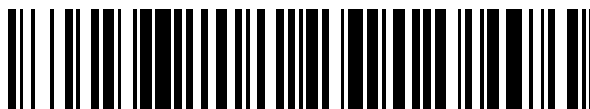


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 832 451**

51 Int. Cl.:

C12N 9/14 (2006.01)

C12N 9/88 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.07.2014** **E 18182700 (7)**

97 Fecha y número de publicación de la concesión europea: **02.09.2020** **EP 3404097**

54 Título: **Expansión de la familia de enzimas de halohidrina deshalogenasa: identificación de nuevas enzimas por extracción de base de datos**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.06.2021

73 Titular/es:

ENZYMICALS AG (100.0%)
Walther-Rathenau-Strasse 49a
17489 Greifswald, DE

72 Inventor/es:

SCHALLMEY, ANETT;
SCHALLMEY, MARCUS;
WESSEL, JULIA;
MENYES, ULF;
WARDENGA, RAINER;
BRUNDIEK, HENRIKE y
BORCHERT, SONJA

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 832 451 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Expansión de la familia de enzimas de halohidrina deshalogenasa: identificación de nuevas enzimas por extracción de base de datos

Las halohidrina deshalogenasas (también denominadas haloalcohol deshalogenasas, haloalcohol/halohidrina epoxidadas o haluro de hidrógeno liasas; EC 4.5.1.-) (HHDH) son enzimas biotecnológicamente relevantes que catalizan la deshalogenación reversible de β -haloalcoholes en la formación de epóxido (1- 3). Además de ser útiles para la producción de haloalcoholes enantiopuros (4-6) y epóxidos (5, 7-9), estas enzimas también pueden ser aplicadas en la formación de nuevos enlaces carbono-carbono, carbono-nitrógeno o carbono-oxígeno. Esto se debe a su promiscua actividad de apertura de anillo epóxido que acepta también otros nucleófilos (por ejemplo, cianuro, azida o nitrito) además de los haluros, lo que produce una amplia variedad de productos (10). Los ejemplos de aplicaciones relevantes de HHDH son la producción de precursores químicos finos C3 o C4 ópticamente puros (5, 8, 11, 12), que incluyen la producción a escala de varias toneladas de ésteres de (R)-4-ciano-3-hidroxitirato enantiopuros para fármacos de estatinas (13) o la producción de alcoholes terciarios quirales (14-17) para los que la síntesis orgánica convencional es bastante compleja.

A pesar de su potencial indicado como biocatalizadores, sólo se hallan disponibles unas pocas HHDH. Aunque ha sido informado que numerosas especies bacterianas poseen actividad HHDH, sólo han sido clonados y expresados de forma recombinante seis genes HHDH, a saber, *hheA* de *Corynebacterium* sp. cepa N-1074 (18), *hheA2* de *Arthrobacter* sp. cepa AD2 (19), *hheB* de *Corynebacterium* sp. cepa N-1074 (18), *hheB2* de *Mycobacterium* sp. cepa GP1 (19), y dos secuencias *hheC* idénticas de *Agrobacterium radiobacter* AD1 (19) y *Rhizobium* sp. cepa NHG3 (20).

Todas las HHDH clonadas pertenecen a la superfamilia de deshidrogenasa/reductasa de cadena corta (SDR) y exhiben varias características importantes de esta diversa clase de enzimas (19, 21, 22). Por ejemplo, todas las HHDH conocidas hacen uso de una tríada catalítica de Ser-Tyr-Arg, comparten el ensamblaje cuaternario homomultimérico comúnmente encontrado y ambas HHDH cristalizadas, a saber, HheA2 (23) y HheC (24), poseen una estructura terciaria similar a otras de proteínas con plegamiento de Rossmann. En contraste con un sitio de unión al cofactor presente en las enzimas SDR, en las estructuras de HheA2 y HheC está presente una bolsa de unión de aniones espaciosa (23, 24). En consecuencia, todas las secuencias de HHDH conocidas forman solo una fracción diminuta pero bien definida dentro de la superfamilia SDR.

Aunque estos pocos HHDH conocidos ya han dado lugar a numerosas aplicaciones interesantes, es de gran interés aumentar el número de enzimas HHDH funcionalmente diversas. Claramente, las secuencias novedosas serían una valiosa adición a la diversidad funcional de la caja de herramientas de la enzima HHDH. Las bases de datos de secuencias públicas contienen un inmenso tesoro de secuencias con información solo limitada o sin información sobre la función de estas secuencias o sus biomoléculas relacionadas. Schallmey *et al.* describen la identificación de nuevas halohidrina deshalogenasas usando la minería de datos de bases de datos de secuencias disponibles públicamente usando información de secuencia específica de HHDH (A).

El objeto de la presente invención es el uso de enzimas recientemente identificadas como HHDH entre las secuencias ya depositadas.

El problema es resuelto mediante el uso de una biomolécula que tiene la secuencia de proteínas o secuencia de ADN codificante de HheD3 como halohidrina deshalogenasa de acuerdo con la reivindicación 1.

a) Un procedimiento para la identificación de enzimas que tienen potencial actividad de halohidrina deshalogenasa a partir de las secuencias descritas en la base de datos de biomoléculas – que no está cubierta en la presente invención – comprende las siguientes etapas: proporcionar al menos una base de datos de biomoléculas,

b) realizar una búsqueda de la base de datos de biomoléculas para identificar al menos una secuencia de biomolécula diana, que comprende, en el nivel de aminoácido, ambos de los siguientes motivos de la secuencia de aminoácidos (1) y (2) cualquier lugar dentro de la secuencia de biomolécula diana

1) siendo el aminoácido Treonina (T) seguido por cuatro aminoácidos variables (x(4)), seguido por aminoácido de fenilalanina (F) o tirosina (Y) seguida por un aminoácido variable (x), seguido por aminoácido Glicina (G):

"T-x(4)-[FY]-x-G"

y

2) el aminoácido Serina (S) seguido por 12 aminoácidos variables (x(12)), seguido por aminoácido Tirosina (Y), seguido por tres aminoácidos variables (x(3)), seguido por el aminoácido Arginina (R):

"S-x(12)-Y-x(3)-R",

en el que en (1) y (2) aminoácidos variables (x) pueden ser independientemente los mismos o diferentes y son seleccionados de aminoácidos naturales y sintéticos.

Las secuencias siempre son indicadas desde el extremo N- a C-terminal para las proteínas o en el caso de ADN desde el extremo terminal 5' a 3'. La indicación de aminoácidos entre corchetes [] significa que los aminoácidos entre corchetes son intercambiables. Por ejemplo, "[FY]" significa que F puede ser reemplazada por Y. Las palabras "motivo" y "patrón de secuencia" se usan como sinónimos. La abreviatura de una letra así como la abreviatura de tres letras de los aminoácidos cumple con el estándar.

Una base de datos de biomoléculas es una fuente de biomoléculas en sí misma o una colección de datos de secuencia de biomoléculas, en particular datos de secuencia de polinucleótidos o polipéptidos. En el contexto de la presente invención, una biomolécula es preferentemente una molécula de polinucleótido que porta información genética, en particular una molécula de ADN. En otra realización preferente de la presente invención, una biomolécula es un polipéptido, en particular una proteína, que comprende varios aminoácidos (proteína ≥ 100 aminoácidos). Por tanto, una base de datos de biomoléculas puede ser una fuente física de biomoléculas, en particular puede ser un banco de genes, en particular una biblioteca de ADNc o genómica, o es una colección de información sobre dichas biomoléculas, en particular una colección de datos de secuencia, en particular secuencias de aminoácidos o secuencias de polinucleótidos, en particular secuencias de ADN.

Un motivo adicional (3) puede estar presente en cualquier parte de la secuencia de biomoléculas diana, en el que el aminoácido Glicina (G) es seguido de seis aminoácidos variables (x (6)), seguido de dos Glicinas (G)

"G-x(6)-G-G"

Especialmente, el procedimiento para la identificación de enzimas que tienen actividad potencial de halohidrina deshalogenasa a partir de secuencias descritas en una base de datos de biomoléculas comprende una etapa adicional de

- c) proporcionar la al menos una biomolécula que tiene la secuencia de biomolécula diana identificada en la etapa b) y ser o codificar al menos parcialmente una proteína con la actividad de halohidrina deshalogenasa y, opcionalmente, una etapa adicional
- d) comprobar/verificar la actividad de halohidrina deshalogenasa de la al menos una biomolécula que tiene la secuencia de biomolécula diana mediante la conversión de halohidrininas vecinas en epóxidos o mediante la apertura del anillo de epóxidos con nucleófilos.

"Comprobar/verificar la actividad de halohidrina deshalogenasa" significa que la biomolécula diana convierte al menos parcialmente el 1,3-dihalo-2-propanol (halo = cloro, bromo, yodo, preferentemente 1,3-dicloro-2-propanol) como estándar halohidrina vecinal en la correspondiente epihalohidrina (halo = cloro, bromo, yodo, preferentemente epiclorhidrina) y/o convierte al menos parcialmente epihalohidrina (halo = cloro, bromo, yodo, preferentemente epiclorhidrina) o glicidol como estándares de epóxidos con iones haluro (Cl⁻, I⁻, Br⁻, preferentemente Cl⁻) o ion azida N⁻ como nucleófilos estándares.

El término "una biomolécula siendo o codificando al menos parcialmente una proteína" significa preferentemente que la biomolécula puede ser una proteína con la actividad de una halohidrina deshalogenasa o, en caso de que la biomolécula sea una secuencia de nucleótidos, al menos una parte de dicha secuencia de nucleótidos codifica dicha proteína, preferentemente una proteína de longitud completa. En consecuencia, en caso de que la biomolécula proporcionada sea una molécula de la secuencia de nucleótidos, al menos una parte de dicha molécula de la secuencia de nucleótidos codifica una proteína con la actividad de una halohidrina deshalogenasa, por lo que posiblemente una parte adicional de dicha molécula de nucleótidos tiene alguna otra función, por ejemplo función reguladora o de replicación. Por lo tanto, el término "proporcionar una biomolécula que tiene la secuencia de biomolécula diana identificada en la etapa b) y que es o codifica al menos parcialmente una proteína con la actividad de una halohidrina deshalogenasa" es preferentemente equivalente al término "proporcionar una biomolécula que tiene la secuencia de biomolécula diana identificada en la etapa b) que puede ser una proteína con actividad de halohidrina deshalogenasa o que puede ser una molécula de la secuencia de nucleótidos que incluye una secuencia que codifica dicha proteína siempre que la proteína tenga la actividad de una halohidrina deshalogenasa".

Con respecto a la identificación, es posible que el motivo (1) y el motivo (2) estén espaciados entre sí por 1 a 250 aminoácidos variables (x(1,250))

"T-x(4)-[FY]-x-G-x(1,250)-S-x(12)-Y-x(3)-R",

más preferentemente por 50 a 200 aminoácidos variables (x(50,200)), de máxima preferencia por 93 a 131 aminoácidos variables (x(93,131)), en el que los aminoácidos variables (x) pueden ser los mismos o diferentes y son seleccionados de los aminoácidos naturales y sintéticos.

Como se mencionó anteriormente, los motivos (1), (2) se suceden entre sí desde el extremo N-terminal al extremo C-terminal o, en el caso de ADN, desde el extremo 5' al 3'.

Si (3) está presente, preferentemente está espaciado del motivo (2) por 1 a 250 aminoácidos variables (x(1,250))

S-x(12)-Y-x(3)-R-x(1,250)-G-x(6)-G-G"

- 5 más preferentemente por 50 a 200 aminoácidos variables (x(50,200)), con máxima preferencia por 76 a 84 aminoácidos variables (x(76,84)), en que los aminoácidos variables (x) pueden ser iguales o diferentes y se seleccionan de aminoácidos naturales y sintéticos.

Con respecto a la identificación, es posible que los motivos (1), (2), (3) se sucedan entre sí desde el extremo N-terminal al extremo C-terminal o, en el caso de ADN, desde el extremo 5' al 3':

- 10 "T-x(4)-[FY]-x-G-x(1,250)-S-x(12)-Y-x(3)-R-x(1,250)-G-x(6)-G-G",

en el que las variantes preferentes de los aminoácidos espaciadores x (1250) entre los motivos (1) y (2) y entre (2) y (3) son los que se definieron anteriormente. El motivo (3) es de uso especial para asegurar que la enzima sea funcional, es decir, que la secuencia esté completa en el extremo C-terminal. Sin estar sujeto a esta teoría: el motivo (3) podría ser esencial para la formación de oligómeros.

- 15 Además, la presencia del motivo (1) también asegura que la enzima sea funcional, es decir, que la secuencia esté completa en el extremo N-terminal. Es posible que el motivo (3) es seguido de 0-40 aminoácidos, preferentemente de 1 a 30 aminoácidos. Además, de 0 a 40 aminoácidos, más preferentemente de 1 a 30 aminoácidos, pueden preceder al motivo (1).

- 20 Las biomoléculas son proteínas o moléculas de ADN y las secuencias de biomoléculas son secuencias de aminoácidos o secuencias de ADN. Se prefieren las proteínas y, por tanto, las secuencias de aminoácidos.

Con respecto a la identificación, es posible que la etapa a) consista en dos etapas a1) y a2) de

a1) proporcionar la al menos una base de datos de biomoléculas (por ejemplo, GenBank, UniProt, PFAM, Prosite)

- 25 a2) proporcionar al menos una secuencia de la biomolécula incógnita de al menos una halohidrina deshalogenasa en que las etapas a1) y a2) son de orden intercambiable y

la etapa b) consiste en dos etapas sucesivas b1) y b2) de,

b1) realizar una búsqueda de la base de datos de biomoléculas con la al menos una secuencia de la biomolécula incógnita para identificar un grupo de secuencias de biomolécula pre-diana,

- 30 b2) seleccionar del grupo de secuencias de biomoléculas pre-diana al menos una secuencia de biomoléculas diana, que comprende, a nivel de aminoácidos, ambos motivos (1) y (2) en cualquier lugar dentro de la al menos una secuencia de biomoléculas diana y opcionalmente comprende el motivo (3) en cualquier lugar dentro de la al menos una secuencia de biomoléculas diana.

Preferentemente, todos estos motivos (1), (2) y (3) deben estar presentes.

- 35 El procedimiento de identificación puede ser aplicable en sus etapas a) a b) a una base de datos de biomoléculas que es una colección de datos de secuencia de secuencias de polinucleótidos, en particular secuencias de ADN o secuencias de aminoácidos, las cuales se buscan en una etapa b1) con herramientas de alineamiento de secuencias. Esto significa especialmente BLAST (Herramienta básica de búsqueda de alineamiento local, BLASTp (p = proteína) y/o BLASTn (n = nucleótido)). Otros programas/herramientas adecuados incluyen alineamiento de secuencias por pares y programas de alineamiento de secuencias múltiples tales como MAFFT, GAP, BESTFIT, MUSCLE, CLUSTALW (y sucesores), T-COFFEE (y sabores), PRANK y FASTA.

- 40 En el caso de que la secuencia de biomolécula incógnita sea una secuencia de ADN, dicha secuencia es preferentemente un marco de lectura abierto (ORF) de la secuencia codificante de biomolécula incógnita o una parte característica de la misma. En el caso de que la secuencia de biomolécula incógnita sea una secuencia de proteína, dicha secuencia es preferentemente la secuencia de aminoácidos ORF en sí misma o una parte característica de la misma.

- 45 Cabe señalar que la variante especial definida anteriormente del procedimiento de identificación está relacionada con el uso de una secuencia de biomoléculas incógnita como punto de partida para la búsqueda dentro de una base de datos de biomoléculas. Sin embargo, son posibles variantes alternativas del procedimiento de identificación cuando se busca en un banco de biomoléculas simplemente sobre la base de los motivos (1) y (2) y si es necesario (3). Estas búsquedas de solo motivos están disponibles, por ejemplo, a través de ScanProsite y MAST (que incluye el conjunto MEME).

El grupo de biomoléculas pre-diana mencionado en las etapas b1), b2) también puede consistir en una sola biomolécula, es decir, la biomolécula diana. Si solo hay una biomolécula pre-diana, no es necesaria la selección de un grupo.

A continuación, la denominación de genes y enzimas de la familia de enzimas HDDH (Hhe) se realiza sobre la base de su agrupación en cualquiera de los subtipos filogenéticos A a G (HheA-HheG) y, si es necesario, subtipos filogenéticos adicionales. Por lo tanto, la numeración de las enzimas dentro de cada subtipo se realizó de manera consecutiva de acuerdo con el tiempo de presentación en cualquiera de las bases de datos públicas.

Debido a que la búsqueda en bases de datos de biomoléculas también da como resultado automáticamente el hallazgo de secuencias ya indicadas como HDDH, las secuencias de HheA, HheA2, HheB, HheB2, HheC (grupo 1a) se excluyen como secuencias de biomoléculas pre-diana o secuencias de biomoléculas diana. HheA, HheA2, HheB, HheB2, HheC son HDDH conocidas y su actividad como HDDH está comprobada. Además, HheA4 tiene una identidad de secuencia en el nivel de aminoácidos del 99% con HheA, incluso si no hay pruebas de actividad de HDDH o denominación como HDDH en la bibliografía para HheA4. Por tanto, una variante especial excluye HheA, HheA2, HheA4, HheB, HheB2, HheC (grupo 2a) como secuencias de biomoléculas pre-diana o secuencias de biomoléculas diana. HheA5 simplemente se nombra como HDDH en las bases de datos de biomoléculas sin una prueba confiable de la actividad de HDDH real. HheA3 pertenece a un grupo anotado automáticamente de secuencias similares a HDDH (CDD, base de datos de dominio conservado) sin prueba confiable de actividad de HDDH real. Por lo tanto, una variante más especial excluye a HheA, HheA2, HheA3, HheA4, HheA5, HheB, HheB2, HheC (grupo 3a) como secuencias de biomoléculas pre-diana o secuencias de biomoléculas diana. La variante más especial además excluye HheGx y HheBx, es decir, en general excluye HheA, HheA2, HheA3, HheA4, HheA5, HheB, HheB2, HheBx, HheC, HheGx (grupo 4a) como secuencias de biomoléculas pre-diana o secuencias de biomoléculas objetivo, porque HheGx tiene una M en lugar de una T en el motivo (1) y la secuencia anotada ABC25271 (GenBank), que corresponde parcialmente a la secuencia HheBx, carece del motivo (1) (que lleva a una enzima sin actividad HDDH) pero la última es simplemente nombrada como HDDH en bases de datos de biomoléculas sin prueba confiable de actividad de HDDH real.

Una posible variante excluye al menos una secuencia de biomoléculas incógnita y/o como secuencia de biomoléculas pre-diana y/o secuencias de secuencia de biomoléculas diana que tienen menos o que codifican menos de o iguales a 186 aminoácidos. En otras palabras, la al menos una secuencia de biomoléculas incógnita y/o secuencia de biomoléculas pre-diana y/o secuencia de biomoléculas diana debe tener más de 186 aminoácidos o tiene que codificar más de 186 aminoácidos.

Otra variante posible excluye la al menos una secuencia de biomolécula incógnita y/o como secuencia de biomolécula pre-diana y/o secuencias de secuencia de biomolécula diana que tienen menos o que codifican menos de o igual a 201 aminoácidos. En otras palabras, la al menos una secuencia de biomoléculas incógnita y/o la secuencia de biomoléculas pre-diana y/o la secuencia de biomoléculas diana tiene que tener más de 201 aminoácidos o tiene que codificar más de 201 aminoácidos.

En un aspecto posible del procedimiento de identificación, en el que es usada la al menos una secuencia de la biomolécula incógnita de al menos un halohidrina deshalogenasa, es decir, son usadas las etapas a1) y a2), la al menos una secuencia de la biomolécula incógnita de al menos una halohidrina deshalogenasa es seleccionado de las secuencias de proteína o secuencias de ADN codificantes de HheA, HheA2, HheA3, HheA4, HheA5, HheB, HheB2, HheBx, HheC, HheGx (grupo 4a), preferentemente de las secuencias de proteína o secuencias de ADN codificantes de HheA, HheA2, HheA3, HheA4, HheA5, HheB, HheB2, HheC (grupo 3a), más preferentemente de las secuencias de proteína o secuencias de ADN codificantes de HheA, HheA2, HheA4, HheB, HheB2, HheC (grupo 2a), de máxima preferencia de las secuencias de proteína o secuencias de ADN codificantes de HheA, HheA2, HheB, HheB2, HheC (grupo 1a).

En aún otro aspecto posible del procedimiento de identificación, en el que al menos una secuencia de la biomolécula incógnita de al menos se usa una halohidrina deshalogenasa, es decir, son usadas las etapas a1) y a2), la al menos una secuencia de la biomolécula incógnita de al menos una halohidrina deshalogenasa es seleccionada de secuencias de proteína o secuencia de ADN codificantes de HheA3, HheA4, HheA5, HheA6, HheB3, HheB4, HheB5, HheB6, HheB7, HheBx, HheD, HheD2, HheD3, HheD4, HheD5, HheD6, HheD7, HheD8, HheD9, HheD10, HheD11, HheD12, HheD13, HheD14, HheD15, HheD16, HheD17, HheE, HheE2, HheE3, HheE4, HheE5, HheEx, HheEx2, HheEx3, HheF, HheG, HheGx (grupo 1b), preferentemente de las secuencias de proteína o secuencia de ADN codificantes de HheA3, HheA5, HheA6, HheB3, HheB4, HheB5, HheB6, HheB7, HheBx, HheD, HheD2, HheD3, HheD4, HheD5, HheD6, HheD7, HheD8, HheD9, HheD10, HheD11, HheD12, HheD13, HheD14, HheD15, HheD16, HheD17, HheE, HheE2, HheE3, HheE4, HheE5, HheEx, HheEx2, HheEx3, HheF, HheG, HheGx (grupo 2b), más preferentemente de secuencias de proteína o secuencia de ADN codificantes de HheA6, HheB3, HheB4, HheB5, HheB6, HheB7, HheBx, HheD, HheD2, HheD3, HheD4, HheD5, HheD6, HheD7, HheD8, HheD9, HheD10, HheD11, HheD12, HheD13, HheD14, HheD15, HheD16, HheD17, HheE, HheE2, HheE3, HheE4, HheE5, HheEx, HheEx2, HheEx3, HheF, HheG (grupo 4b).

Preferentemente, las secuencias de biomolécula pre-diana tienen un grado de identidad de secuencia de al menos 15%, preferentemente 20% a la al menos una secuencia de la biomolécula incógnita, calculada en el nivel de aminoácidos. También es preferente cuando la secuencias de biomolécula pre-diana tienen un grado de identidad de secuencia de como máximo 80%, preferentemente de como máximo 70%, más preferentemente de como máximo 60% con la al menos una secuencia de la biomolécula incógnita, calculada en el nivel de aminoácidos. Especialmente, si las secuencias excluidas como biomoléculas pre-diana o biomoléculas diana son usadas como biomolécula diana, todas las secuencias con más de 80% de identidad de secuencias con estas biomoléculas diana, calculada en el nivel de aminoácidos, se excluyen como biomoléculas pre-diana o biomoléculas diana.

En aún otro aspecto posible del procedimiento de identificación son excluidas secuencias de biomolécula pre-diana y diana que tienen más o que codifican más de 350 aminoácidos, más preferentemente que tienen más o codifican más de 280 aminoácidos, más preferentemente que tienen más de o codifican más de 262 aminoácidos.

Cabe señalar que los aminoácidos en cada uno de los motivos (1)-(3) se pueden reemplazar sin perder la función HDDH. Esto se aplica, por ejemplo, si el aminoácido de reemplazo tiene una estructura similar, propiedades similares, etc., que el aminoácido reemplazado.

Con respecto a los reemplazos de aminoácidos en el motivo (1), es posible que

- el aminoácido treonina (T) pueda ser reemplazado con el aminoácido Metionina (M) o con aminoácido Serina (S) y/o
- el aminoácido fenilalanina (F) o tirosina (Y) puedan ser reemplazados con un aminoácido seleccionado de Isoleucina (I) o Leucina (L)

y/o

- el aminoácido glicina (G) pueda ser reemplazado con aminoácido Serina (S).

El motivo resultante (1) con el reemplazo es en consecuencia:

"[TSM]-x(4)-[FYIL]-x-[GS]".

Sin embargo, es preferente el uso del motivo (1) sin reemplazos de aminoácidos.

Con respecto a los reemplazos de aminoácidos en el motivo (2), también es posible que

- el aminoácido Tirosina (Y) pueda ser reemplazado con el aminoácido Prolina (P)

El motivo resultante (2) con reemplazos en consecuencia es:

"S-x(12)-[YP]-x(3)-[R]".

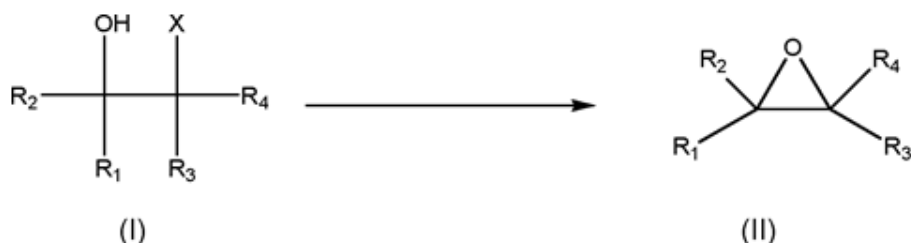
Sin embargo, es preferente el uso del motivo (2) sin reemplazos de aminoácidos.

Uso de secuencias de biomolécula diana como HDDH

Las secuencias de biomoléculas diana identificadas pueden ser usadas como HDDH en la conversión de halohidrinas en epóxidos y/o en la apertura del anillo de epóxidos en alcoholes β -sustituidos. De acuerdo con la invención, una biomolécula que tiene la secuencia de proteína o la secuencia de ADN codificante de HheD3 que tiene NCBI, número de acceso de GenBank ABM93639 se usa como halohidrina deshalogenasa.

Conversión de halohidrinas a epóxidos

En detalle, la biomolécula que tiene la secuencia de proteínas o secuencia de ADN codificante de HHeD3 que tiene NCBI, número de acceso GenBank ABM93639 es usada como halohidrina deshalogenasa en la conversión de al menos una parte de la cantidad de una halohidrina de fórmula (I) en un epóxido de fórmula (II)



en el que

X es un átomo de halógeno seleccionado de Cl, Br, I;

R₁, R₂, R₃, R₄ son independientemente hidrógeno, halógeno (F, Cl, Br, I) o alquilo C1-C12, C2-C12-alquenilo, alquinilo C2-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o arilo C5-C12 sustituido o no sustituido, o -alquil C1-C3-C(=O)-O-R₅ o -alquil C1-C3-O-R₅, con

R₅ seleccionado de hidrógeno, alquilo C1-C6, alquenilo C2-C6, alquinilo C2-C6, alcoxi C1-C6, alquilarilo C5-C13, arilalquilo C5-C13, ariloxi C5-C13, arilalcoxi C5-C13, heteroalquilo C5-C13, arilo C5-C13, heteroarilo C5-C13, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o heteroarilo cíclico C5-C13, en el que el caso de más de un sistema anular alifático o aromático, los sistemas anulares están condensados o separados y R₅ puede tener al menos un sustituyente adicional R₆ con

R₆ seleccionado de hidrógeno, hidroxilo, -C(=O)-NH₂, halógeno (F, Cl, Br, I), alquilo-C1-C3, -C(=O)-alquilo C1-C3-, -C(=O)-alquilo-C1-C3, -NH-C(=O)-alquilo-C1-C3, -alquil C1-C3-O-alquilo-C1-C3, alquenilo C2-C3, -O-alquenilo C2-C3, cicloalquilo C5-C7, heteroalquilo cíclico C5-C7, arilo C5-C7 o heteroarilo C5-C7, preferentemente seleccionado de hidrógeno y metilo,

y/o

dos de R₁, R₂, R₃, R₄ forman en conjunto una estructura cíclica seleccionada de cicloalquilo C3-C12, cicloalquenilo C4-C12, cicloalquinilo C6-C12, heteroalquilo cíclico C3-C12, heteroalquenilo cíclico C4-C12 o heteroalquinilo cíclico C6-C12, preferentemente un grupo cicloalquilo C4-C8, o cicloalquenilo C4-C8, más preferentemente un grupo cicloalquilo C5-C7, o cicloalquenilo C5-C7 grupo, con máxima preferencia un grupo cicloalquilo C5-C6, o cicloalquenilo C5-C6 en el que en el caso de más de un sistema anular, los sistemas anulares están condensados o separados, y en el que el número de átomos de C indicado para la estructura cíclica siempre incluye los 2 átomos de C de la estructura de halohidrina que portan el OH y el X, y

en el que cada grupo cíclico en R₁, R₂, R₃, R₄ o estructura cíclica formada por dos de R₁, R₂, R₃, R₄ opcionalmente tiene al menos un sustituyente adicional diferente de hidrógeno, preferentemente seleccionado de R₆, con máxima preferencia halógeno (F, Cl, Br, I) o alquilo C1-C3. Este opcional al menos un sustituyente adicional también se puede denominar como sustituyente R₇, que es especialmente (R₇)_n con n que es cero o un número entero entre 1 y 4, preferentemente n = 0, 1 o 2.

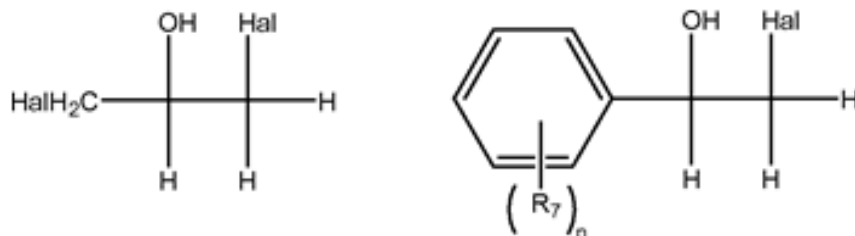
Los términos alqueno/alquenilo, cicloalqueno/cicloalquenilo se refieren a hidrocarburos que contienen al menos un doble enlace carbono-carbono. Los términos alquino/alquinilo, cicloalquino/cicloalquinilo se refieren a hidrocarburos que contienen al menos un triple enlace carbono-carbono.

Con respecto a R₁, R₂, R₃, R₄: cada uno de R₁, R₂, R₃, R₄, que no forma una estructura cíclica con otro de R₁, R₂, R₃, R₄, es decir, cada uno de estos residuos R₁, R₂, R₃, R₄, que no está unido covalentemente a otro de estos residuos R₁, R₂, R₃, R₄, puede tener al menos un átomo de carbono reemplazado por un heteroátomo. En otras palabras: cada R₁, R₂, R₃, R₄ es un hidrocarburo, es decir, alquilo C1-C12, alquenilo C2-C12, alquinilo C2-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o arilo C5-C12 sustituido o no sustituido pueden tener al menos un átomo de carbono reemplazado por un heteroátomo.

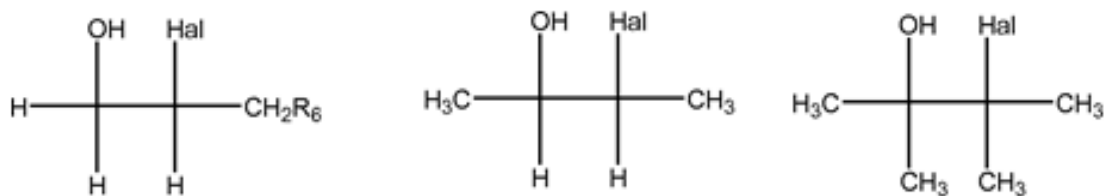
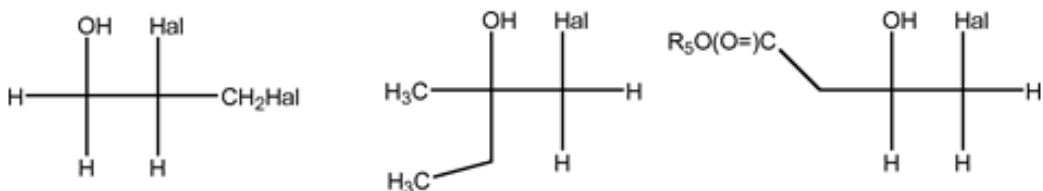
En general, es decir, con respecto a cualquiera de los residuos R₁-R₇: la expresión "hetero" o "heteroátomo" significa preferentemente un heteroátomo seleccionado de nitrógeno (N), azufre (S), oxígeno (O), selenio (Se) o fósforo (P), más preferentemente un heteroátomo seleccionado de N, S u O.

Con respecto a todas las sustancias de fórmula (I), los átomos de C que portan el OH y X siempre tienen hibridación sp³. Simultáneamente, los átomos de C que forman el anillo epóxido en la fórmula (II) también siempre tienen hibridación sp³.

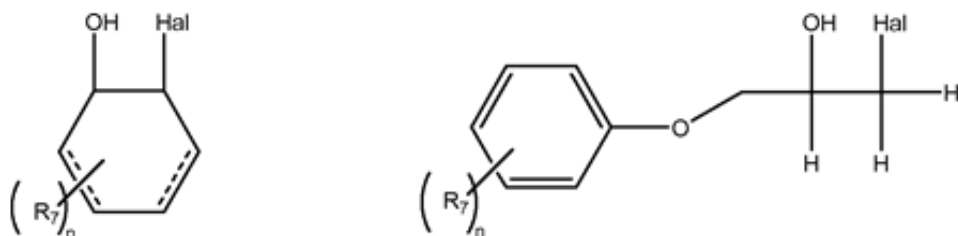
La halohidrina de fórmula (I) es seleccionada preferentemente de:



5



10



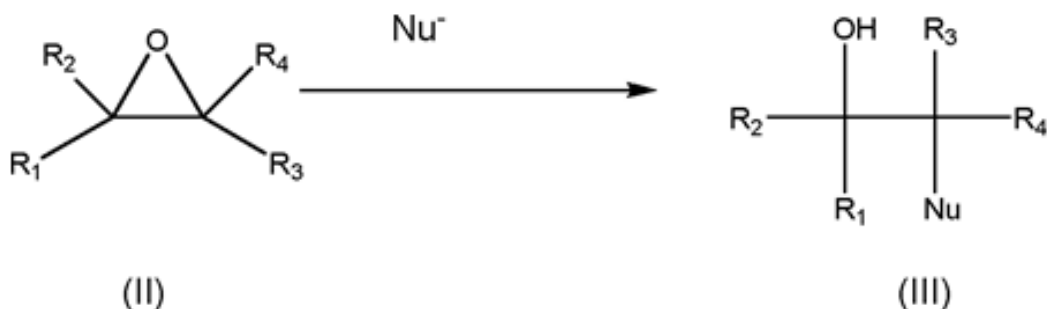
en el que Hal es Cl o Br y en el caso de que más de un Hal preferentemente es idéntico en sustancia, y R_7 es seleccionado de halógeno (F, Cl, Br, I) o alquilo C1-C3 y cada ----- representa independientemente un enlace doble o simple y n es cero o un número entero entre 1 y 4, preferentemente 0, 1 o 2.

15

Apertura de anillo de epóxidos a alcoholes

La al menos una biomolécula que tiene la secuencia de proteínas o secuencia de ADN codificante de HHeD3 que tiene NCBI, número de acceso GenBank ABM93639 también es usada como halohidrina deshalogenasa en la conversión de al menos una parte de un epóxido de fórmula (II) en un alcohol de fórmula (III)

5



en el que

$\text{R}_1, \text{R}_2, \text{R}_3, \text{R}_4$

son independientemente hidrógeno, halógeno (F, Cl, Br, I) o alquilo C1-C12, alquenilo C2-C12, alquinilo C2-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o arilo C5-C12 sustituido o no sustituido, o -alquil C1-C3-C(=O)-O-R₅ o -alquil C1-C3-O-R₅, con

R_5

seleccionado de hidrógeno, alquilo C1-C6, alquenilo C2-C6, alquinilo C2-C6, alcoxi C1-C6, alquilarilo C5-C13, arilalquilo C5-C13, ariloxi C5-C13, arilalcoxi C5-C13, heteroalquilo C5-C13, arilo C5-C13, heteroarilo C5-C13, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o heteroarilo cíclico C5-C13, en el que el caso de más de un sistema anular alifático o aromático, los sistemas anulares están condensados o separados y R₅ puede tener al menos un sustituyente adicional R₆ con

R_6

seleccionado de hidrógeno, hidroxilo, -C(=O)-NH₂, halógeno (F, Cl, Br, I), -alquilo-C1-C3, -C(=O)-alquilo C1-C3, -C(=O)-alquilo-C1-C3, -NH-C(=O)-alquilo-C1-C3, -alquil C1-C3-O-alquilo-C1-C3, alquenilo C2-C3, -O-alquenilo-C2-C3, cicloalquilo C5-C7, heteroalquilo cíclico C5-C7, arilo C5-C7 o heteroarilo C5-C7, preferentemente seleccionado de hidrógeno y metilo,

y/o

dos de $\text{R}_1, \text{R}_2, \text{R}_3, \text{R}_4$ forman en conjunto una estructura cíclica seleccionada de cicloalquilo C3-C12, cicloalquenilo C4-C12, cicloalquinilo C6-C12, heteroalquilo cíclico C3-C12, heteroalquenilo cíclico C4-C12 o heteroalquinilo cíclico C6-C12, preferentemente un grupo cicloalquilo C4-C8, o cicloalquenilo C4-C8, más preferentemente un grupo cicloalquilo C5-C7, o cicloalquenilo C5-C7, con máxima preferencia un grupo cicloalquilo C5-C6, o cicloalquenilo C5-C6 en el que el caso de más de un sistema anular, los sistemas anulares están condensados o separados, y en el que el número de átomos de C indicados para la estructura cíclica siempre incluye los 2 átomos de C de la estructura de halohidrina que portan el OH y el X, y

en el que cada grupo cíclico en $\text{R}_1, \text{R}_2, \text{R}_3, \text{R}_4$ o estructura cíclica formada por dos de $\text{R}_1, \text{R}_2, \text{R}_3, \text{R}_4$ opcionalmente tiene al menos un sustituyente adicional diferente de hidrógeno, preferentemente seleccionado de R₆, con máxima preferencia halógeno (F, Cl, Br, I) o alquilo C1-C3. Este al menos un sustituyente adicional opcional también puede ser denominado sustituyente R₇, que es especialmente (R₇)_n con n que es cero o un número entero entre 1 y 4, preferentemente n = 0, 1 o 2.

Nu es seleccionado de R'NO₂, N₃, NO₂, SCN, OCN, CN, halógeno (F, Cl, Br, I, especialmente Cl, Br, I) y R"CO₂, en el que R' es seleccionado de alquilo C1-C4, C1-C4-alquenilo, preferentemente CH₃,

R" es seleccionado de hidrógeno, alquilo C1-C4, alquenilo C1-C4, preferentemente hidrógeno,

y Nu⁻ es un anión de Nu.

En una realización especial, Nu es seleccionado de N₃, NO₂, SCN, OCN, CN, halógeno (Cl, Br, I) y R"CO₂, es decir,

se excluye $R'NO_2$. Preferentemente, los hidrocarburos en R' y/o en R'' son lineales/no ramificados.

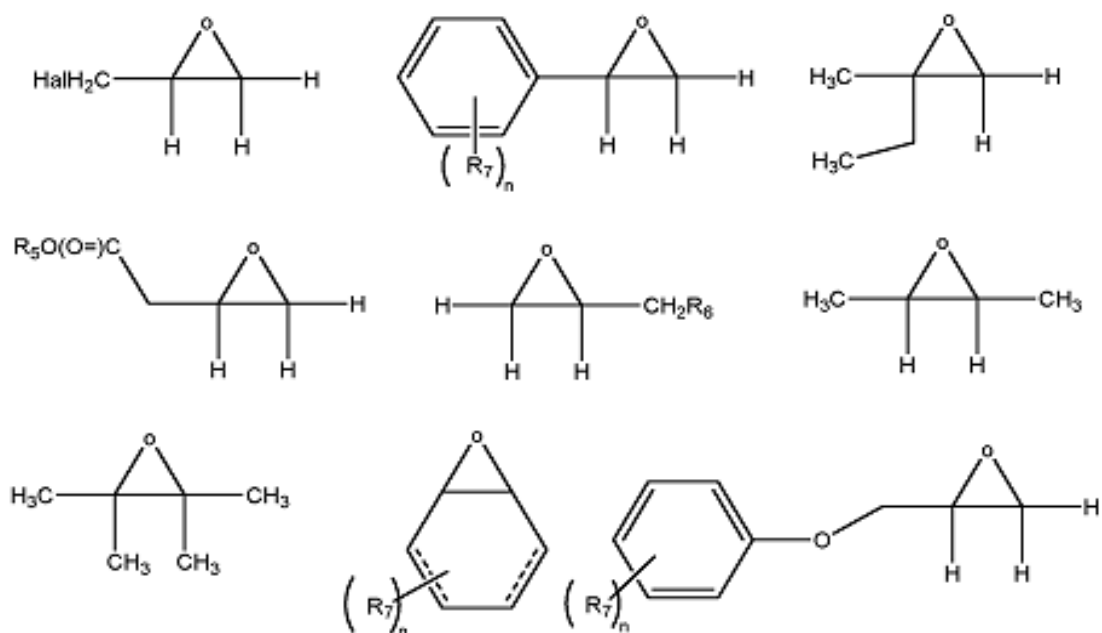
Los términos alqueno/alquénilo, cicloalqueno/cicloalquénilo se refieren a hidrocarburos que contienen al menos un doble enlace carbono-carbono como ya se explicó anteriormente. Los términos alquino/alquínilo, cicloalquino/cicloalquínilo se refieren a hidrocarburos que contienen al menos un triple enlace carbono-carbono.

- 5 Con respecto a R_1, R_2, R_3, R_4 : cada uno de R_1, R_2, R_3, R_4 , que no forma una estructura cíclica con otro de R_1, R_2, R_3, R_4 , es decir, cada uno de estos residuos R_1, R_2, R_3, R_4 , que no está covalentemente unido a otro de estos residuos R_1, R_2, R_3, R_4 , puede tener al menos un átomo de carbono reemplazado por un heteroátomo. En otras palabras, cada R_1, R_2, R_3, R_4 es un hidrocarburo, es decir, alquilo C1-C12, alquénilo C1-C12, alquínilo C1-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalquénilo C3-C12, cicloalquínilo C3-C12 o arilo C5-C12 sustituido o no sustituido puede tener
- 10 al menos un átomo de carbono reemplazado por un heteroátomo.

En general, es decir, con respecto a cualquiera de los residuos R_1 - R_7 : La expresión "hetero" o "heteroátomo" significa preferentemente un heteroátomo seleccionado de nitrógeno (N), azufre (S), oxígeno (O), selenio (Se) o fósforo (P), más preferentemente un heteroátomo seleccionado de N, S u O.

- 15 Con respecto a todas las sustancias de fórmula (II), los átomos de C que forman el anillo epóxido siempre tienen hibridación sp^3 . Además, los átomos de C que portan OH y Nu en la fórmula (III) siempre tienen hibridación sp^3 .

El epóxido de fórmula (II) es seleccionado preferentemente de



- 20 en el que Hal es Cl o Br y en el caso de más de un Hal es preferentemente idéntico en sustancia, y R_7 es seleccionado de hidrógeno, halógeno (F, Cl, Br, I) o alquilo C1-C3 cada --- representa independientemente un enlace doble o simple y n es cero o un número entero entre 1 y 4, preferentemente 0, 1 o 2.

En cada fórmula (I), (II) y/o (III) al menos dos de R_1, R_2, R_3, R_4 son preferentemente hidrógeno o $-CH_3$, más preferentemente uno de R_1, R_2 y uno de R_3, R_4 son hidrógeno. Esto se aplica a las reacciones separadas descritas anteriormente y a la reacción combinada explicada a continuación.

25 Reacción combinada

En una realización preferente de la invención, la biomolécula que tiene secuencia de proteína o secuencia de ADN codificante de HHeD3 que tiene NCBI, número de acceso GenBank ABM93639 se usa como halohidrina deshalogenasa en la conversión de al menos una parte de la cantidad de halohidrina de fórmula (I) en un epóxido de

fórmula (II) y en la conversión de al menos una parte del epóxido de fórmula (II) en un alcohol de fórmula (III). Preferentemente, esto se hace como reacción en un recipiente, síntesis en un recipiente, lo que significa que la sustancia inicial de fórmula (I) se somete a reacciones sucesivas, lo que finalmente da como resultado sustancias de fórmula (III) en un solo reactor.

- 5 Con respecto a las realizaciones preferentes de (I) y (III) en la reacción combinada, aplica lo mismo descrito anteriormente con respecto a las reacciones separadas.

En variantes posibles de la conversión de al menos una parte de la cantidad de halohidrina de fórmula (I) en un epóxido de fórmula (II) y/o en la conversión de al menos una parte de un epóxido de fórmula (II) en un alcohol de fórmula (III), es decir, en reacciones separadas de (I) a (II) o de (II) a (III) o en una reacción combinada de (I) a (III), la al menos una biomolécula es una proteína seleccionada de HheA6, HheB3, HheB4, HheB5, HheB6, HheB7, HheD, HheD2, HheD3, HheD4, HheD5, HheD6, HheD7, HheD8, HheD9, HheD10, HheD11, HheD12, HheD13, HheD14, HheD15, HheD16, HheD17, HheE, HheE2, HheE3, HheE4, HheE5, HheEx, HheEx2, HheEx3, HheF, HheG, en el que solo HHeD3 está abarcada por la invención.

Descripción detallada de la invención

- 15 A continuación, la invención y su desarrollo son descritos con mayor detalle.

Características de secuencia de las HHDH

En primer lugar, con el fin de identificar nuevas enzimas HHDH, fueron inspeccionadas todas las secuencias conocidas de HHDH en busca de residuos distintivos que distinguen las HHDH de otras enzimas SDR. Todas las HHDH conocidas poseen una tríada catalítica de Ser-Tyr-Arg (SYR) alineada con los residuos catalíticos Ser-Tyr-Lys (SYK) presentes en las secuencias SDR. Por lo tanto, la presencia de Arg en la tríada catalítica de HHDH se puede usar como criterio inicial para filtrar las nuevas HHDH putativas de otras secuencias SDR. Además, las estructuras cristalinas de HheA2 (23) y HheC (24) muestran que ambas HHDH poseen un bolsillo de unión de aniones espacioso que está formado en parte por residuos que se alinean con residuos de un motivo rico en Gly responsable de la unión del cofactor de nucleótidos en a las enzimas con plegamiento Rossmann, tal como las enzimas SDR. Específicamente, el residuo grande F12 tanto en HheA2 como en HheC es esencial para la formación de la bolsa de unión del anión HHDH y reemplaza un Gly o Ala pequeño en el motivo de unión de nucleótidos "TGx (3) - [GA] -x-G" de las enzimas SDR clásicas (22-24). En las enzimas conocidas HheB/B2, un residuo de Tyr está alineado con la Phe de las otras HHDH conocidas. Por lo tanto, en todas las enzimas HHDH conocidas, los aminoácidos aromáticos grandes Phe (F) o Tyr (Y) alteran el motivo de unión del cofactor rico en Gly comúnmente observado de las enzimas SDR. En consecuencia, los inventores afirman que la presencia tanto de la tríada catalítica HHDH (semejante al motivo (2), ver a continuación) como del Phe/Tyr aromático (incluido en el motivo (1), ver a continuación) indica secuencias con actividad HHDH. Especialmente la definición del motivo (1), con un aminoácido aromático en lugar del motivo rico en Gly normal, es nueva y no se ha realizado antes. En combinación con el motivo (2), el motivo (1) es suficiente para discriminar las secuencias HHDH verdaderas de las secuencias SDR no relacionadas.

Extracción de base de datos

Para identificar nuevas secuencias de HHDH, fueron iniciadas búsquedas BLASTp para recolectar secuencias homólogas de GenBank que después pueden ser evaluadas para determinar la presencia de características conservadas de la secuencia de HHDH.

A partir de la colección GenBank de secuencias no redundantes (nr) y secuencias ambientales no redundantes (env_nr), muchas secuencias nuevas de HHDH (Tabla 1) contenían la tríada catalítica de HHDH conocidas en combinación con el aromático Phe o Tyr específico para enzimas HHDH conocidas. Para las nuevas secuencias de HHDH obtenidas, no se pueden recuperar informes sobre su actividad de los registros GenBank asociados, ya que estos registros se anotan principalmente como "enzima SDR" u "oxidorreductasa" pero nunca como enzima HHDH.

TABLA. 1 Fuentes y números de acceso de HDDH previamente conocidos con actividad probada (*), de enzimas previamente denominadas HDDH sin prueba de actividad HDDH o de secuencias con alta identidad de secuencia en comparación con HDDH conocidas (**) y HDDH nuevas

HDDH	Organismo o fuente	Acceso (NCBI, GenBank)
HheA*	Corynebacterium sp.	BAA14361
HheA2*	Arthrobacter sp. AD2	AAK92100
HheA3**	Parvibaculum lavamentivorans DS-1	ABS64560
HheA4**	Arthrobacter sp. JBH1	AFI98638
HheA5**	Tistrella mobilis KA081020-065	AFK51877
HheA6	bacterium Ec32	CDO61292
HheB*	Corynebacterium sp.	BAA14362
HheB2*	Mycobacterium sp. GP1	AAK73175
HheB3	metagenoma marino	SEQ ID NÚM. 1
HheB4	metagenoma marino	EBP61646
HheB5	metagenoma marino	SEQ ID NÚM. 2
HheB6	metagenoma marino	SEQ ID NÚM. 3
HheB7	metagenoma marino	SEQ ID NÚM. 4
HheBx	Ant4E12 bacteria marino no cultivada	SEQ ID NÚM. 15
HheC*	Agrobacterium tumefaciens	AAK92099
HheD	Dechloromonas aromatica RCB	AAZ44846
HheD2	gamma proteobacterium HTCC2207	EAS46473
HheD3	Methylobium petroleiphilum PM1	ABM93639
HheD4	metagenoma marino	SEQ ID NÚM. 5
HheD5	Thauera sp. MZ1T	YP_002355872
HheD6	Marinobacter nanhaiticus D15-8W	ENO15189
HheD7	Thauera sp. 27	ENO82779
HheD8	Thauera aminoaromatica S2	ENO87252
HheD9	Thauera phenylacetica B4P	ENO98837
HheD10	Limnohabitans sp. Rim28	SEQ ID NÚM. 6
HheD11	Curvibacter lanceolatus	SEQID NO.7
HheD12	Thiothrix disciformis	WP_020394200
HheD13	Pseudomonas pelagia	WP_022962804
HheD14	Betaproteobacteria bacterium	ESS13801
HheD15	Gammaproteobacteria bacterium	ETN91936
HheD16	candidatus Competibacter	CDI00977

HHDH	Organismo o fuente	Acceso (NCBI, GenBank)
HheD17	Methylibium sp. T29	EWS52496
HheE	metagenoma marino	EBP63112
HheE2	metagenoma marino	ECW41905
HheE3	metagenoma marino	SEQ ID NÚM. 8
HheE4	metagenoma marino	SEQ ID NÚM. 9
HheE5	gamma proteobacterium IMCC3088	SEQ ID NÚM. 10
HheEx	marine gamma proteobacterium	SEQ ID NÚM. 11
HheEx2	metagenoma marino	SEQ ID NÚM. 12
HheEx3	metagenoma marino	SEQ ID NÚM. 13
HheF	bacteria no cultivada	SEQ ID NÚM. 14
HheG	Ilumatobacter coccineus YM16-304	BAN03849
HheGx	Microbacterium yannicii	WP_019180129
Las secuencias indicadas con ** en la Tabla 1 solo son denominadas como HHDH o son similares a las HHDH conocidas anteriormente (*) pero su actividad nunca fue probada (para más detalles, véase anteriormente).		

5 Varias de las secuencias nuevas (SEQ ID NÚM. 11-13 y 15) en la Tabla 1 corresponden/son similares a las secuencias enumeradas en las bases de datos, en las que las secuencias enumeradas en la base de datos pierden aminoácidos en el extremo N-terminal o C-terminal, que se suplementaron aquí por los inventores. Con respecto a las secuencias novedosas SEQ ID NÚM. 1-10 y 14, las secuencias enumeradas en las bases de datos contienen un exceso de aminoácidos, que fueron suprimidos por los inventores. Para estas secuencias enumeradas en la base de datos, no se pueden recuperar informes sobre su actividad de los registros GenBank asociados, ya que estos registros se anotan principalmente como "enzima SDR" u "oxidorreductasa". Solo la secuencia enumerada en la base de datos correspondiente a HheBx (GenBank: ABC25271) está anotada como HHDH pero, debido a las piezas de secuencia que faltan, no puede actuar como HHDH. Los accesos correspondientes de las secuencias enumeradas en la base de datos se dan en la Tabla 2 a continuación:

TABLA 2. Accesos correspondientes de secuencias enumeradas pero incompletas o demasiado largas

HHDH	Acceso (NCBI, GenBank)	SEQ ID NÚM.
HheB3	EBL02020	1
HheB5	ECR06649	2
HheB6	EDB56284	3
HheB7	EDD65701	4
HheBx	ABC25271	15
HheD4	ECY18578	5
HheD10	WP_019427705	6
HheD11	WP_019574066	7
HheE3	EDF62577	8
HheE4	EDH34310	9
HheE5	EGG28524	10

HheEx	EAW33023	11
HheEx2	EBC61576	12
HheEx3	EBF34592	13
HheF	BAH89601	14

Especialmente HheBx solo es funcional con una secuencia de aminoácidos de acuerdo con la SEQ ID NÚM. 15. La variación de la base de datos enumerada ABC25271 no es funcional porque falta el motivo (1).

Descripción de las figuras

- 5 **FIG. 1** Conversiones de diferentes haloalcoholes con extractos libres de células de expresión de HHDH recombinante. Las conversiones se comparan para la formación de epóxidos correspondientes a partir de 1,3-dibromo-2-propanol (blanco), 1,3-dicloro-2-propanol (negro) y 2-clorofeniletanol (líneas) después de deducir las actividades de fondo de los controles de vector vacío.

Verificación experimental de actividad HHDH

- 10 Con el fin de investigar si las secuencias identificadas realmente representan enzimas con actividad de HHDH verdadera, se ordenó una selección de 19 genes sintéticos optimizados por codones para la expresión heteróloga de HheA3, HheA5, HheB3, HheB4, HheB5, HheB6, HheB7, HheD, HheD2, HheD3, HheD4, HheD5, HheE, HheE2, HheE3, HheE4, HheE5, HheF o HheG en *E. coli*.

- 15 Para confirmar la actividad de HHDH, los 19 genes sintéticos de HHDH fueron clonados en el vector de expresión de pET-28a bien establecido. La expresión heteróloga de la enzima soluble fue optimizada para cada HHDH mediante la variación de parámetros tal como huésped de expresión [*E. coli* BL21 (DE3) o C43 (DE3)] o temperatura de expresión (20, 30 o 37 °C). La optimización de las condiciones de expresión produjo bandas visibles de enzima soluble para la mayoría de las HHDH después de la tinción de Coomassie de geles de poliacrilamida que no estaban presentes en los controles de vector vacíos.

- 20 Para evaluar si estas enzimas recombinantes poseen actividad de HHDH verdadera, fue llevado a cabo un ensayo de actividad colorimétrica para la liberación de haluro de cloroalcoholes 1,3-dicloro-2-propanol, 2-clorofeniletanol y 1,3-dibromo-2-propanol con células libre de extracto (CFE) de células que contienen las enzimas expresadas de forma recombinante. Posteriormente, se usó el análisis de cromatografía de gases para confirmar la formación de los epóxidos correspondientes. Para todas las HHDH seleccionadas, se pudo detectar la liberación de haluro así como la
- 25 formación de epóxido para al menos uno de los sustratos. La Fig. 1 resume las conversiones obtenidas de los diferentes CFE que contienen HHDH (que se calcula a partir de la cantidad de epóxido formado después de la deducción de los controles de vector vacíos). Las HHDH, HheA2, HheB2 y HheC conocidas, fueron incluidas en las mediciones para comparación.

Huellas digitales para la recuperación exclusiva de secuencias de HHDH

- 30 Una vez confirmado que el enfoque identifica específicamente secuencias que exhiben una actividad de HHDH verdadera, también se intentó optimizar la rutina de búsqueda para identificar incluso secuencias relacionadas más lejanas y, al mismo tiempo, simplificar el procedimiento general.

- 35 En las búsquedas BLASTp iniciales de secuencias de HHDH novedosas, la gran mayoría de los resultados estuvo dominada por secuencias de SDR y, por lo tanto, se podrían haber pasado por alto las HHDH relacionadas lejanamente. Para evitar esto, se puede usar PHI-BLAST para detectar secuencias relacionadas más lejanamente que coinciden con un patrón definido por el usuario (25). Para reconocer tal patrón, se inspeccionó el alineamiento MAFFT de todas las secuencias de HHDH nuevas y conocidas para detectar la presencia de motivos conservados.

- 40 Como se esperaba del protocolo de identificación descrito, todas las HHDH identificadas necesariamente poseían la tríada catalítica Ser-Tyr-Arg similar al patrón "S-x (12) -Y-x (3) -R". Además, el Phe o Tyr aromático conservado es parte de un patrón específico de HHDH "T-x(4)-[FY]-x-G" que contrasta el motivo SDR "T-G-x(3)-[GA]-x-G". En consecuencia, cualquiera de los patrones específicos de HHDH puede servir como patrón semilla para búsquedas PHI-BLAST.

El nuevo motivo (1) (secuencia de aminoácidos HHDH) por lo tanto es

1) "T-x(4)-[FY]-x-G"

- 45 y el nuevo motivo (secuencia de aminoácidos de HHDH) (2) por lo tanto es

2) "S-x(12)-Y-x(3)-R",

Con cualquier patrón como semilla para búsquedas PHI-BLAST, las secuencias de cada tipo filogenético de HDDH, a saber, HheB, HheC, HheD, HheE, HheF y HheG, fueron usadas para consultar la base de datos nr. Dentro de estos resultados, se incluyeron nuevamente las enzimas SDR pero también todas las HDDH verificadas experimentalmente que se habían identificado anteriormente. Por tanto, cada patrón solo es suficiente para recuperar las secuencias de HDDH disponibles independientemente de la secuencia incógnita.

Al aumentar el valor de corte de E (hasta 1000) y usando cualquiera de las otras secuencias de HDDH como incógnita, no pudo ser identificada ninguna secuencia de HDDH adicional.

Aunque cada patrón por sí solo reduce en gran medida el número total de secuencias para evaluar para la característica de secuencia complementaria, aun así, las secuencias SDR siempre fueron incluidas en todos los resultados. Para la recuperación específica solo de secuencias HDDH, es posible combinar ambos patrones para identificar secuencias que deben concordar con ambas características de secuencia mencionadas anteriormente. Como ambos patrones están separados por 93 a 131 residuos en las HDDH verificadas experimentalmente, ambos motivos se pueden combinar en el patrón "T-x(4)-[FY]-x-G-x(93,131)-S-x(12)-Y-x(3)-R", o, más general "T-x(4)-[FY]-x-G-x(1,250)-S-x(12)-Y-x(3)-R".

A continuación, usando el patrón combinado con cualquier secuencia HDDH como incógnita PHI-BLAST, solo son recuperadas las nuevas HDDH (Tabla 1) como aciertos relevantes (valor E <0,1). La variación de la longitud de brecha entre 1 y 250 residuos no recupera ninguna secuencia relevante adicional (con un valor E <0,1) con ambas características de la secuencia HDDH. La consulta de la base de datos env_nr actualizada con patrones separados o combinados no produjo la recuperación de nuevas secuencias de HDDH putativas adicionales.

Además, las HDDH nuevas y conocidas, así como otras enzimas SDR, poseen un motivo conservado (3) de "G-x (6) -G-G" que es crucial para su oligomerización. Como algunos informes sugieren que al menos HheC se inactiva si se inhibe la formación del tetrámero, este tercer motivo puede ser útil para discernir las enzimas HDDH activas.

De este modo, ha sido identificado tercer motivo (3):

3) "G-x(6)-G-G"

Dado que el patrón (3) está separado por 76 a 84 residuos en las HDDH verificadas experimentalmente, los tres motivos pueden ser combinados en el patrón "T-x(4)-[FY]-x-G-x(93,131)-S-x(12)-Y-x(3)-R-x(76,84)-G-x(6)-G-G", o, más general "T-x(4)-[FY]-x-G-x(1,250)-S-x(12)-Y-x(3)-R-x(1,250)-G-x(6)-G-G".

En resumen, fue posible reducir en gran medida las secuencias para analizar a partir de búsquedas de homología haciendo uso de patrones de secuencia específicos de HDDH restrictivos, especialmente "T-x(4)-[FY]-x-G" y/o "S-x(12)-Y-x(3)-R". La combinación de patrones (1), (2) y (3), más preferentemente (1) y (2), es capaz de discernir secuencias no HDDH de los resultados de búsqueda de homología ($E \geq 0.1$) y, en consecuencia, es altamente útil para la identificación precisa de secuencias HDDH presentes en bases de datos públicas a partir de la gran cantidad de secuencias SDR similares pero diferentes.

Material y procedimientos

Extracción de base de datos: Las secuencias de HheA, HheB y HheC (Tabla 1) fueron usadas como incógnitas para búsquedas BLASTp (26) de la base de datos nr y env_nr de GenBank (versión 200) (27). Para cada incógnita, fueron recuperadas 20.000 secuencias y usadas junto con HheA, HheB y HheC para la construcción de alineamientos de secuencia múltiple (MSA) con MAFFT (FFT-NS-2) (28). En los alineamientos resultantes, las secuencias fueron descartadas si poseían los residuos catalíticos típicos Ser-Tyr-Lys de las enzimas SDR que se alinearon con la tríada catalítica Ser-Tyr-Arg de HheA/HheA2 (S134-Y148-R151), HheB/HheB2 (S127-Y139-R143) y HheC (S132-Y145-R149). Después de la eliminación de esas secuencias de SDR putativas, la mezcla de secuencias reducidas de secuencias restantes fue alineada nuevamente usando MAFFT (FFT-NS-2). Posteriormente, solo fueron seleccionadas secuencias con una tríada catalítica de Ser-Tyr-Arg también presente en enzimas HDDH conocidas para la construcción de un nuevo MAFFT (L-INS-i) (29) MSA. A partir de este último alineamiento, solo se consideraron las secuencias para nuevas HDDH putativas si poseían un Phe o Tyr aromático que se alinea con F12, Y27 y F12 de HheA/HheA2, HheB/HheB2 y HheC, respectivamente. Todas las nuevas secuencias de HDDH se usaron como incógnitas en rutinas de búsqueda posteriores que consistieron en búsquedas BLASTp, construcción de MSA y su inspección para detectar la presencia de la tríada catalítica de HDDH típica en combinación con el residuo aromático conservado cerca del extremo N-terminal como se describe anteriormente. Esta estrategia se continuó hasta que no se pudieron identificar secuencias de HDDH nuevas adicionales en la base de datos nr5 y env_nr.

Síntesis y clonación génica: Los genes sintéticos que codifican 19 HDDH (HheA3, HheA5, HheB3, HheB4, HheB5, HheB6, HheB7, HheD, HheD2, HheD3, HheD4, HheD5, HheE, HheE2, HheE3, HheE4F, HheE5, Hhe) fueron solicitados a Life Technologies (Darmstadt, Germany) después de la traducción inversa de las secuencias de proteínas curadas y la optimización de codones para *E. coli* con GeneOptimizer (30). Los genes sintéticos fueron escindidos de los plásmidos recibidos mediante restricción con NdeI y HindIII o XhoI seguido de ligación con ADN T4 ligasa (todas

las enzimasificadoras de ADN de New England Biolabs, Frankfurt, Germany) en pET-28a linealizado (Merck, Darmstadt, Germany). Después de la transformación en *E. coli* DH5 α (Life Technologies), fue preparado ADN plasmídico recombinante con el kit NucleoSpin Plasmid (Macherey Nagel, Düren, Germany) y enviado para secuenciación en GATC Biotech (Konstanz, Germany).

- 5 **Ensayos de expresión y actividad:** Los plásmidos que contienen los diferentes genes HHDH fueron transformados en *E. coli* BL21 (DE3) (Life Technologies) o C43 (DE3) (Lucigen Corporation, Middleton, WI, USA), para la expresión de proteínas heterólogas. Cada medio TB de 20 ml que contiene 50 mg L⁻¹ de kanamicina e IPTG 0,2 mM fue inoculado con 10% v/v de un precultivo de *E. coli* respectivo e incubado a 20, 30 o 37 °C durante 7 a 24 h. Posteriormente, los cultivos fueron centrifugados y los gránulos almacenados a -20 °C. A modo de comparación, las enzimas conocidas
- 10 HheA2, HheB2 y HheC fueron expresadas de forma recombinante en *E. coli* Top10 (Life Technologies) a partir del vector pBAD (Life Technologies) en medio TB con 100 mg de ampicilina L⁻¹ y L-arabinosa 0,02%. Para preparar extractos libres de células (CFE), los gránulos celulares fueron resuspendidos en 1,2 ml de tampón Tris•SO₄ 25 mM, pH 8,0, y disgregados mediante sonicación. Después de la centrifugación, cada 200 ml de estos CFE fueron añadidos a 600 ml de tampón Tris•SO₄ 25 mM que contiene uno de los sustratos 1,3-dicloro-2-propanol, 2-clorofeniletanol o 1,3-
- 15 dibromo-2-propanol en concentración final 10 mM para ensayos de actividad. Las reacciones fueron incubadas a 30 °C y fueron tomadas muestras de 50 ml después de 5, 20 y 60 minutos de incubación para controlar la actividad de la deshalogenasa usando el ensayo de liberación de haluro según lo descrito en otra parte (31). Además, después de 60 min, fueron extraídos 650 ml restantes por reacción una vez con cada 600 ml de metil terc-butil éter. Los extractos orgánicos fueron secados sobre sulfato de magnesio y se analizaron en un cromatógrafo de gases GC2010 (Shimadzu, Japan) equipado con una columna Supreme 5ms (CS Chromatography Service, Germany). El análisis de GC de las reacciones que contienen 1,3-dicloro-2-propanol o 1,3-dibromo-2-propanol se realizó con un gradiente de temperatura a partir de 40 °C durante 1 min, calentando a 10 °C min⁻¹ a 120 °C y posteriormente con 20 °C min⁻¹ a 300 °C. En el caso de reacciones que contienen 2-clorofeniletanol, el gradiente de temperatura fue iniciado a 80 °C durante 1 min, calentando con 10 °C min⁻¹ a 160 °C y finalmente con 20 °C min⁻¹ a 300 °C. Los sustratos fueron eluidos a los 6,6 min (1,3-dicloro-2-propanol), 8,6 min (2-clorofeniletanol) y 9,4 min (1,3-dibromo-2-propanol) mientras que los productos correspondientes fueron detectados a los 3,7 min (epiclorhidrina), 5,8 min (óxido de estireno) y 4,8 min (epibromhidrina), respectivamente.

- Huellas digitales de HHDH:** Los residuos conservados se dedujeron de los alineamientos MAFFT de todas las HHDH o con secuencias SDR homólogas. Los patrones "T-x(4)-[FY]-x-G" (motivo 1), "S-x(12)-Y-x(3)-R" (motivo 2), o sus combinaciones se usaron como semillas en las búsquedas PHI-BLAST (25) de la base de datos GenBank nr y env_nr (versión 200). Aquí, HheB, HheC, HheD, HheE, HheF y HheG se usaron como incógnitas con un umbral E de 10. Todos los aciertos se alinearon con MAFFT (FFT-NS-2) y el MSA resultante se inspeccionó en busca de secuencias con una tríada catalítica HHDH tríada (motivo 2) en combinación con el aromático Phe o Tyr (dentro del motivo 1) presente en las enzimas HHDH conocidas como se describe anteriormente.

35

Referencias

1. Janssen DB, Majeric-Elenkov M, Hasnaoui G, Hauer B, Lutje Spelberg JH. 2006. Enantioselective formation and ring-opening of epoxides catalysed by halohydrin dehalogenases. *Biochem. Soc. Trans.* 34:291-295.
2. Schallmeyer M, Floor RJ, Szymanski W, Janssen DB. 2012. 7.8 Hydrolysis and Reverse Hydrolysis: Halohydrin Dehalogenases, p. 143-155. In Editors-in-Chief: Erick M. Carreira, Hisashi Yamamoto (eds.), *Comprehensive Chirality*. Elsevier, Amsterdam.
3. You Z-Y, Liu Z-Q, Zheng Y-G. 2012. Properties and biotechnological applications of halohydrin dehalogenases: current state and future perspectives. *Appl. Microbiol. Biotechnol.* 97:9-21.
4. Nakamura T, Nagasawa T, Yu F, Watanabe I, Yamada H. 1992. Resolution and some properties of enzymes involved in enantioselective transformation of 1,3-dichloro-2-propanol to (R)-3-chloro-1,2-propanediol by *Corynebacterium* sp. strain N-1074. *J Bacteriol* 174:7613-7619.
5. Lutje Spelberg JH, van Hylckama Vlieg JET, Bosma T, Kellogg RM, Janssen DB. 1999. A tandem enzyme reaction to produce optically active halohydrins, epoxides and diols. *Tetrahedron Asymmetry* 10:2863-2870.
6. Haak RM, Tarabiono C, Janssen DB, Minnaard AJ, de Vries JG, Feringa BL. 2007. Synthesis of enantiopure chloroalcohols by enzymatic kinetic resolution. *Org. Biomol. Chem.* 5:318-323.
7. Haak RM, Berthiol F, Jerphagnon T, Gayet AJA, Tarabiono C, Postema CP, Ritleng V, Pfeffer M, Janssen DB, Minnaard AJ, Feringa BL, de Vries JG. 2008. Dynamic Kinetic Resolution of Racemic β -Haloalcohols: Direct Access to Enantioenriched Epoxides. *J. Am. Chem. Soc.* 130:13508-13509.
8. Jin H-X, Hu Z-C, Liu Z-Q, Zheng Y-G. 2012. Nitrite-mediated synthesis of chiral epichlorohydrin using halohydrin dehalogenase from *Agrobacterium radiobacter* AD1. *Biotechnol. Appl. Biochem.* 59:170-177.
9. Majeric Elenkov M, Primožič I, Hrenar T, Smolko A, Dokli I, Salopek-Sondi B, Tang L. 2012. Catalytic activity of halohydrin dehalogenases towards spiroepoxides. *Org. Biomol. Chem.* 10:5063-5072.
10. Hasnaoui-Dijoux G, Majeric Elenkov M, Lutje Spelberg JH, Hauer B, Janssen DB. 2008. Catalytic Promiscuity of Halohydrin Dehalogenase and its Application in Enantioselective Epoxide Ring Opening. *ChemBioChem* 9:1048-1051.
11. Nakamura T, Nagasawa T, Yu F, Watanabe I, Yamada H. 1994. A new enzymatic synthesis of (R)-gamma-chloro-beta-hydroxybutyronitrile. *Tetrahedron* 50:11821-11826.
12. Lutje Spelberg JH, Tang L, Kellogg RM, Janssen DB. 2004. Enzymatic dynamic kinetic resolution of epihalohydrins. *Tetrahedron Asymmetry* 15:1095-1102.
13. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S, Grate J, Gruber J, Whitman JC, Sheldon RA, Huisman GW. 2007. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* 25:338-344.
14. Majeric Elenkov M, Hauer B, Janssen DB. 2006. Enantioselective Ring Opening of Epoxides with Cyanide Catalysed by Halohydrin Dehalogenases: A New Approach to Non-Racemic β -Hydroxy Nitriles. *Adv. Synth. Catal.* 348:579-585.
15. Majeric Elenkov M, Hoeffken HW, Tang L, Hauer B, Janssen DB. 2007. Enzyme-Catalyzed Nucleophilic Ring Opening of Epoxides for the Preparation of Enantiopure Tertiary Alcohols. *Adv. Synth. Catal.* 349:2279-2285.
16. Fuchs M, Simeo Y, Ueberbacher BT, Mautner B, Netscher T, Faber K. 2009. Enantiocomplementary Chemoenzymatic Asymmetric Synthesis of (R)- and (S)-Chromanemethanol. *Eur. J. Org. Chem.* 2009:833-840.
17. Molinaro C, Guilbault A-A, Kosjek B. 2010. Resolution of 2,2-Disubstituted Epoxides via Biocatalytic Azidolysis. *Org. Lett.* 12:3772-3775.
18. Yu F, Nakamura T, Mizunashi W, Watanabe I. 1994. Cloning of two halohydrin hydrogen-halide-lyase genes of *Corynebacterium* sp. strain N-1074 and structural comparison of the genes and gene products. *Biosci. Biotechnol. Biochem.* 58:1451-1457.
19. Van Hylckama Vlieg JET, Tang L, Lutje Spelberg JH, Smilda T, Poelarends GJ, Bosma T, van Merode AEJ, Fraaije MW, Janssen DB. 2001. Halohydrin Dehalogenases Are Structurally and Mechanistically Related to Short-Chain Dehydrogenases/Reductases. *J. Bacteriol.* 183:5058-5066.
20. Higgins TP, Hope SJ, Effendi AJ, Dawson S, Dancer BN. 2005. Biochemical and molecular characterisation of the 2,3-dichloro-1-propanol dehalogenase and stereospecific haloalkanoic dehalogenases from a versatile *Agrobacterium* sp. *Biodegradation* 16:485-492.

21. De Jong RM, Dijkstra BW. 2003. Structure and mechanism of bacterial dehalogenases: different ways to cleave a carbon-halogen bond. *Curr. Opin. Struct. Biol.* 13:722-730.
22. Kavanagh KL, Jörnvall H, Persson B, Oppermann U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* 65:3895-3906.
- 5 23. De Jong RM, Kalk KH, Tang L, Janssen DB, Dijkstra BW. 2006. The X-Ray Structure of the Haloalcohol Dehalogenase HheA from *Arthrobacter* sp. Strain AD2: Insight into Enantioselectivity and Halide Binding in the Haloalcohol Dehalogenase Family. *J. Bacteriol.* 188:4051-4056.
24. De Jong RM, Tiesinga JJW, Rozeboom HJ, Kalk KH, Tang L, Janssen DB, Dijkstra BW. 2003. Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. *EMBO J.* 22:4933-4944.
- 10 25. Zhang Z, Miller W, Schäffer AA, Madden TL, Lipman DJ, Koonin EV, Altschul SF. 1998. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* 26:3986-3990.
26. 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.
- 15 27. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. *Nucleic Acids Res.* 41:D36-D42.
28. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059-3066.
- 20 29. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33:511 -518.
30. Raab D, Graf M, Notka F, Schödl T, Wagner R. 2010. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst. Synth. Biol.* 4:215-225.
- 25 31. Schallmeyer M, Floor RJ, Hauer B, Breuer M, Jekel PA, Wijma HJ, Dijkstra BW, Janssen DB. 2013. Biocatalytic and Structural Properties of a Highly Engineered Halohydrin Dehalogenase. *ChemBioChem* 14:870-881. A Schallmeyer A, Schallmeyer M, Wardenga R, *BIOspektrum*, 2013, 19, 7, pp. 816-817.

<110> Enzymicals AG

Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen

<120> Expansión de la familia de enzimas de halohidrina deshalogenasas: identificación de nuevas enzimas por extracción de base de datos

5 <130> EZY13965EP1

<160> 15

<170> PatentIn version 3.3

<210> 1

10 <211> 227

<212> PRT

<213> metagenoma marino

<400> 1

ES 2 832 451 T3

Met Ser Lys Arg Leu Glu Gly Lys Arg Val Leu Val Thr Gln Ser Asn
 1 5 10 15
 Asn Tyr Met Gly Pro Ala Thr Val Glu Leu Phe Glu Lys Glu Gly Ala
 20 25 30
 Ile Val Thr Ala Asp Ser Ser Asp Leu Thr Glu Thr Glu Lys Cys Gln
 35 40 45
 Asp Ile Val Glu Ser Ala Gly Ile Ile Asp Ile Leu Ile Ala Asn Leu
 50 55 60
 Ala Ser Glu Asn Phe Ser Gly Ile Pro Thr Ser Asp Leu Thr Asp Glu
 65 70 75 80
 Asp Trp Asn Thr Thr Phe Asp Val Met Val His Pro Leu His Lys Leu
 85 90 95
 Ser Arg Ala Val Ile Pro Gln Met Val Glu Arg Gln Ala Gly Lys Ile
 100 105 110
 Ile Val Tyr Gly Ser Ala Ser Ala Leu Lys Gly Met Arg Thr Leu Ala
 115 120 125
 Ala Tyr Ser Ala Ala Arg Ala Ala Gln Val Gly Tyr Val Gln Ser Leu
 130 135 140
 Gly Val Glu Leu Ala Pro His Asn Ile Gln Val Asn Leu Ile Ala Gln
 145 150 155 160
 Asn Tyr Val Glu Asn Pro Ile Tyr Tyr Pro Glu Ser Leu Arg Ser Gln
 165 170 175
 Glu Lys Phe Gln Asn Ser Leu Lys Arg Gln Val Pro Leu Gly Arg Leu
 180 185 190
 Ala Arg Ala Glu Glu Asp Ala Gln Phe Ala Leu Phe Leu Ala Ser Asn
 195 200 205
 Glu Ser Asp Phe Phe Val Gly Gln Ala Ile Pro Phe Ser Gly Gly Trp
 210 215 220
 Val Gln Arg
 225

<210> 2

<211> 224

<212> PRT

<213> metagenoma marino

5

<400> 2

```

Met Gly Arg Leu Asp Gly Lys Arg Val Leu Val Thr Gln Ala Asp Asp
 1             5             10             15

Tyr Met Gly Pro Val Thr Leu Glu Val Phe Ala Glu Glu Gly Ala Glu
      20             25             30

Val Ile Ala Asp Asn Ser Asp Leu Thr Asp Pro Ser Arg Ala Ala Ala
      35             40             45

Leu Ile Glu Glu Thr Gly His Ile Asp Val Leu Val Ala Asn Leu Ala
      50             55             60

Ala Pro Ala Asn Leu Gly Val Ala Ala Ala Asp Met Pro Asp Asp Ile
65             70             75             80

Trp Gln Thr Met Phe Asp Val Met Val His Pro Leu His Arg Leu Thr
      85             90             95

Lys Ala Val Leu Pro Gln Met Val Glu Arg Gln Lys Gly Lys Ile Leu
      100            105            110

Val Tyr Gly Ser Ala Thr Gly Val Lys Gly Met Ala Gly Ile Thr Ala
      115            120            125

Tyr Ser Ala Ala Arg His Ala Gln Val Gly Tyr Val Arg Ser Thr Gly
      130            135            140

Ala Glu Ile Ala Gly His Asn Ile Gln Met Asn Leu Ile Ala Gln Asn
145            150            155            160

Phe Val Glu Asn Pro Val Tyr Phe Pro Pro Gln Phe Thr Glu Thr Pro
      165            170            175

Glu Phe Lys Glu Leu Leu Lys Gly Val Pro Ala Gly Arg Leu Ala Thr
      180            185            190

Ala Arg Glu Asp Ala Leu Phe Ala Val Phe Leu Ala Ser Asp Glu Ser
      195            200            205

Asp Phe Phe Val Gly Gln Ala Ile Pro Phe Ser Gly Gly Trp Ala Gln
      210            215            220

```

<210> 3

<211> 226

<212> PRT

5 <213> metagenoma marino

<400> 3

```

Met Asp Lys Arg Leu Ala Gly Lys Arg Val Leu Ile Thr Gln Ala Glu
 1          5          10          15

Asp Tyr Met Gly Pro Ala Ile Thr Glu Leu Phe Ala Glu His Gly Ala
 20          25          30

Glu Ile Ile Ala Asp Thr Arg Asp Leu Thr Glu Asp Gly Ala Val Glu
 35          40          45

Ser Leu Ile Lys Ser Ala Gly Glu Ile Asp Val Leu Ile Ala Asn Leu
 50          55          60

Ala Ala Pro Ala His Leu Gly Leu Ser Val Thr Asp Thr Asp Asp Thr
 65          70          75          80

Thr Trp Glu Thr Ala Phe Asp Val Met Val His Pro Leu His Arg Ile
 85          90          95

Phe Arg Ser Val Leu Pro Ala Met Tyr Glu Arg Lys Arg Gly Lys Ile
100          105          110

Val Val Ile Gly Ser Ala Thr Gly Leu Lys Ala Leu Glu Gly Val Val
115          120          125

Ala Tyr Gly Ala Ala Arg His Ala Gln Val Gly Tyr Val Arg Ala Ala
130          135          140

Gly Ile Glu Ala Ala Arg His Asn Val Gln Ile Asn Leu Ile Ala Gln
145          150          155          160

```

ES 2 832 451 T3

Asn Phe Val Glu Asn Pro Ala Tyr Phe Pro Pro Glu Phe Thr Ala Ser
165 170 175

Asn Glu Phe Lys Gln Leu Leu Ser Ala His Val Pro Leu Gly Arg Leu
180 185 190

Ala Thr Ala Arg Glu Asp Ala Leu Phe Ala Leu Phe Leu Ala Ser Asp
195 200 205

Glu Ser Asp Phe Phe Val Gly Gln Ala Ile Pro Phe Ser Gly Gly Trp
210 215 220

Ala Gln
225

<210> 4

<211> 226

<212> PRT

5 <213> metagenoma marino

<400> 4

ES 2 832 451 T3

Met Gly Asp Arg Leu Ala Gly Lys Arg Val Leu Val Thr Gln Ala Asp
1 5 10 15

Asn Tyr Met Gly Pro Ala Thr Ile Glu Leu Phe Thr Glu Glu Gly Ala
20 25 30

Glu Val Leu Ala Asp His Ser Asp Leu Thr Val Ala Gly Arg Cys Ser
35 40 45

Ala Leu Ile Lys Glu His Gly Arg Ile Asp Val Leu Val Ala Asn Leu
50 55 60

Ala Ser Pro Asn Phe Thr Gly Ile Pro Ala Thr Glu Leu Ala Asp Glu
65 70 75 80

Asp Trp His Cys Thr Phe Asn Met Met Val His Pro Leu His Gln Leu
85 90 95

Cys Gln Ala Val Val Pro Gln Met Ile Glu Arg Gln Ala Gly Lys Ile
100 105 110

Val Val Phe Gly Ser Ala Thr Ala Leu Lys Gly Met Pro Thr Val Thr
115 120 125

Ala Tyr Ser Ala Ala Arg Ala Ala Gln Ala Gly Tyr Val Arg Ser Leu
130 135 140

Gly Val Glu Ile Ala Lys His Asn Ile Gln Ala Asn Leu Ile Ala Gln
145 150 155 160

Asn Tyr Val Glu Asn Pro Val Tyr Tyr Pro Pro Glu Leu Arg Glu Lys
165 170 175

Glu Ser Phe Gln Lys Ser Leu Ala Arg Gln Val Pro Leu Gly Arg Leu
180 185 190

Ala Thr Ala Arg Glu Asp Ala Leu Phe Ala Leu Phe Leu Ala Ser Asp
195 200 205

Glu Ser Asp Phe Phe Val Gly Gln Ser Ile Ala Phe Ser Gly Gly Trp
210 215 220

Ala Gln
225

<210> 5

<211> 250

<212> PRT

<213> metagenoma marino

5

<400> 5

```

Met Ser Ser Asn Asp Leu Lys Gly Lys Arg Ile Leu Val Thr Gln Ala
 1          5          10          15

Asp Thr Phe Met Gly Pro Thr Leu Cys Glu Val Phe Thr Glu Lys Gly
 20          25          30

Ala Glu Val Ile Arg Asp Asn Gln Leu Leu Thr Glu Pro Thr Leu Pro
 35          40          45

Ala Gln Leu Val Ala Ala Ala Gly Ser Ile Asp Val Leu Val Ile Asn
 50          55          60

Leu Ala Ile Pro Ala Pro Phe Thr Lys Leu Glu Lys Val Asp Asp Asn
 65          70          75          80

Glu Trp Glu Ser Val Phe Ser Ala Val Val Asn Pro Met Pro Arg Leu
 85          90          95

Val Ser Ala Val Leu Pro Gln Met Ile Glu Arg Gln Ser Gly Lys Ile
 100         105         110

Leu Val Met Gly Ser Ala Ser Ala Leu Arg Gly Met Lys Arg Ala Ser
 115         120         125

```

ES 2 832 451 T3

Thr Tyr Ser Ala Ala Arg Gly Ala Gln Leu Ala Tyr Val Gln Ala Ala
130 135 140

Gly Val Glu Met Ala Pro His Asn Ile Gln Ile Asn Ala Ile Ala Gln
145 150 155 160

Asn Phe Val Asp Asn Pro Thr Tyr Phe Pro Glu Glu Thr Lys Ala Asn
165 170 175

Pro Ala Phe Gln Asp Arg Leu Lys Arg Glu Val Pro Leu Gly Arg Leu
180 185 190

Val Ser Met Arg Glu Asp Ala Gln Phe Ala Ala Tyr Leu Cys Ser Asp
195 200 205

Ala Ala Asp Cys Phe Val Gly Gln Val Phe Pro Val Ser Gly Gly Trp
210 215 220

Ala Gln Leu Lys Asn Glu Gln His Gly Met Ile Ile Ser Phe Ile Asp
225 230 235 240

Ser Leu Ser Ala Lys Val Ile Arg Cys His
245 250

<210> 6

<211> 225

<212> PRT

5 <213> Limnohabitans sp. Rim28

<400> 6

ES 2 832 451 T3

Met	Ile	Asp	Leu	Arg	Gly	Gln	Arg	Ile	Leu	Val	Thr	Gln	Ala	Gln	Asp
1				5					10					15	
Phe	Met	Gly	Pro	Ala	Leu	Cys	Gln	Glu	Leu	Arg	Ala	Cys	Gly	Ala	Glu
			20					25					30		
Val	Ile	Ala	Asp	Asp	Arg	Val	Leu	Thr	Ala	Pro	Gln	Asp	Ala	Gln	Ala
		35					40					45			
Met	Ile	Lys	Glu	Ala	Gly	Pro	Leu	Ala	Ala	Leu	Val	Ile	Asn	Leu	Ala
	50					55					60				
Leu	Pro	Ala	Pro	Ser	Thr	Pro	Val	Thr	Gln	Ile	Asp	Glu	Ala	Glu	Trp
65					70					75					80
His	Gln	Val	Phe	Glu	Val	Met	Val	Asp	Pro	Leu	Pro	Arg	Leu	Val	Arg
				85					90					95	

ES 2 832 451 T3

Ala Val Val Pro Gly Met Lys Ala Arg Gly Gly Gly Lys Ile Val Val
100 105 110

Met Gly Ser Ala Ser Ala Leu Arg Gly Met Lys Arg Ala Ala Ser Tyr
115 120 125

Ser Ala Ala Arg Gly Ala Gln Leu Ala Tyr Val Gln Ala Ala Gly Val
130 135 140

Glu Leu Ala Pro Asp Asn Ile Gln Leu Asn Ala Val Ala Gln Asn Phe
145 150 155 160

Val Asp Asn Pro Thr Tyr Phe Pro Pro Glu Val Gln Ala Asn Pro Arg
165 170 175

Phe Gln Glu Arg Leu Lys Arg Glu Val Pro Leu Gly Arg Leu Val Lys
180 185 190

Ala Glu Glu Asp Ala Arg Phe Val Ala Tyr Leu Cys Ser Asp Ala Ala
195 200 205

Ser Cys Phe Val Gly Gln Val Phe Pro Met Ser Gly Gly Trp Ala Val
210 215 220

Arg
225

<210> 7

<211> 227

5 <212> PRT

<213> *Curvibacter lanceolatus*

<400> 7

ES 2 832 451 T3

```

Met Ser Ser Gln Cys Leu Ala Gly Leu Arg Val Leu Ile Thr Gln Ala
1          5          10          15

Asn Glu Phe Met Gly Pro Thr Leu Cys Glu Val Phe Ala Glu Gln Gly
          20          25          30

Ala Val Val Leu Ala Asp Asp Gly Pro Leu Thr Asp Pro Gln Ala Pro
          35          40          45

Ala Arg Leu Val Ala Gly His Gly Pro Ile Asp Val Leu Val Ala Asn
          50          55          60

Leu Ala Val Pro Ala Pro Ser Thr Pro Ala His Gln Val Ser Glu Gln
65          70          75          80

Glu Trp Arg Asp Thr Phe Ala Ala Leu Val Asp Pro Leu Pro Arg Leu
          85          90          95

Cys Gln Ala Val Leu Pro Ala Met Met Ala Arg Arg Ser Gly Arg Ile
          100          105          110

Leu Val Met Gly Ser Ala Ala Ala Leu Arg Gly Met Lys Arg Ser Ser
          115          120          125

Thr Tyr Ser Ala Ala Arg Gly Ala Gln Leu Ala Tyr Val Gln Ala Leu
          130          135          140

Gly Val Glu Val Ala Ser Tyr Asn Ile Gln Val Asn Ala Ile Ala Gln
145          150          155          160

Asn Leu Val Asp Asn Pro Thr Tyr Phe Pro Pro Glu Val Gln Ala Asp
          165          170          175

Pro Arg Phe Gln Glu Arg Leu Arg Arg Glu Val Pro Leu Gly Arg Leu
          180          185          190

Val Ser Ala Arg Glu Asp Ala Glu Phe Ala Ala Tyr Leu Cys Ser Ala
          195          200          205

Pro Ala Ala Cys Phe Val Gly Gln Val Phe Pro Val Cys Gly Gly Trp
          210          215          220

Val Gly Arg
225

```

<210> 8

<211> 225

<212> PRT

<213> metagenoma marino

<400> 8

Met Arg Leu Glu Asn Lys Arg Val Met Val Thr Gln Ser Asp Asp Tyr
1 5 10 15

Met Gly Pro Ala Ile Thr Ser Leu Phe Ser Thr Glu Gly Ala Gln Val
20 25 30

Thr Thr Arg Glu Lys Pro Val Pro Thr Gly Lys Ala Phe Ser Thr Trp
35 40 45

Val Arg Glu Met Pro Val Tyr Asp Val Val Val Ala Asn Leu Ala His
50 55 60

5

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

<210> 9

<211> 226

<212> PRT

5 <213> metagenoma marino

<400> 9

Met Thr Arg Leu Glu His Lys Lys Val Leu Ile Thr Gln Ser Asp Asp
 1 5 10 15
 Tyr Met Gly Pro Ala Ile Ala Asp Leu Phe Ala Ala Glu Ala Ala Arg
 20 25 30
 Val Thr Ala Arg Pro Gly Leu Val Pro Phe Gly Thr Gln Phe Ala Arg
 35 40 45

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

<210> 10

<211> 222

5 <212> PRT

<213> gamma proteobacterium IMCC3088

<400> 10


```

Met Ser Gln Phe Ser Gly Lys Ser Val Trp Val Thr Ser Ala Asp Arg
1           5           10           15

Tyr Met Gly Pro Ser Ile Ala Asp Glu Phe Glu Arg Leu Gly Ala Ile
          20           25           30

Val Thr Arg Asp Met His Val Leu Tyr Asp Asp His Tyr Leu Arg Glu
          35           40           45

Thr Leu Ala Glu Ile Pro Asp Ile Val Ile Ala Asn Leu Ala Glu Pro
          50           55           60

Pro Arg Lys Asp Ala Leu Glu Ala Ile Gln Asp Asp Asp Trp Asn Leu
65           70           75           80

Leu Phe Asp His Leu Val His Pro Leu Met Arg Ile Val Arg His Val
          85           90           95

Ser Gly Pro Met Lys Ala Arg Gly His Gly Lys Ile Val Ala Ile Thr
          100          105          110

Ser Ala Ala Pro Leu Arg Gly Ile Pro Phe Ala Ser Gly Tyr Cys Ala
          115          120          125

Ala Arg Gly Ala Gln Asn Ala Phe Ile Lys Gly Ala Gly Leu Glu Leu
          130          135          140

Ala Lys Phe Gly Val Gln Ala Asn Ala Ile Gly Gln Asn Tyr Ile Glu
145          150          155          160

Asn Asp Thr Tyr Tyr Pro Pro Glu Leu Met Gln Asp Pro Arg Phe Ile
          165          170          175

Ser Asn Leu Ser Ser Gln Val Pro Thr Lys Lys Val Gly Arg Gly Leu
          180          185          190

Glu Thr Ala Lys Leu Ala Ala Tyr Leu Ala Asp Pro Asp Val Glu His
          195          200          205

Val Val Gly Gln Ile Ile Pro Leu Ala Gly Gly Trp Thr Thr
          210          215          220

```

<210> 11

<211> 226

5 <212> PRT

<213> gamma proteobacterium HTCC2143 marina

<400> 11

```

Met Arg Tyr Ser Asn Lys Arg Val Leu Val Thr Thr Ser Asn Arg Tyr
 1              5              10              15

Met Gly Pro Asp Ile Ala Thr Arg Phe Glu Glu Glu Gly Ala Glu Val
      20              25              30

Ile Arg Ser Glu Asn Arg Leu Leu Asp Gln Gln Ser Val Asp Ala Leu
      35              40              45

Ile Glu Ser Val Gly Glu Val Asp Val Leu Val Ala Asn Phe Ala Glu
      50              55              60

Pro Pro Arg Thr Ala Pro Val Glu Lys Ile Ser Asp Glu Asp Trp Phe
65              70              75              80

Ile Leu Phe Asp Thr Leu Val His Pro Leu Met Arg Thr Val Arg Gly
      85              90              95

Phe Ala Gly Val Met Lys Glu Arg Arg Arg Gly Lys Ile Ile Ala Val
      100             105             110

Thr Ser Ala Ala Pro Leu Arg Gly Ile Pro Asn Ala Ala Gly Tyr Cys
      115             120             125

Ala Ala Arg Gly Ala Gln Asn Ala Phe Val Arg Ala Val Gly Leu Glu
      130             135             140

Leu Ala Lys Ser Asn Val Gln Phe Asn Ala Ile Ala Gln Asn Tyr Ile
145             150             155             160

Ser Asn Asp Thr Tyr Tyr Pro Asp Asp Phe Val Ala Asn Glu Lys Phe
      165             170             175

Lys Ala His Leu Lys Asp Phe Val Pro Thr Lys Lys Val Ala Asp Gly
      180             185             190

Ala Glu Thr Ala Glu Leu Ala Leu Tyr Leu Ala Ser Asp Leu Cys Thr
      195             200             205

His Leu Val Gly Gln Val Ile Pro Phe Ser Gly Gly Trp Ala Thr Thr
      210             215             220

Thr Gly
225

```

5

<210> 12

<211> 219

<212> PRT

<213> metagenoma marino

<400> 12

5

```

Met Ser Arg Arg Thr Ala Leu Val Thr Cys Ile Asp Ser Tyr Met Gly
1          5          10          15

Pro Ala Ile Ser Glu Lys Phe Glu Glu Ile Gly Ile Asn Val Ile Ala
          20          25          30

Gly Pro Ser Asp Leu His Ser Glu Ala Gln Ile Asn Lys Leu Val Ala
          35          40          45

Asp Ala Gly His Ile Asp Ile Leu Val Ala Asn Leu Ala Glu Pro Pro
          50          55          60

Met Thr Gly Pro Val Gln Asn Ile Asp Ser Asp Asn Trp Asn Leu Leu
65          70          75          80

Phe Asp Ser Leu Val His Pro Leu Met Tyr Leu Val Arg Ala Val Thr
          85          90          95

Pro Gln Met Leu Glu Arg Lys Ser Gly Lys Ile Ile Ala Ile Thr Ser
          100          105          110

Ala Ala Pro Leu Arg Gly Ile Pro Asn Asn Ala Ala Tyr Cys Ala Ala
          115          120          125

Arg Gly Ala Gln Asn Gly Phe Ile Lys Ala Ala Gly Leu Glu Leu Ala
          130          135          140

Arg Ser Asn Ile Gln Val Asn Ala Ile Ala Gln Asn Tyr Ile Asn Asn
145          150          155          160

Asn Thr Tyr Tyr Pro Asp Asp Ile Leu Asp Asn Gln Lys Phe Leu Asn
          165          170          175

His Val Lys Arg Asn Val Pro Ser Asn Lys Ile Gly Ala Ala Ser Glu
          180          185          190

Thr Ala Gln Leu Ala Ala Tyr Leu Ser Arg Glu Asp Cys Arg His Met
          195          200          205

Val Gly Gln Val Ile Pro Leu Ala Gly Gly Trp
          210          215

```

<210> 13

<211> 219

<212> PRT

5 <213> metagenoma marino

<400> 13

Met	Ala	Val	Leu	Val	Thr	His	Gly	Asp	Arg	Tyr	Met	Gly	Pro	Ala	Ile
1				5					10					15	

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

Gln Pro Thr Ser Ala Glu Ala Ala Glu Ala Leu Val Glu Gly Gln Glu
35 40 45

Ala Leu Asp Ile Leu Ile Ala Asn Leu Ala Tyr Pro Pro Thr Pro Ala
50 55 60

Leu Val Asp Ala Ile Asp Ser Asp Gly Trp His Ala Leu Phe Asp His
65 70 75 80

Leu Val His Pro Leu Met Gly Leu Val Arg Gly Ala Ala Lys Arg Met
85 90 95

Lys Ala Gln Gly His Gly Arg Ile Ile Ala Ile Thr Ser Ala Ala Pro
100 105 110

Leu Arg Gly Ile Pro Lys Asn Gly Ala Tyr Cys Ala Ala Arg Gly Ala
115 120 125

Gln Asn Ala Phe Leu Arg Ala Ala Gly Leu Glu Leu Ala Arg Asp Gly
130 135 140

Ile Ala Val Ala Ala Ile Ala Gln Asn Tyr Ile Glu Asn Asp Thr Tyr
145 150 155 160

Tyr Pro Pro Glu Leu Leu Asp Asp Glu Arg Phe Leu Ala Arg Met Arg
165 170 175

Glu Val Val Pro Ala Gln Gln Ile Gly Ala Pro Ser Ala Thr Ala Asn
180 185 190

Leu Ala Tyr Phe Leu Ala Thr Glu Ala Glu Phe Met Pro Gly Gln Ile
195 200 205

Phe Pro Leu Ser Gly Gly Trp Thr Thr Thr Leu
210 215

<210> 14

<211> 231

<212> PRT

5 <213> bacteria no cultivada

<400> 14

Met Thr Glu Gln Pro Gln Lys Asn Gly Tyr Gly Leu Ser Gly Lys Arg
1 5 10 15

ES 2 832 451 T3

```

Val Val Ile Thr Gln Ala Ala Gly Phe Met Gly Pro Ser Leu Val Glu
    20                      25                      30

Ala Phe Ser Arg Glu Gly Ala Glu Val Ile Pro Asp His Arg Asp Leu
    35                      40                      45

Thr His Asp Lys Ala Ala Asp Asn Leu Val Ser Glu Phe Lys Glu Ile
    50                      55                      60

Asp Ile Leu Leu Ile Asn Leu Ala Ser Gln Arg Gln Arg Ile Glu Ala
    65                      70                      75                      80

Thr Glu Ile Ser Asp Gln Gln Phe Leu Gln Pro Phe Glu Glu Met Val
    85                      90                      95

Tyr Pro Leu Phe Arg Leu Gly Arg Ser Val Leu Pro Gln Met Ile Ala
    100                     105                     110

Arg Arg Arg Gly Lys Ile Ile Val Ile Gly Ser Ala Ala Pro Leu Arg
    115                     120                     125

Pro Phe Ala Asn Ala Thr Gly Tyr Ala Thr Ala Arg Gly Ala Gln Leu
    130                     135                     140

Ala Trp Val Lys Ala Val Gly Ala Glu Val Ala Gln His Asn Val Gln
    145                     150                     155                     160

Val Asn Gly Ile Ala Gln Ile Phe Val Glu Asn Gln Glu Tyr Phe Pro
    165                     170                     175

Pro Ala Tyr Leu Gln Thr Asp Glu Phe Lys Gln Arg Ile Ala Gln Val
    180                     185                     190

Pro Ala Gly Arg Leu Gly Ser Ala Ala Glu His Ala Ala Leu Ser Leu
    195                     200                     205

Phe Leu Ala Ser Asp Gln Cys Asn Phe Ile Ser Gly Gln Val Val Pro
    210                     215                     220

Phe Ala Gly Gly Trp Thr Thr
    225                     230

```

<210> 15

<211> 227

<212> PRT

5 <213> bacteria marina no cultivada Ant4E12

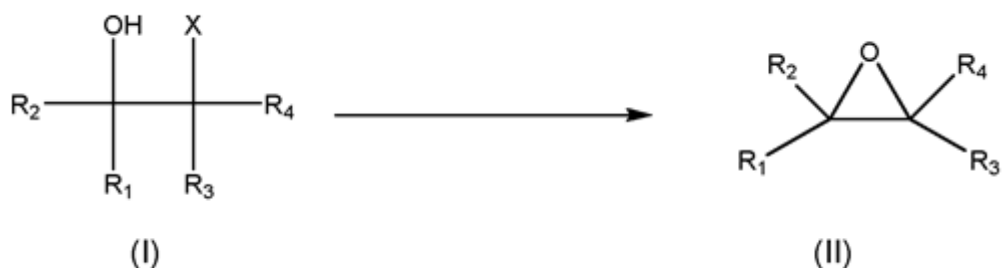
<400> 15

ES 2 832 451 T3

Met	Ser	Gln	Arg	Leu	Ala	Asn	Lys	Arg	Val	Val	Val	Thr	Asn	Ala	Asn	1	5	10	15
Asp	Phe	Met	Gly	Ala	Pro	Thr	Ile	Glu	Val	Phe	Glu	Glu	Glu	Gly	Ala	20	25	30	
Glu	Val	Ile	Ala	Asp	Tyr	Ser	Asn	Leu	Met	Asp	Ser	Asp	Ala	Ala	Ser	35	40	45	
Thr	Leu	Ile	Glu	Met	Val	Gly	Gln	Val	Asp	Val	Leu	Val	Val	Asn	Leu	50	55	60	
Ala	Ala	Pro	Ser	Phe	Gly	Gly	Ile	Pro	Thr	His	Glu	Leu	Asn	Asp	Glu	65	70	75	80
Asn	Phe	Ala	Glu	Cys	Phe	Asp	Val	Met	Val	Tyr	Pro	Leu	His	Arg	Leu	85	90	95	
Thr	Arg	Ala	Val	Leu	Pro	Gln	Met	Ile	Glu	Arg	Lys	Ala	Gly	Lys	Ile	100	105	110	
Val	Val	Tyr	Gly	Ser	Ala	Thr	Ala	Leu	Lys	Gly	Met	Lys	Thr	Leu	Ala	115	120	125	
Ala	Tyr	Ser	Ala	Ala	Arg	Gly	Ala	Gln	Val	Ser	Tyr	Val	Gln	Ser	Val	130	135	140	
Gly	Ile	Glu	Val	Ala	Pro	Gln	Asn	Ile	Gln	Met	Asn	Leu	Ile	Ala	Gln	145	150	155	160
Asn	Tyr	Val	Glu	Asn	Pro	Asp	Tyr	Tyr	Pro	His	Glu	Leu	Thr	Ser	Gln	165	170	175	
Glu	Ser	Phe	Lys	Ser	Ser	Leu	Lys	Arg	Gln	Val	Pro	Leu	Gly	Arg	Leu	180	185	190	
Ala	Thr	Ser	Arg	Glu	Asp	Ala	Gln	Leu	Ala	Ala	Phe	Leu	Ala	Ser	Asn	195	200	205	
Glu	Ser	Asp	Phe	Phe	Val	Gly	Gln	Ala	Ile	Pro	Phe	Ala	Gly	Gly	Trp	210	215	220	
Val	Gln	Arg														225			

REIVINDICACIONES

1. Uso de una biomolécula que tiene la secuencia de proteínas o secuencia de ADN codificante de HheD3 que tiene NCBI, número de acceso GenBank ABM93639 como halohidrina deshalogenasa en la conversión de al menos una parte de la cantidad de una halohidrina de fórmula (I) en un epóxido de fórmula (II)



5

en las que

X es un átomo de halógeno seleccionado de Cl, Br, I;

10

R₁, R₂, R₃, R₄ son independientemente hidrogeno, halógeno (F, Cl, Br, I) o alquilo C1-C12, alquenilo C2-C12, alquinilo C2-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o arilo C5-C12 sustituido o no sustituido, o

alquil-C1-C3-C(=O)-O-R₅ o -alquil-C1-C3-O-R₅, con

15

R₅ seleccionado de hidrogeno, alquilo C1-C6, alquenilo C2-C6, alquinilo C2-C6, alcoxi C1-C6, alquilarilo C5-C13, arilalquilo C5-C13, ariloxi C5-C13, arilalcoxi C5-C13, heteroalquilo C5-C13, arilo C5-C13, heteroarilo C5-C13, cicloalquilo C3-C12, cicloalquenilo C3-C12, cicloalquinilo C6-C12 o heteroalquilo cíclico C5-C13, en el que en el caso de más de un sistema anular alifático o aromático, los sistemas anulares están condensados o separados y R₅ puede tener al menos un sustituyente adicional R₆ con

20

R₆ seleccionado de hidrógeno, hidroxilo, -C(=O)-NH₂, halógeno (F, Cl, Br, I), -alquilo-C1-C3, -C(=O)-alquilo-C1-C3, -C(=O)-alquilo-C1-C3, -NH-C(=O)-alquilo-C1-C3, -alquil-C1-C3-O-alquilo-C1-C3, alquenilo C2-C3, -O-alquenilo C2-C3, cicloalquilo C5-C7, heteroalquilo cíclico C5-C7, arilo C5-C7 o heteroarilo C5-C7

y/o

25

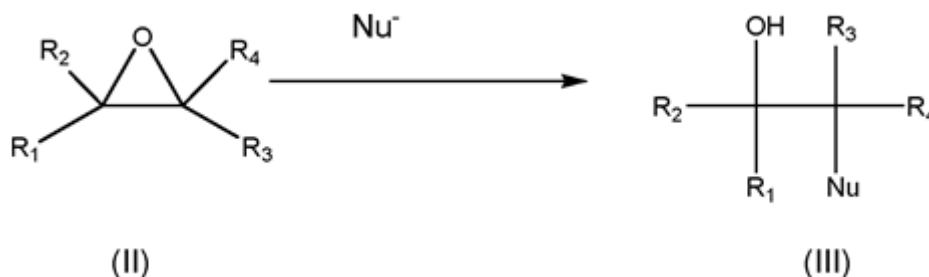
dos de R₁, R₂, R₃, R₄ forman en conjunto una estructura cíclica seleccionada de cicloalquilo C3-C12, cicloalquenilo C4-C12, cicloalquinilo C6-C12, heteroalquilo cíclico C3-C12, heteroalquenilo cíclico C4-C12 o heteroalquinilo cíclico C6-C12, preferentemente un grupo cicloalquilo C4-C8, o cicloalquenilo C4-C8, más preferentemente un grupo cicloalquilo C5-C7, o cicloalquenilo C5-C7, con máxima preferencia un grupo cicloalquilo C5-C6, o cicloalquenilo C5-C6 en el que en el caso de más de un sistema anular, los sistemas anulares están condensados o separados, y

en el que el número de átomos de C indicado para la estructura cíclica siempre incluye los 2 átomos de C de la estructura de halohidrina que portan el OH y el X, y

30

en el que cada grupo cíclico en R₁, R₂, R₃, R₄ o estructura cíclica formada por dos de R₁, R₂, R₃, R₄ opcionalmente tiene al menos un sustituyente adicional diferente de hidrógeno, preferentemente seleccionado de R₆, con máxima preferencia halógeno (F, Cl, Br, I) o alquilo C1-C3 (R₇).

2. Uso de una biomolécula que tiene la secuencia de proteínas o secuencia de ADN codificante de HheD3 que tiene NCBI, número de acceso GenBank ABM93639 como halohidrina deshalogenasa en la conversión de al menos una parte de un epóxido de fórmula (II) en un alcohol de fórmula (III)



en las que

R₁, R₂, R₃, R₄ son independientemente hidrógeno, halógeno (F, Cl, Br, I) o alquilo C1-C12, alqueno C2-C12, alquino C2-C12, alcoxi C1-C12, cicloalquilo C3-C12, cicloalqueno C3-C12, cicloalquino C6-C12 o arilo C5-C12 sustituido o no sustituido, o

-alquil-C1-C3-C(=O)-O-R₅ o -alquil-C1-C3-O-R₅, con

R₅ seleccionado de hidrógeno, alquilo C1-C6, alqueno C2-C6, alquino C2-C6, alcoxi C1-C6, alquilarilo C5-C13, arilalquilo C5-C13, ariloxi C5-C13, arilalcoxi C5-C13, heteroalquilo C5-C13, arilo C5-C13, heteroarilo C5-C13, cicloalquilo C3-C12, cicloalqueno C3-C12, cicloalquino C6-C12 o heteroarilo cíclico C5-C13, en el que en el caso de más de un sistema anular alifático o aromático, los sistemas anulares están condensados o separados y R₅ puede tener al menos un sustituyente adicional R₆ con

R₆ seleccionado de hidrógeno, hidroxilo, -C(=O)-NH₂, halógeno (F, Cl, Br, I), -alquilo C1-C3, -C(=O)-alquilo-C1-C3, -C(=O)-O-alquilo-C1-C3, -NH-C(=O)-alquilo-C1-C3, -alquil-C1-C3-O-alquilo-C1-C3, alqueno C2-C3, -O-alqueno-C2-C3, cicloalquilo C5-C7, heteroalquilo cíclico C5-C7, arilo C5-C7 o heteroarilo C5-C7

y/o

dos de R₁, R₂, R₃, R₄ forman en conjunto una estructura cíclica seleccionada de cicloalquilo C3-C12, cicloalqueno C4-C12, cicloalquino C6-C12, heteroalquilo cíclico C3-C12, heteroalqueno cíclico C4-C12 o heteroalquino cíclico C6-C12, preferentemente un grupo cicloalquilo C4-C8, o cicloalqueno C4-C8, más preferentemente un grupo cicloalquilo C5-C7, o cicloalqueno C5-C7, con máxima preferencia un grupo cicloalquilo C5-C6, o cicloalqueno C5-C6 en el que en el caso de más de un sistema anular, los sistemas anulares están condensados o separados, y en el que el número de átomos de C indicado para la estructura cíclica siempre incluye los 2 átomos de C de la estructura de halohidrina que portan el OH y el X, y

en el que cada grupo cíclico en R₁, R₂, R₃, R₄ o estructura cíclica formada por dos de R₁, R₂, R₃, R₄ opcionalmente tiene al menos un sustituyente adicional diferente de hidrógeno, preferentemente seleccionado de R₆, con máxima preferencia halógeno (F, Cl, Br, I) o alquilo C1-C3 (R₇) y

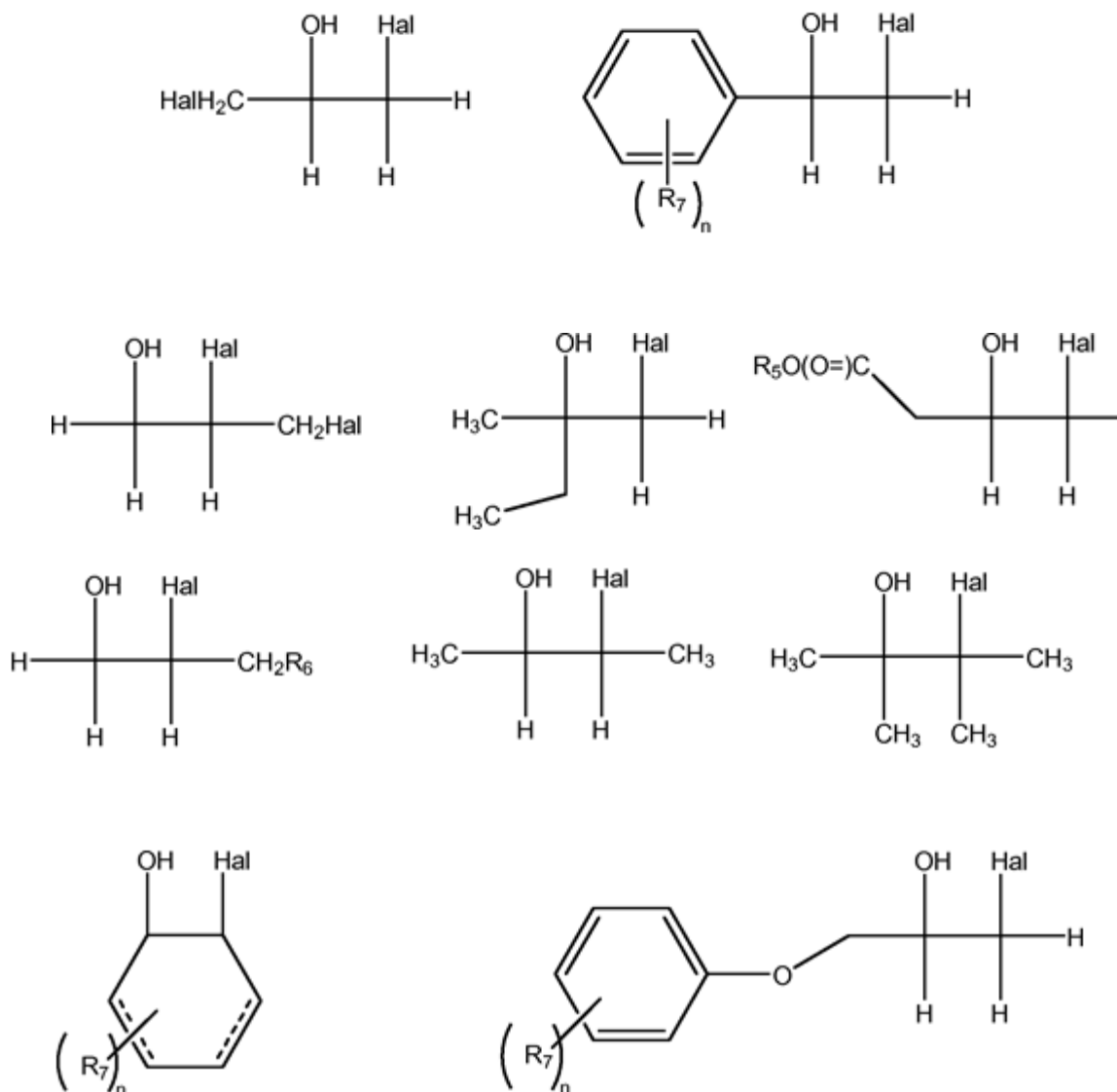
Nu es seleccionado de R'NO₂, N₃, NO₂, SCN, OCN, CN, halógeno (F, Cl, Br, I) y

R''CO₂, en el que

R' es seleccionado de alquilo C1-C4, -alqueno, preferentemente CH₃,

R'' es seleccionado de hidrógeno, alquilo C1-C4, -alqueno, preferentemente hidrógeno, siendo Nu⁻ un anión de Nu.

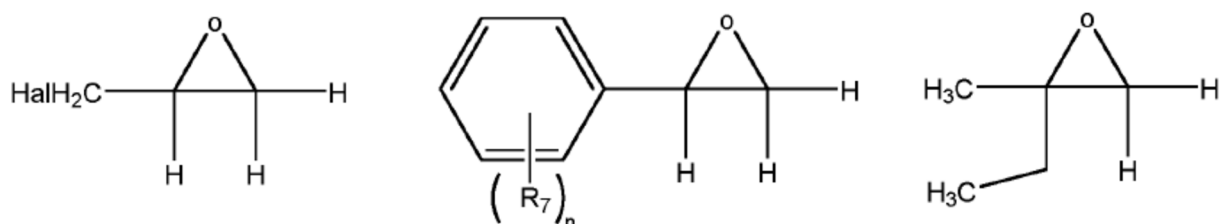
3. Uso de la biomolécula como halohidrina deshalogenasa en la conversión de al menos una parte de la cantidad de halohidrina de fórmula (I) en un epóxido de fórmula (II) y/o en la conversión de al menos una parte de un epóxido de fórmula (II) en un alcohol de fórmula (III) de acuerdo con la reivindicación 1 o 2, **caracterizado porque** cada uno de R₁, R₂, R₃, R₄, que no forma una estructura cíclica con otro de R₁, R₂, R₃, R₄ puede tener al menos un átomo de carbono reemplazado por un heteroátomo.
4. Uso de la biomolécula como halohidrina deshalogenasa en la conversión de al menos una parte de la cantidad de halohidrina de fórmula (I) en un epóxido de fórmula (II) de acuerdo con la reivindicación 1 en el que la halohidrina de fórmula (I) es seleccionada de



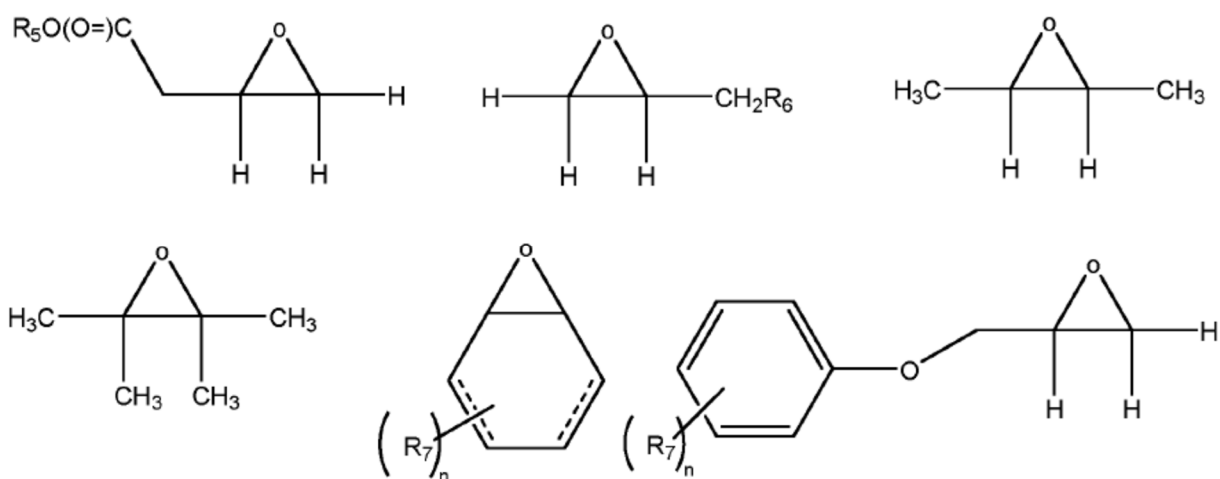
5

en las que Hal es Cl o Br y en el caso de más de un Hal es preferentemente idéntico en sustancia, y R_7 es seleccionado de halógeno (F, Cl, Br, I) o alquilo C1-C3 y cada --- independientemente representa un enlace doble o simple y n es cero o un número entero entre 1 y 4, preferentemente 0, 1 o 2.

- 10 5. Uso de la biomolécula como halohidrina deshalogenasa en la conversión de al menos una parte del epóxido de fórmula (II) en un alcohol de fórmula (III) de acuerdo con la reivindicación 2 en el que el epóxido de fórmula (II) es seleccionado de



15



en las que Hal es Cl o Br y en el caso de más de un Hal es preferentemente idéntico en sustancia, y R_7 es seleccionado de hidrógeno, halógeno (F, Cl, Br, I) o alquilo C1-C3 y cada --- representa independientemente un enlace doble o simple y n es cero o un número entero entre 1 y 4, preferentemente 0, 1 o 2.

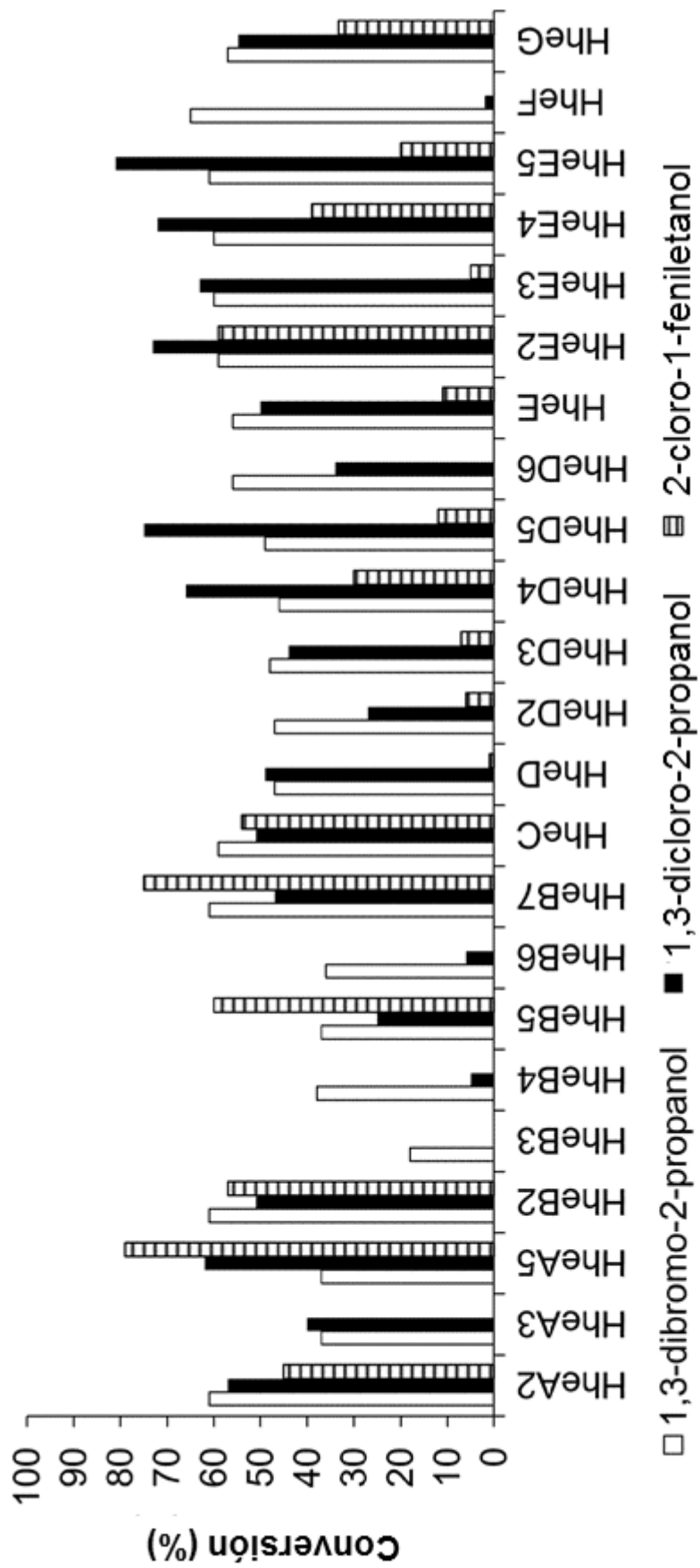


Fig. 1