN, Ferrer M. (2013). Single residues dictate the co-evolution of dual esterases - MCP hydrolases from the α/β hydrolase family. *Biochem. J*. 454:157-166.

Alcalde M. (2015). Engineering the ligninolytic enzyme consortium. *Trends Biotechnol*. pii: S0167-7799(14)00256-X.

Alexander E, Stock A, Breiner HW, Behnke A, Bunge J, Yakimov MM, Stoeck T. (2009). Microbial eukaryotes in the hypersaline anoxic L'Atalante deep-sea basin. *Environ. Microbiol*.11:360-381.

Anderson RE, Sogin ML, Baross JA. (2014). Evolutionary Strategies of Viruses, Bacteria and Archaea in Hydrothermal Vent Ecosystems Revealed through Metagenomics. *PLoS One* 9:e109696.

Angelaccio S. (2013). Extremophilic SHMTs: from structure to biotechnology. *Biomed. Res. Int*. 2013:851428.

Anton BP, Chang YC, Brown P, Choi HP, Faller LL, Guleria J, Hu Z, Klitgord N, Levy-Moonshine A, Maksad A, Mazumdar V, McGettrick M, Osmani L, Pokrzywa R, Rachlin J, Swaminathan R, Allen B, Housman G, Monahan C, Rochussen K, Tao K, Bhagwat AS, Brenner SE, Columbus L, de Crécy-Lagard V, Ferguson D, Fomenkov A, Gadda G, Morgan RD, Osterman AL, Rodionov DA, Rodionova IA, Rudd KE, Söll D, Spain J, Xu SY, Bateman A, Blumenthal RM, Bollinger JM, Chang WS, Ferrer M, Friedberg I, Galperin MY, Gobeill J, Haft D, Hunt J, Karp P, Klimke W, Krebs C, Macelis D, Madupu R, Martin MJ, Miller JH, O'Donovan C, Palsson B, Ruch P, Setterdahl A, Sutton G, Tate J, Yakunin A, Tchigvintsev D, Plata G, Hu J, Greiner R, Horn D, Sjölander K, Salzberg SL, Vitkup D, Letovsky S, Segrè D, DeLisi C, Roberts RJ, Steffen M, Kasif S. (2013). The COMBREX project: design, methodology, and initial results. *PLoS Biol*. 11:e1001638.

Arpigny JL, Jaeger KE. (1999). Bacterial lipolytic enzymes: classification and properties. *Biochem. J*. 343:177–183.

Azim A, Sharma SK, Olsen CE, Parmar VS. (2001). Lipase catalysed synthesis of optically enriched α-haloamides. *Bioorg. Med. Chem.* 9:1345-1348.

Barth S, Fischer M, Schmid RD, Pleiss J. (2004). The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics* 20:2845-2847.

Bastard K, Smith AA, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, Mariage A, Boutard M, Debard A, Lechaplais C, Pelle C, Pellouin V, Perchat N, Petit JL, Kreimeyer A, Medigue C, Weissenbach J, Artiguenave F, De Berardinis V, Vallenet D, Salanoubat M. (2014). Revealing the hidden functional diversity of an enzyme family. *Nat. Chem. Biol*.10:42-49.

Baumann M, Stürmer R, Bornscheuer UT (2001). A high-throughput-screening method for the identification of active and enantioselective hydrolases. *Angew. Chem. Int. Ed.* 40:4201–4204.

Beloqui A, Polaina J, Vieites JM, Reyes-Duarte D, Torres R, Golyshina OV, Chernikova TN, Waliczek A, Aharoni A, Yakimov MM, Timmis KN, Golyshin PN, Ferrer M. (2010). Novel hybrid esterase-haloacid dehalogenase enzyme. *Chembiochem* 11:1975-1978.

Blin K, Kazempour D, Wohlleben W, Weber T. (2014). Improved lanthipeptide detection and prediction for antiSMASH. *PLoS One* 9:e89420.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. (2013). antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*. 41:W204-212.

Blomberg R, Kries H, Pinkas DM, Mittl PR, Grütter MG, Privett HK, Mayo SL, Hilvert D. (2013). Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503:418-421.

Bommarius AS, Paye MF. (2013). Stabilizing biocatalysts. *Chem Soc Rev*; 42:6534-6365.

Borin S, Brusetti L, Mapelli F, D'Auria G, Brusa T, Marzorati M, Rizzi A, Yakimov M, Marty D, De Lange GJ, Van der Wielen P, Bolhuis H, McGenity TJ, Polymenakou PN, Malinverno E, Giuliano L, Corselli C, Daffonchio D. (2009). Sulfur cycling and methanogenesis primarily drive microbial colonization of the highly sulfidic Urania deep hypersaline basin. *Proc. Natl. Acad. Sci. USA* 106:9151-9156.

Brustad, E.M.; Arnold, F.H. (2011). Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol*. 15:201-210.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37:D233-D238.

Casanueva A, Tuffin M, Cary C, Cowan DA. (2010). Molecular adaptations to psychrophily: the impact of 'omic' technologies. *Trends Microbiol*. 18:374-381.

Chang C, Sustarich J, Bharadwaj R, Chandrasekaran A, Adams PD, Singh AK. (2013). Droplet-based microfluidic platform for heterogeneous enzymatic assays. *Lab Chip* 13:1817-1822.

Charlaou JL, Donwal JP, Fouquet Y, Jean-Baptiste P, Jean-Luc C, Jean-Pierre D, Yves F, Jean-Baptiste P, Holm N. (2002). Geochemistry of high H_2 and CH_4 vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36 degress 14'N, MAR). *Chem. Geol.* 191:345-359.

Chistoserdova, L. (2014). Is metagenomics resolving identification of functions in microbial communities? *Microb. Biotechnol.*7:1-4.

Choi SL, Rha E, Lee SJ, Kim H, Kwon K, Jeong YS, Rhee YH, Song JJ, Kim HS, Lee SG. (2014). Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. *ACS Synth. Biol*. 3:163-171.

Cita MB. (2006) Exhumation of Messinian evaporites in the Deep-sea and creation of Deep anoxic brine-filled collapsed basins. *Sed. Geol*. 188:357-378.

Copa JL, Solivery J, Caballero A. (2006). Utilización de felúrico esterasa en masas de harina de trigo. P200302033.

Corella JP, El Amrani A, Sigró J, Morellón M, Rico E, Valero-Garcés BL. (2011). Recent evolution of Lake Arreo, northern Spain: influences of land use change and climate. *J. Paleolimnol.* 46:469-485.

Daffonchio D, Borin S, Brusa T, Brusetti L, van der Wielen PW, Bolhuis H, Yakimov MM, D'Auria G, Giuliano L, Marty D, Tamburini C, McGenity TJ, Hallsworth JE, Sass AM, Timmis KN, Tselepides A, de Lange GJ, Hübner A, Thomson J, Varnavas SP, Gasparoni F, Gerber HW, Malinverno E, Corselli C, Garcin J, McKew B, Golyshin PN, Lampadariou N, Polymenakou P, Calore D, Cenedese S, Zanon F, Hoog S; Biodeep Scientific Party. (2006). Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. *Nature* 440:203-207.

Davids T, Schmidt M, Böttcher D, Bornscheuer UT. (2013). Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* 17:215-220.

De Vos, D., Xu, Y., Hulpiau, P., Vergauwen, B., and Van Beeumen, J.J. (2007). Structural investigation of cold activity and regulation of aspartate carbamoyltransferase from the extreme psychrophilic bacterium *Moritella profunda*. *J. Mol. Biol*. 365:379-395.

DeLong EF. (2005). Microbial community genomics in the ocean. *Nat. Rev. Microbiol.* 3:459-469.

Eloe EA, Malfatti F, Gutierrez J, Hardy K, Schmidt WE, Pogliano K, Pogliano J, Azam F, (2011*). Appl. Environ. Microbiol.* 77:8145-8153.

Esteban-Torres, M., Santamaría, L., de las Rivas, B., Muñoz, R. (2014). Characterisation of a cold-active and salt-tolerant esterase from *Lactobacillus plantarum* with potential application during cheese ripening". *Int. Dairy J.* 39:312-315.

Feller G. (2013). Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica* (Cairo) 2013:512840.

Fernández-Arrojo L, Guazzaroni ME, López-Cortés N, Beloqui A, Ferrer M. (2010). Metagenomic era for biocatalyst identification. *Curr. Opin. Biotechnol*. 21:725-733.

Ferrer M, Guazzaroni ME, Richter M, García-Salamanca A, Yarza P, Suárez-Suárez A, Solano J, Alcaide M, van Dillewijn P, Molina-Henares MA, López-Cortés N, Al-Ramahi Y, Guerrero C, Acosta A, de Eugenio LI, Martínez V, Marques S, Rojo F, Santero E, Genilloud O, Pérez-Pérez J, Rosselló-Móra R, Ramos JL. (2011). Taxonomic and functional metagenomic profiling of the microbial community in the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo, Central Spain). *Microb. Ecol.* 62:824-837.

Ferrer M, Werner J, Chernikova TN, Bargiela R, Fernández L, La Cono V, Waldmann J, Teeling H, Golyshina OV, Glöckner FO, Yakimov MM, Golyshin PN; MAMBA Scientific Consortium. (2012). Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study. *Environ. Microbiol.* 14:268-281.

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins Dos Santos VA, Yakimov MM, Timmis KN, Golyshin PN. (2005). Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* 12:895-904.

Filker S, Stock A, Breiner HW, Edgcomb V, Orsi W, Yakimov MM, Stoeck T. (2013). Environmental selection of protistan plankton communities in hypersaline anoxic deep-sea basins, Eastern Mediterranean Sea. *Microb. Open* 2:54-63.

Fojan P, Jonson PH, Petersen MT, Petersen SB. (2000). What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* 82:1033-1041.

Fraaije MW, Scrutton NS. (2013). Special issue: Catalytic mechanisms by biological systems--introduction. *FEBS J*. 280:2947.

Frodie FJ, Heck KL Jr. (2011). Response of coastal fishes to the Gulf of Mexico oil disaster. *PLoS One* 6:e21609.

Furubayashi M, Ikezumi M, Kajiwara J, Iwasaki M, Fujii A, Li L, Saito K, Umeno D. (2014). A high-throughput colorimetric screening assay for terpene synthase activity based on substrate consumption. *PLoS One* 9:e93317.

Glogauer A, Martini VP, Faoro H, Couto GH, Müller-Santos M, Monteiro RA, Mitchell DA, de Souza EM, Pedrosa FO, Krieger N. (2011). Identification and characterization of a new true lipase isolated through metagenomic approach. *Microb. Cell Fact.*10:54-68.

Gomez-Escribano JP, Bibb MJ. (2012). Streptomyces coelicolor as an expression host for heterologous gene clusters. *Methods Enzymol*. 517:279-300.

Gonzalez-Lozano MC, Mendez-Rodriguez LC, Maeda-Martinez AM, Murugan G, Vazquez-Botello A.(2010). Evaluation of toxicity of polluted marine sediments from Bahia Salina Cruz, Mexico. *J Environ. Sci. Health A Tox. Hazard Subst. Environ. Eng.* 45:121-127.

Guazzaroni M-E, Ferrer M. (2011) (Meta) genomics in System biology. Chapter 54. *En*: Handbook of Molecular Microbial Ecology, Volume I: Metagenomics and Complementary Approaches, First Edition (Editor: Frans J. de Bruijn). Wiley-Blackwell, John Wiley & Sons, Inc-2, pp. 475-489.

Guazzaroni ME, Silva-Rocha R, Ward RJ. (2015). Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb. Biotechnol*. 8:52-64

Guazzaroni, M-E, Fernández-Arrojo L, López-Cortés N, Ferrer M. (2011). The role of new (meta-) metabolomic technologies in medical systems microbiology. *En: ebook* "The Role of New Technologies in Medical Microbiological Research and Diagnosis (Editors: John P. Hays and W.B. van Leeuwen). Bentham ebook, The Netherlands (eISBN: 978-1-60805-316-2, 2012).

Guisan JM, Fernández G, Fernández R, Mateo C, Ceínos C, Ramón E. (2002). Procedimiento de hidrólisis enzimática de ésteres del ácido (3SR, 4RS)-4-(4-Flouorofenil)-6-Oxoperidin-3-carboxílico con biocatalizadores de lipasa o esterasas inmovilizadas. WO/09802556.

Guri M, Durand L, Cueff-Gauchard V, Zbinden M, Crassous P, Shillito B, Cambon-Bonavita MA. (2*012).* Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp Rimicaris exoculata. *ISME J*. 6:597-609.

Harrison EH. (2000). Lipases and carboxylesterases: possible roles in the hepatic utilization of vitamin A. *J. Nutr.* 130(2S Suppl):340S-344S.

Harrison JP, Gheeraert N, Tsigelnitskiy D, Cockell CS. (2013). The limits for life under multiple extremes. *Trends Microbiol.* 21:204-212.

Hausmann S, Jaeger K-E. (2010). Lipolytic enzymes from bacteria. *En*: Timmis KN (ed) Handbook of hydrocarbon and lipid microbiology v. 2 (part 4). Springer, Berlin, pp 1099–1126.

Hosokawa M, Hoshino Y, Nishikawa Y, Hirose T, Yoon DH, Mori T, Sekiguchi T, Shoji S, Takeyama H. (2014). Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosens. Bioelectron.* pii:S0956-5663(14)00655-1.

Hügler M, Petersen JM, Dubilier N, Imhoff JF, Sievert SM. (2011). Pathways of carbon and energy metabolism of the epibiotic community associated with the deep-sea hydrothermal vent shrimp Rimicaris exoculata. *PLoS One* 6:e16018.

Jacquiod S, Franqueville L, Cécillon S, Vogel TM, Simonet P. (2013). Soil bacterial community shifts after chitin enrichment an integrative metagenomic approach. *PLoS One* 8:e79699.

Jaeger KE, Eggert T. (2002). Lipases for biotechnology. *Curr. Opin. Biotechnol.* 13, 390-397.

Jan C, Petersen JM, Werner J, Teeling H, Huang S, Glöckner FO, Golyshina OV, Dubilier N, Golyshin PN, Jebbar M, Cambon-Bonavita MA. (2014). The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of Zetaproteobacteria. *Environ. Microbiol*. 16:2723-2738.

Janes LE, Kazlauskas RJ, Quick E. (1997). A Fast Spectrophotometric method to measure the enantioselectivity of hydrolases. *J. Org. Chem*. 62:4560–4561.

Jemli S, Ayadi-Zouari D, Hlima HB, Bejar S. (2014). Biocatalysts: application and engineering for industrial purposes. *Crit Rev Biotechnol*. 6:1-13

Jeon JH, Kim JT, Lee HS, Kim SJ, Kang SG, Choi SH, Lee JH. (2011). Novel lipolytic enzymes identified from metagenomic library of deep-sea sediment. *Evid. Based Complement. Alternat. Med*. 2011:271419.

Karpushova A, Brümmer F, Barth S, Lange S, Schmid RD. (2005). Cloning, recombinant expression and biochemical characterisation of novel esterases from *Bacillus* sp. associated with the marine sponge Aplysina aerophoba. *Appl. Microbiol. Biotechnol.* 67:59-69.

Kim D, Kim SN, Baik KS, Park SC, Lim CH, Kim JO, Shin TS, Oh MJ, Seong CN. (2011). Screening and characterization of a cellulase gene from the gut microflora of abalone using metagenomic library. *J. Microbiol.* 49:141-145.

Kintses B, Hein C, Mohamed MF, Fischlechner M, Courtois F, Leine C, Hollfelder F. (2012). Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.* 19:1001-1009.

Konn C, Testemale D, Querellou J, Holm NG, Charlou JL. (2011). New insight into the contributions of thermogenic processes and biogenic sources to the generation of organic compounds in hydrothermal fluids. *Geobiology* 9:79-93.

Koudelakova T, Chaloupkova R, Brezovsky J, Prokop Z, Sebestova E, Hesseler M, Khabiri M, Plevaka M, Kulik D, Kuta Smatanova I, Rezacova P, Ettrich R, Bornscheuer UT, Damborsky J. (2013). Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chem. Int. Ed. Engl*. 52:1959-1963.

Kourist R, Hari Krishna S, Patel JS, Bartnek F, Hitchman TS, Weiner DP, Bornscheuer UT. (2007). Identification of a metagenome-derived esterase with high enantioselectivity in the kinetic resolution ofarylaliphatic tertiary alcohols. *Org. Biomol. Chem*. 5:3310-3313.

Kube M, Chernikova TN, Al-Ramahi Y, Beloqui A, Lopez-Cortez N, Guazzaroni ME, Heipieper HJ, Klages S, Kotsyurbenko OR, Langer I, Nechitaylo TY, Lünsdorf H, Fernández M, Juárez S, Ciordia S, Singer A, Kagan O, Egorova O, Petit PA, Stogios P, Kim Y, Tchigvintsev A, Flick R, Denaro R, Genovese M, Albar JP, Reva ON, Martínez-Gomariz M, Tran H, Ferrer M, Savchenko A, Yakunin AF, Yakimov MM, Golyshina OV, Reinhardt R, Golyshin PN. (2013). Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nat. Commun*. 4:2156.

Kunze M, Lattermann C, Diederichs S, Kroutil W, Büchs J. (2014). Minireactor-based high-throughput temperature profiling for the optimization of microbial and enzymatic processes. *J. Biol. Eng*. 8:22.

Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J, Colwell RR, Danchin A, Dawyndt P, Dedeurwaerdere T, DeLong EF, Detter JC, De Vos P, Donohue TJ, Dong XZ, Ehrlich DS, Fraser C, Gibbs R, Gilbert J, Gilna P, Glöckner FO, Jansson JK, Keasling JD, Knight R, Labeda D, Lapidus A, Lee JS, Li WJ, Ma J, Markowitz V, Moore ER, Morrison M, Meyer F, Nelson KE, Ohkuma M, Ouzounis CA, Pace N, Parkhill J, Qin N, Rossello-Mora R, Sikorski J, Smith D, Sogin M, Stevens R, Stingl U, Suzuki K, Taylor D, Tiedje JM, Tindall B, Wagner M, Weinstock G, Weissenbach J, White O, Wang J, Zhang L, Zhou YG, Field D, Whitman WB, Garrity GM, Klenk HP. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol*. 12:e1001920.

La Cono V, Smedile F, Bortoluzzi G, Arcadi E, Maimone G, Messina E, Borghini M, Oliveri E, Mazzola S, L'Haridon S, Toffin L, Genovese L, Ferrer M, Giuliano L, Golyshin PN, Yakimov MM. (2011). Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: Prokaryotes and environmental settings. *Environ. Microbiol.* 13:2250-2268.

Laffon B, Aguilera F, Ríos-Vázquez J, Valdiglesias V, Pásaro E.Mutat Res Genet Toxicol Environ Mutagen. (2014). Follow-up study of genotoxic effects in individuals exposed to oil from the tanker Prestige, seven years after the accident. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 760:10-16.

Li, G.; Wang, K.; Liu, Y.H. (2008). Molecular cloning and characterization of a novel pyrethroid hydrolyzing esterase originating from the metagenome. *Microb. Cell Fact.* 7:38-47.

Lim J, Vrignon J, Gruner P, Karamitros CS, Konrad M, Baret J-C. (2013). Ultra-high throughput detection of single cell b-galactosidase activity in droplets using micro-optical lens array. *App. Phys. Lett.* 103:203704.

Liu JY, Bian HP, Tang Y, Bai YP, Xu JH. (2014). Double substituted variant of *Bacillus amyloliquefaciens* esterase with enhanced enantioselectivity and high activity towards 1-(3',4'-methylenedioxyphenyl) ethyl acetate. *Appl. Microbiol. Biotechnol.* en prensa.

Loeschcke A, Markert A, Wilhelm S, Wirtz A, Rosenau F, Jaeger KE, Drepper T. (2013). TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth. Biol.* 2:22-33.

Ma J, Wu L, Guo F, Gu J, Tang X, Jiang L, Liu J, Zhou J, Yu H. (2013). Enhanced enantioselectivity of a carboxyl esterase from *Rhodobacter sphaeroides* by directed evolution. *Appl. Microbiol. Biotechnol.* 97:4897-4906.

Marsh JA, Teichmann SA, Forman-Kay JD. (2012). Probing the diverse landscape of protein flexibility and binding. *Curr. Opin. Struct. Biol.* 22:643-650.

Martín-García F, Mendieta-Moreno JI, Marcos-Alcalde I, Gómez-Puertas P, Mendieta J. (2013) Simulation of catalytic water activation in mitochondrial F1-ATPase using a hybrid quantum mechanics/molecular mechanics approach: An alternative role for ß-Glu 188. *Biochemistry* 52:959-966.

Martinez A, Kolvek SJ, Yip CL, Hopke J, Brown KA, MacNeil IA, Osburne MS. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* 70:2452-2463.

Martínez-Martínez M, Lores I, Peña-García C, Bargiela R, Reyes-Duarte D, Guazzaroni ME, Peláez AI, Sánchez J, Ferrer M. (2014). Biochemical studies on a versatile esterase that is most catalytically active with polyaromatic esters. *Microb. Biotechnol.* 7:184-191.

Martínez-Martínez, M.; Alcaide, M.; Tchigvintsev, A.; Reva, O.; Polaina, J.; Bargiela, R.; Guazzaroni, M.E.; Chicote, A.; Canet, A.; Valero, F.; Rico Eguizabal, E., Guerrero, M.C., Yakunin, A.F., Ferrer, M. (2013). Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol.* 79:3553-3562.

Mate DM, Alcalde M. (2014). Laccase engineering: From rational design to directed evolution. *Biotechnol Adv*. pii: S0734-9750(14)00194-3.

Mate DM, Gonzalez-Perez D, Falk M, Kittl R, Pita M, De Lacey AL, Ludwig R, Shleev S, Alcalde M. (2013). Blood tolerant laccase by directed evolution. *Chem. Biol*. 20:223-231.

Medina I, Alcántara D, González MJ, Torres P, Lucas R, Roque J, Plou FJ, Morales JC. (2010). Antioxidant activity of resveratrol in several fish lipid matrices: effect of acylation and glucosylation. *J. Agric. Food. Chem*. 58:9778-9786.

Mendieta-Moreno JI, Walker R, Lewis J, Gómez-Puertas P, Mendieta J, Ortega J. (2014) FIREBALL/AMBER: An efficient local-orbital DFT QM/MM method for biomolecular systems. *J. Chem. Theor. Comp*. 10:2185–2193.

Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. (2011). How many species are there on Earth and in the ocean? *PLoS Biol.* 9:e1001127.

Muñoz Solano D, Hoyos P, Hernáiz MJ, Alcántara AR, Sánchez-Montero JM. (2012) Industrial biotransformations in the synthesis of building blocks leading to enantiopure drugs. *Bioresour. Technol*. 115:196-207.

Najah M, Mayot E, Mahendra-Wijaya IP, Griffiths AD, Ladame S, Drevelle A. (2013). New glycosidase substrates for droplet-based microfluidic screening. *Anal. Chem*. 85:9807-9814.

Navarro-González I, Sánchez-Ferrer Á, García-Carmona (2012). Molecular characterization of a novel arylesterase from the wine-associated acetic acid bacterium *Gluconobacter oxidans*. *J. Agric. Food Chem.* 60:10789-10795.

Navarro-Gonzalez I, Periago MJ. (2012). Bacterial lipolitic enzymes: properties, structure, technology applications and framework law. *AN.VET. (Murcia)* 28:45-65.

Niehaus F, Gabor E, Wieland S, Siegert P, Maurer KH, Eck J. (2011). Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. *Microb. Biotechnol*. 4:767-776.

Nobili A, Gall MG, Pavlidis IV, Thompson ML, Schmidt M, Bornscheuer UT. (2013). Use of 'small but smart' libraries to enhance the enantioselectivity of an esterase from *Bacillus stearothermophilus* towards tetrahydrofuran-3-yl acetate. *FEBS J.* 280:3084-3093.

Nyyssönen M, Tran HM, Karaoz U, Weihe C, Hadi MZ, Martiny JBH, Martiny AC, Brodie EL. (2013). Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. *Front. Microbiol.* 4:282.

Okamura Y, Kimura T, Yokouchi H, Meneses-Osorio M, Katoh M, Matsunaga T, Takeyama H. (2010). Isolation and characterization of a GDSL esterase from the metagenome of a marine sponge-associated bacteria. *Mar. Biotechnol*. 12:395-402.

Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. (2014). Enzyme Promiscuity: Engine of Evolutionary Innovation. *J. Biol. Chem.* 289:30229-30236.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, Pelletier E, Fink D, Borowski C, Bach W, Dubilier N. (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 476:176-180.

Phelan RM, DiPardo BJ, Townsend CA. (2012). A high-throughput screen for the engineered production of β-lactam antibiotics. *ACS Chem. Biol.* 7:835-840.

Pietra F. (2012). On 3LEZ, a deep-sea halophilic protein with in vitro class-a β-lactamase activity: molecular-dynamics, docking, and reactivity simulations. *Chem. Biodivers.* 9:2659-2684

Plou FJ, Cruces MA, Ferrer M, Fuentes G, Comelles F, Pastor E, Bernabé M, Parra JL, Ballesteros A. (2012). Enzymatic acylation of di- and trisaccharides with fatty acids: choosing the appropriate enzyme, support and solvent. *J. Biotechnol.* 96:55-66.

Plou FJ, Ferrer M, Ballesteros, A. (2003). Transesterification with biological catalysts. *En*: *Encyclopedia of Catalysis*, Ed.: István Horváth. John Wiley & Sons, Inc, New York. Vol 6, Wiley Interscience, New York. Pág. 483-506.

Ponsard J, Cambon-Bonavita MA, Zbinden M, Lepoint G, Joassin A, Corbari L, Shillito B, Durand L, Cueff-Gauchard V, Compere P. (2013). Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *ISME J.* 7:96-109.

Prieur D, Erauso G, Jeanthon C. (1995). Hyperthermophilic life at deep-sea hydrothermal vents. *Planet Space Sci*. 43:115-122.

Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH. (2012). Are uncultivated bacteria really uncultivable? *Microbes Environ*. 27:356-366.

Qureshi SA. (2007). β-Lactamase: an ideal reporter system for monitoring gene expression in live eukaryotic cells. *BioTechniques* 42:91-96.

Reetz MT, Prasad S, Carballeira JD, Gumulya Y, Bocola M. (2010). Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem Soc.*132:9144-9152.

Reetz MT, Zonta A, Schimossek K, Liebeton K, Jaeger KE. (1997). Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution. *Angew. Chem. Int. Ed. Engl*. 36:2830–2832.

Reyes-Duarte D, Ferrer M, García-Arellano H. (2012). Functional-based screening methods for lipases esterases and phospholipases in metagenomic libraries. Georgina Sandoval (ed.) Lipases and Phospholipases: Methods and Protocols Methods in Molecular Biology vol. 861 DOI 10.1007/978-1-61779-600-5_6 © Springer Science+Business Media New York 2012.

Reyes-Duarte, D, Ferrer M, García-Arellano H. (2012). Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods Molecular Biology* 861:101-113.

Reyes-Duarte D, Ferrer M, García-Arellano H. (2012). *In*: Functional-based screening methods for lipases esterases and phospholipases in metagenomic libraries. Georgina Sandoval (ed.) Lipases and Phospholipases: Methods and Protocols Methods in Molecular Biology vol. 861 DOI 10.1007/978-1-61779-600-5_6 © Springer Science+Business Media New York 2012.

Reyes-Duarte D, Ferrer M, García-Arellano H. (2012). Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. In: *Lipases and Phospholipases*; Georgina Sandoval Ed.; Humana Press Springer Protocols: Hatfield; Vol. *861, pp*. 101-113.

Ro HS, Hong HP, Kho BH, Kim S, Chung BH. (2004). Genome-wide cloning and characterization of microbial esterases. *FEMS Microbiol. Lett.* 233:97-105.

Rothschild LJ, Mancinelli RL. (2001). Life in extreme environments. *Nature* 409:1092-1101.

Schallmey M, Koopmeiners J, Wells E, Wardenga R, Schallmey A. (2014). Expanding the halohydrin dehalogenase enzyme family: Identification of novel enzymes by database mining. *Appl. Environ. Microbiol*. pii:AEM.01985-14.

Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM. (2008). Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89:S91-107.

Selvin J, Kennedy J, Lejon DP, Kiran GS, Dobson AD. (2012). Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge *Haliclona simulans. Microb. Cell Fact.* 11:72.

Sharma R, Chisti Y, Banerjee UC. (2001). Production, purification, characterization, and applications of lipases. *Biotechnol Adv.* 19:627-662.

Sheng J, Ji XF, Wang F, Sun M. (2014). Engineering of *Yarrowia lipolytica* lipase Lip8p by circular permutation to alter substrate and temperature characteristics. *J. Ind. Microbiol. Biotechnol.* 41:757-762.

Shin DS, Didonato M, Barondeau DP, Hura GL, Hitomi C, Berglund JA,Getzoff ED, Cary SC, Tainer JA. (2009). Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *J. Mol. Biol*. 385:1534-1555.

Shirai T, Hung VS, Morinaka K, Kobayashi T, Ito S. (2008). Crystal structure of GH13 alpha-glucosidase GSJ from one of the deepest sea bacteria. *Proteins* 73:126-133.

Sim SJ, Baik KS, Park SC, Choe HN, Seong CN, Shin TS, Woo HC, Cho JY, Kim D. (2012). Characterization of alginate lyase gene using a metagenomic library constructed from the gut microflora of abalone. *J. Ind. Microbiol. Biotechnol.* 39:585-593.

Sineva EV, Davydov DR. (2010). Cytochrome P450 from *Photobacterium profundum* SS9, a piezophilic bacterium, exhibits a tightened control of water access to the active site. *Biochemistry* 49:10636-10646.

Singh BK. (2010). Exploring microbial diversity for biotechnology: the way forward. *Trends Biotechnol*. 28:111-116.

Sirim D, Wagner F, Wang L, Schmid RD, Pleiss J. (2011). The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database* 2011:bar006.

Smedile F, Messina E, La Cono V, Tsoy O, Monticelli LS, Borghini M, Giuliano L, Golyshin PN, Mushegian A, Yakimov MM. (2013). Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ. Microbiol.* 15:167-182.

Stock A, Breiner HW, Pachiadaki M, Edgcomb V, Filker S, La Cono V, Yakimov MM, Stoeck T. (2012). Microbial eukaryote life in the new hypersaline deep-sea basin Thetis. *Extremophiles* 16:21-34.

Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson AD, O'Gara F. (2014). Identification of novel phytase genes from an agricultural soil-derived metagenome. *J. Microbiol. Biotechnol*. 24:113-118.

Tchigvintsev A, Tran H, Popovic A, Kovacic F, Brown G, Flick R, Hajighasemi M, Egorova O, Somody JC, Tchigvintsev D, Khusnutdinova A, Chernikova TN, Golyshina OV, Yakimov MM, Savchenko A, Golyshin PN, Jaeger KE, Yakunin AF. (2014). The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Appl. Microbiol. Biotechnol.* en prensa.

Terrón-González L, Medina C, Limón-Mortés MC, Santero E. (2013). Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Sci. Rep.* 3:1107.

Timmis K, de Lorenzo V, Verstraete W, Garcia JL, Ramos JL, Santos H, Economidis I, Nogales B, Timmis JK, Fonseca C, Pruzzo C, Karagouni A, Panopoulos N, Dixon B. (2014). Pipelines for New Chemicals: a strategy to create new value chains and stimulate innovation-based economic revival in Southern European countries. *Environ. Microbiol*. 16:9-18.

Torres P, Poveda A, Jimenez-Barbero J, Ballesteros A, Plou FJ. (2010). Regioselective lipase-catalyzed synthesis of 3-o-acyl derivatives of resveratrol and study of their antioxidant properties. *J. Agric. Food Chem.* 58:807-813.

Torres P, Reyes-Duarte D, Ballesteros A, Plou FJ. (2012). Lipase-catalyzed modification of phenolic antioxidants. *Methods Mol. Biol.* 861:435-43

Turner NJ, Truppo MD. (2013). Biocatalysis enters a new era. *Curr. Opin. Chem. Biol*. 17:212-214.

Valero F. (2012). Heterologous expression systems for lipases: a review. *Methods Mol. Biol*. 861:161-178.

Van Rossum T, Kengen SW, van der Oost J. (2013). Reporter-based screening and selection of enzymes. *FEBS J*. 280:2979-2996.

Velasco-Bucheli R, Del Cerro C, Hormigo D, Acebal C, Arroyo M, García JL, de la Mata I. (2015) Draft genome sequence of *actinoplanes utahensis* nrrl 12052, a microorganism involved in industrial production of pharmaceutical intermediates. *Genome Announc*. 3, pii: e01411-14.

Verastegui Y, Cheng J, Engel K, Kolczynski D, Mortimer S, Lavigne J, Montalibet J, Romantsov T, Hall M, McConkey BJ, Rose DR, Tomashek JJ, Scott BR, Charles TC, Neufeld JD. (2014). Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. *MBio* 5:e01157-14.

Vergne-Vaxelaire C, Bordier F, Fossey A, Besnard-Gonnet M, Debard A, Mariage A, Pellouin V, Perret A, Petit JL, Stam M, Salanoubat M, Weissenbach J, De Berardinis V, Zaparucha A. (2013).

Nitrilase activity screening on structurally diverse substrates: providing biocatalytic tools for organic synthesis. *Adv. Synth. Catal.* 355:1763-1777.

Vester JK, Glaring MA, Stougaard P. (2014). Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. *Microb. Cell Fact*. 13:72.

Vieites JM, Guazzaroni ME, Beloqui A, Golyshin PN, Ferrer M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* 33:236-255.

Wahl M, Goecke F, Labes A, Dobretsov S, Weinberger F. (2012). The second skin: ecological role of epibiotic biofilms on marine organisms. *Front. Microbiol*. 3:292.

Wallmann K, Suess E, Winckler G, Cita M, Westbrook G. (1997) Salty brines on the Mediterranean sea floor. *Nature* 387:31-32

Westbrook GK, Reston TJ. (2002) The accretionary complex of the Mediterranean Ridge: tectonics, fluid flow and the formation of brine lakes-an introduction to the special issue of Marine Geology. *Mar. Geol*. 186:1-8.

Woodley JM. (2013). Protein engineering of enzymes for process applications. *Curr. Opin. Chem. Biol.* 17:310-316.

Wu, Q.; Soni, P.; Reetz, M.T. (2013). Laboratory evolution of enantiocomplementary *Candida antarctica* lipase B mutants with broad substrate scope. *J. Am. Chem. Soc.* 135:1872-1881.

Yakimov MM, La Cono V, Slepak VZ, La Spada G, Arcadi E, Messina E, Borghini M, Monticelli LS, Rojo D, Barbas C, Golyshina OV, Ferrer M, Golyshin PN, Giuliano L. (2013). Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci. Rep.* 3:3554.

Yakimov MM, Timmis KN, Golyshin PN. *Curr Opin Biotechnol*. (2007). Obligate oil-degrading marine bacteria. *Curr. Opin. Biotechnol.*18:257-266.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12:635-645.

Yoon S, Kim S, Park S, Hong E, Kim J, Kim S, Yoo TH, Ryu Y. (2014). Improving the enantioselectivity of an esterase toward (S)-ketoprofen ethyl ester through protein engineering. *J. Mol. Catal. B-Enzym.* 100:25-31.

Zbinden M, Shillito B, Le Bris N, de Villardi de Montlaur C, Roussel E, Guyot F, Gaill F, Cambon-Bonavita M-A. (2008). New insights on the metabolic diversity among the epibiotic microbial community of the hydrothermal shrimp *Rimicaris exoculata*. *J. Experim. Mar. Biol. Ecol*. 359:131–140.