UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

DIVISION DE ESTUDIOS DE POSGRADO E INVESTIGACION
DEPTO. BIOQUIMICA, FACULTAD DE MEDICINA

DIVERSIDAD, TAXONOMÍA Y EVOLUCIÓN DE LAS DESHIDROGENASAS/REDUCTASAS DE CADENA MEDIA

T E S I S

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS BIOMEDICAS (AREA: BIOQUIMICA)

P R E S E N T A :

M. en C. HECTOR RIVEROS ROSAS

MEXICO, D.F. AGOSTO DEL 2004

Director de tesis: Dr. Enrique Piña Garza

Tesis Doctoral realizada en el
Laboratorio 34, de “Integración metabólica”
Departamento de Bioquímica,
Facultad de Medicina, UNAM

Comité tutoral:

Dr. Rafael Villalobos Molina
Depto. Farmacobiología, CINVESTAV-Sede Sur, México, D.F.

Dr. Juan Pablo Pardo Vázquez
Depto. Bioquímica, Fac. Medicina, UNAM. Cd. Universitaria, México, D.F.

Dr. Enrique Piña Garza
Depto. Bioquímica, Fac. Medicina, UNAM. Cd. Universitaria, México, D.F.

H.R.R. recibió apoyo como becario de la Dirección General de Estudios de Posgrado (DGEP), UNAM, y del Consejo Nacional de Ciencia y Tecnología (CONACyT), México.

El proyecto recibió financiamiento parcial de los siguientes donativos:

IN212695, IN214101, de la Dirección General Asuntos del Personal Académico, UNAM

34823-M del CONACyT, México

PAEP99-102359, PAEP2000-102360 de la DGEP, UNAM

Jurado:

Presidente:
Dr. Mario Calcagno Montans
Depto. Bioquímica, Fac. Medicina, UNAM.
Cd. Universitaria, México, D.F.

Secretario:
Dr. Enrique Piña Garza
Depto. Bioquímica, Fac. Medicina, UNAM.
Cd. Universitaria, México, D.F.

Vocal:
Dr. Diego González Halphen
Instituto de Fisiología Celular, UNAM.
Cd. Universitaria, México, D.F.

Vocal:
Dr. Julio Collado Vides
Centro de Investigaciones sobre Fijación del Nitrógeno, UNAM
Cuernavaca, Morelos.

Vocal:
Dr. Lorenzo Segovia Forcella
Instituto de Biotecnología, UNAM
Cuernavaca, Morelos

Suplente:
Dr. Rafael Villalobos Molina
Depto. Farmacobiología, CINVESTAV-Sede Sur, México, D.F.

Suplente:
Dr. Juan Pablo Pardo Vázquez
Depto. Bioquímica, Fac. Medicina, UNAM.
Cd. Universitaria, México, D.F.

A Adriana

Mi adorable y amada esposa

Por su incansable estímulo, apoyo y ayuda

fuente de inspiración de mi vida,

consuelo para mi alma

A nuestra hija Adriana Mariana

Quien sin proponerselo nos abrió los ojos

y nos cambio la vida

Promesa de vida, capullo de mi alma

A mis padres

Por su cariño, apoyo y ejemplo

A mis suegros

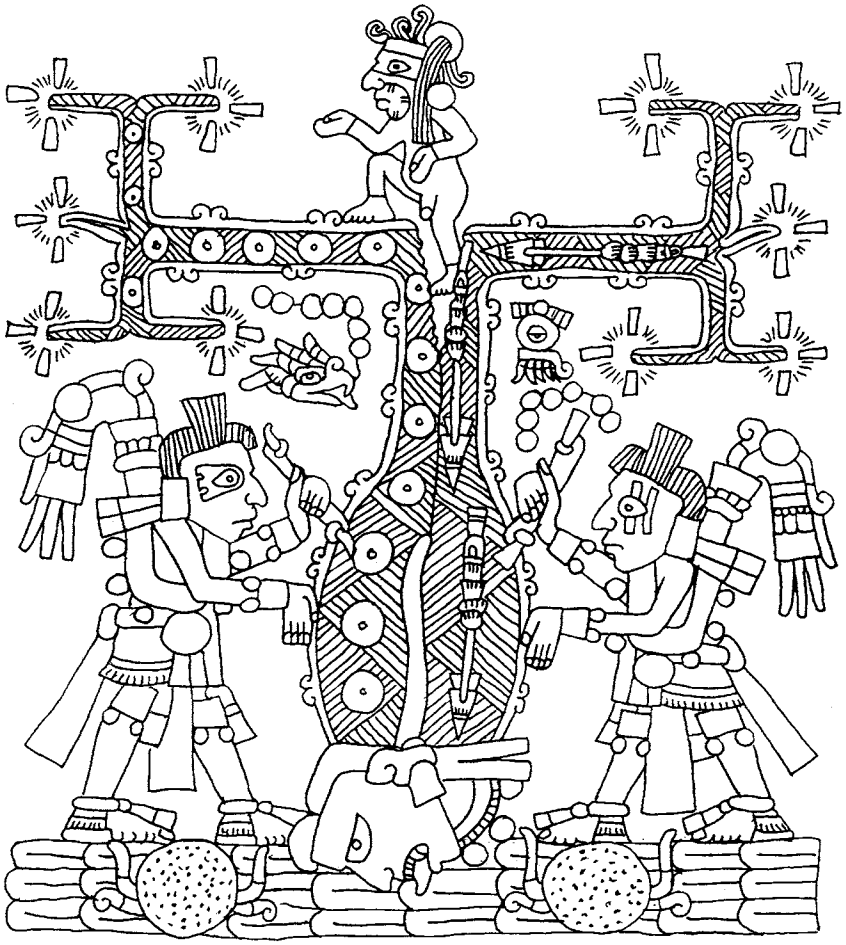
Por su desinteresado cariño y ejemplo

A todos a quienes estimo

y cuya lista es tan larga . . . .

que prefiero no comenzarla

Quiero expresar mi sincero agradecimiento a la Universidad Nacional Autónoma de México, por todo lo que representa, incluyendo por supuesto a sus valiosos miembros, quienes de una u otra manera, han influido positivamente en mi formación académica y a lo largo del desarrollo de este trabajo. En especial, quisiera hacer patente mi agradecimiento a las siguientes personas (y pido de antemano una disculpa por todas aquellas omisiones injustas que siempre se cometen al elaborar un texto como este). Así, quiero agradecer en primer lugar al Dr. Enrique Piña Garza, por su amistad, apoyo, libertad y confianza que me ha brindado en todos estos años que he tenido oportunidad de trabajar en su laboratorio; al Dr. Rafael Villalobos Molina, por su desinteresado apoyo, aliento y amistad, y por ser el primero en demostrarme que la bioquímica es lo más interesante de las ciencias biológicas; al Dr. Juan Pablo Pardo Vázquez, por sus atinadas observaciones, su sincera amistad y por compartir el gusto de emplear las matemáticas como una herramienta para resolver problemas; al Dr. Mario Calcagno Montans, también por una larga amistad, y el gusto de escucharle siempre un comentario interesante; al Dr. Diego González Halphen, por brindarme su apoyo y confianza, además de permitir contarme entre sus colaboradores; a los Drs. Julio Collado Vides y Lorenzo Segovia Forcella por la cuidadosa revisión del trabajo y sus valiosos comentarios y sugerencias; a los Drs. Alfredo Saavedra Molina y Martha Zentella Mayer, por la colaboración, apoyo y amistad que siempre nos han brindado; al Dr. Héctor Riveros Rotge (mi padre), por su ejemplo y por hacerme ver que la ciencia es divertida e inculcarme el placer que implica trabajar en lo que a uno le gusta; al Dr. Javier Garfias y Ayala, por su ejemplo, apoyo y consejo, en los más de 13 años de larga amistad y colaboración. Finalmente, quiero mencionar muy especialmente a la Biol. Adriana Julián Sánchez (mi esposa), por su singular e importante apoyo, fuente de inspiración, confianza y ayuda durante el desarrollo de este insospechadamente largo trabajo, amén del placer de compartir durante ya casi 15 años, cada instante y momento de nuestras vidas; el crecer y madurar al lado de una inteligente, bella y agradable mujer, es un placer que se disfruta lento, se vive intensamente y si se cultiva, se sublima.



Ideograma del árbol cósmico , que era entre las culturas mesoamericanas una de las formas de concebir el *axis mundi*, parte primordial del aparato cósmico, necesaria para el funcionamiento del ciclo de la vida y la muerte. La raíz del árbol primordial está representada por una cabeza antropomorfo-femenina, que a partir del caos, su cuerpo origina la tierra, el cielo y los árboles que comunican los diferentes niveles del cosmos. El tronco está formado por dos ramales en oposición complementaria, en donde se unen las fuerza calientes y las fuerzas frías que intervienen en la creación de los seres mundanos. El torzal o *malinalli* del tronco, expresa la idea de movimiento helicoidal propio de las fuerzas divinas. Así, a partir del desorden proteico o primordial, el árbol cósmico se proyecta desde su posición axial hacia los cuatro extremos del mundo para formar cuatro árboles que, como estructuras, soportan el cosmos. [La Figura corresponde a la lámina 37 del *Códice Vindobonensis* y fue tomada de: López Austin A. Ícono y mito: su convergencia. Ciencias (No. 74): 4-15 (abril-junio 2004)].

Nota: El ideograma aquí representado no tiene ninguna relación con el presente trabajo, aunque estética y conceptualmente son bellos, pero si se sustituyen los conceptos de árbol y cosmos por proteínas y célula, se obtiene una interpretación inesperadamente interesante.

indice

Resumen

Un detallado análisis *in silico*, de la estructura y función de 583 proteínas que pertenecen a la superfamilia de las deshidrogenasas/reductasas de cadena media (MDR), fue realizado mediante una extensa minería en las bases de datos públicas y a través del programa BLASTP, que utilizado en forma iterativa, permitió la identificación de todos los miembros conocidos de esta superfamilia. Considerando el patrón filogenético, así como las similitudes en secuencia y función de las proteínas que pertenecen a las MDR, éstas se clasificaron con base en tres diferentes categorías taxonómicas: 1) *subfamilias*, que consisten en grupos cerrados, conformados idealmente por conjuntos de proteínas ortólogas que desarrollan la misma función; 2) *familias*, cada una compuesta por un conjunto monofilético de subfamilias que poseen una identidad significativa al comparar sus secuencias entre ellas y que pueden o no, compartir sustratos y/o mecanismos de reacción; y 3) *macrofamilias*, cada una compuesta por un conjunto monofilético de familias de proteínas presentes en los tres dominios en que se distribuyen los seres vivos, e incluye al menos a una subfamilia de proteínas que posee una actividad perteneciente a una vía metabólica muy antigua. En este contexto, una *superfamilia*, es un grupo monofilético de familias (y/o macrofamilias) de proteínas, que comparten una identidad en secuencia escasamente detectable, pero que muestran una conformación tridimensional muy similar.

La superfamilia de las MDR está conformada por tres macrofamilias, con ocho familias y 49 subfamilias. Estas subfamilias exhiben una gran diversidad funcional que incluye miembros no-catalíticos, distribuidos en diferentes especies, grupos taxonómicos y compartimentos subcelulares. Esto resulta de una constante enzimogénesis y proteinogénesis dentro de cada reino y resalta la enorme plasticidad que poseen los miembros de la superfamilia de las MDR. Así, a través de la evolución un gran número de nuevas funciones fueron adquiridas por los miembros de las MDR en taxas específicos. De hecho, la generación de nuevas funciones desempeñadas por las proteínas, puede considerarse como la esencia misma de la evolución de proteínas. Los mecanismos de evolución de proteínas dentro de las MDR no están constreñidos a conservar la especificidad del sustrato y/o la química de la catálisis; en consecuencia, la diversidad funcional de las MDR es tan compleja como su diversidad en secuencias.

Las MDR conforman una superfamilia de proteínas muy antigua que existe desde el último ancestro común. Esta posee al menos dos (tal vez tres) diferentes actividades ancestrales relacionadas al metabolismo del formaldehído y la fermentación alcohólica. Los miembros eucarióticos de ésta superfamilia están más relacionados a sus homólogos en bacterias que a los de arquea. Por otra parte, la transferencia horizontal de genes, aparentemente frecuente dentro

de los diferentes grupos de bacterias, es un evento raro entre organismos pertenecientes a dominios diferentes.

Por último, la enorme plasticidad observada entre los miembros de las MDRs, no parece ser una propiedad exclusiva de esta superfamilia; ochenta por ciento de todas las proteínas secuenciadas hasta ahora, pertenece a alguna de las grandes superfamilias de proteínas ya reportadas. Un análisis exhaustivo de alguna otra de las principales superfamilias de proteínas seguramente mostrará resultados semejantes. De hecho, el sistema taxonómico propuesto en éste trabajo para las MDR (macrofamilia, familia y subfamilia) puede ser fácilmente extrapolado a otras superfamilias de proteínas.

Abstract

A comprehensive, structural and functional, *in silico* analysis of the medium-chain dehydrogenase/reductase (MDR) superfamily, including 583 proteins, was carried out by use of extensive database mining and the BLASTP program in an iterative manner to identify all known members of the superfamily. Based on phylogenetic, sequence, and functional similarities, the protein members of the MDR superfamily were classified into three different taxonomic categories: 1) subfamilies, consisting of a closed group containing a set of ideally orthologous proteins that perform the same function; 2) families, each comprising a cluster of monophyletic subfamilies that possess significant sequence identity among them and might share or not common substrates or mechanisms of reaction; and 3) macrofamilies, each comprising a cluster of monophyletic protein families with protein members from the three domains of life, which includes at least one subfamily member that displays activity related to a very ancient metabolic pathway. In this context, a superfamily is a group of homologous protein families (and/or macrofamilies) with monophyletic origin that shares at least a barely detectable sequence similarity, but showing the same 3D fold.

The MDR superfamily encloses three macrofamilies, with eight families and 49 subfamilies. These subfamilies exhibit great functional diversity including non-catalytic members with different subcellular, phylogenetic, and species distribution. This results from constant enzymogenesis and proteinogenesis within each kingdom, and highlights the huge plasticity that MDR superfamily members possess. Thus, through evolution a great number of taxa-specific new functions were acquired by MDRs. The generation of new functions fulfilled by proteins, can be considered as the essence of protein evolution. The mechanisms of protein evolution inside MDR are not constrained to conserve substrate specificity and/or chemistry of catalysis. In consequence, MDR functional diversity is more complex than sequence diversity.

MDR is a very ancient protein superfamily that existed in the last universal common ancestor. It had at least two (probably three) different ancestral activities related to formaldehyde metabolism and alcoholic fermentation. Eukaryotic members of this superfamily are more related to bacterial than to archaeal members; horizontal gene transfer among the domains of life appears to be a rare event in modern organisms inside this protein superfamily.

The huge plasticity observed inside MDR members, does not seem to be an exclusive property of this superfamily; eighty percent of all sequenced proteins belong to any of the great protein superfamilies reported. A comprehensive analysis of any other of the main protein superfamilies probably will shows similar results. Indeed, the taxonomic system proposed by MDR

proteins (superfamily, macrofamily, family and subfamily) can be easily extrapolated to other protein superfamilies.

Presentación[1]

La idea “una secuencia de aminoácidos determina un plegamiento o estructura única y ésta a su vez una sola función”, es uno de los grandes paradigmas que permitió un notable avance en el desarrollo de la Bioquímica moderna. Los contribuciones pioneras de Edward Tatum y Christian Anfinsen, en las décadas de 1940 y 1950 fueron determinantes para ello. Sin embargo, gracias a la gran cantidad de información que se ha acumulado sobre la estructura y función de las proteínas, este paradigma ha comenzado a ser retado y nos obliga a replantearlo con base en una nueva e insospechada plasticidad funcional y estructural de las proteínas.

Hasta hace poco tiempo se consideró que las proteínas que pertenecen a una misma superfamilia de proteínas comparten la misma arquitectura y el mismo mecanismo de reacción (Gerlt y Babbit, 2000). Esta idea está también ampliamente arraigada entre los desarrolladores de las bases de datos de proteínas, para quienes es importante indicar una probable función para las proteínas en las cuales se conoce únicamente su secuencia. Sin embargo, en un estudio reciente, publicado por nuestro grupo de trabajo (Riveros-Rosas y cols., 2003), se muestran datos en donde se ilustra la gran diversidad de funciones desarrolladas por los miembros de una superfamilia de proteínas, las alcohol deshidrogenasas dependientes de zinc, actualmente denominadas como deshidrogenasas/reductasas de cadena media (MDR). Esta superfamilia incluye algunos de los grupos de enzimas más estudiadas, tales como las típicas alcohol deshidrogenasas dependientes de zinc características de animales y levaduras, las formaldehído deshidrogenasas, las sorbitol deshidrogenasas, las treonina deshidrogenasas o las quinona oxidorreductasas.

Así, en dicho trabajo, mediante un análisis *in silico* muy detallado de la estructura y función de 583 proteínas pertenecientes a la superfamilia de las MDR, se obtuvieron las siguientes conclusiones:

1) Su actividad es sorprendentemente diversa, ya que si bien la mayoría son deshidrogenasas y/o reductasas, también están presentes dismutasas, esterasas, metilformato sintasas, liasas, y proteínas no catalíticas como proteínas que se unen

[1] Esta presentación corresponde básicamente al documento enviado por invitación a la Sociedad Iberoamericana de Información Científica, con sede en Buenos Aires, Argentina: Riveros-Rosas H, Julián-Sánchez A, Piña E. (2004). Proteínas: rompiendo paradigmas. Publicado en formato electrónico en: Siic”Salud”, sección: *Los Autores comunican.* (ISSN: 1667-9008). http://www.siicsalud.com/dato/dat037/04322001.htm

a receptores nucleares, proteínas unidas al ADN, proteínas involucradas con el tráfico de proteínas de membrana y proteínas estructurales como la ζ-cristalina.

2) Su distribución subcelular es también muy diversa, ya que además del citoplasma, pueden encontrarse en mitocondria, núcleo e inclusive en membranas.

3) Su contenido de zinc es variable, algunas requieren 2 Zn^{2+}/subunidad, otras 1 Zn^{2+}/subunidad, y otras carecen de él.

4) Con respecto al uso de cofactores, la mayoría utilizan NADH, NADPH, o ambos, aunque algunas no requieren de ningún cofactor.

5) En cuanto a su organización, la mayoría forma estructuras monofuncionales, ya sea en forma de dímeros o tetrámeros, pero existen también trímeros y monómeros funcionales. Algunas están unidas a otras proteínas y/o dominios, formando proteínas multifuncionales.

Toda esta información resalta la gran plasticidad y diversidad de funciones que desarrollan las proteínas que pertenecen a las MDR y que es resultado de una intensa enzimogénesis y proteinogénesis (Riveros-Rosas y cols., 2003). Así, a lo largo de la evolución, un gran número de nuevas funciones fueron adquiridas por las MDR dentro de diferentes taxa o grupos de organismos. De hecho, la capacidad para desarrollar nuevas funciones puede ser considerada como la esencia misma de la evolución de las proteínas. De esta manera, los mecanismos de evolución de las proteínas no están restringidos a conservar la especificidad del sustrato o el mecanismo de reacción (Gerlt y Babbit, 2001), y demuestran, al menos para el caso de las MDR, que una misma arquitectura del sitio activo de una proteína puede participar en dos o más reacciones completamente diferentes. De hecho, la presencia de proteínas multifuncionales parece ser ahora la regla y no la excepción (James y Tawfik, 2003). En este contexto, las actividades secundarias desarrolladas por una proteína pueden servir como el punto de partida para desarrollar nuevas funciones (O'Brien y Herschlag, 1990; James y Tawfik, 2001), y varios ejemplos se muestran en los trabajos antes mencionados.

Por otra parte, esta enorme plasticidad dentro de las MDR no parece ser exclusiva de esta superfamilia, ya que si se toma en cuenta que cerca del 80% de todas las proteínas conocidas hasta ahora, pertenecen a alguna de las grandes superfamilias de proteínas ya identificadas (Wolf y cols., 2000), es claro que al explorar en detalle alguna de estas otras superfamilias se obtendrán, muy probablemente, conclusiones similares. De hecho, el sistema taxonómico propuesto por nuestro grupo de trabajo (Riveros-Rosas y cols., 2003) para las MDR (superfamilia, macrofamilia, familia y subfamilia), puede ser fácilmente extrapolado a otras superfamilias.

En resumen, tenemos que adaptarnos a la idea de que una misma proteína puede desarrollar varias funciones de manera simultánea. Así, la diversidad funcional dentro de una superfamilia de proteínas, puede ser tan grande como su diversidad en secuencias.

Por esta razón, nos parece indispensable analizar con detalle, las modificaciones que han experimentado en su secuencia algunas de las enzimas pertenecientes a esta superfamilia de proteínas, con la idea de obtener información útil que nos permita comenzar a entender como actúan las fuerzas que dirigen la evolución de proteínas.

Aquí, es importante aclarar que la actividad de alcohol deshidrogenasa (ADH) dependiente de NAD(P) se encuentra ampliamente distribuida en la naturaleza, y es llevada a cabo por tres superfamilias de enzimas, las cuales se originaron en forma independiente durante el transcurso de la evolución (Reid y Fewson, 1994). Ellas comparten entre sí un 20% de identidad o menos en su secuencia y tienen estructuras y mecanismos de reacción distintos. La primer superfamilia corresponde a las ADHs dependientes de hierro, que es el grupo más pequeño y menos estudiado de deshidrogenasas: se encuentran característicamente en bacterias, levaduras y protozoarios (Scopes, 1983; Williamson y Paquin, 1987; Conway e Ingram, 1989). La segunda superfamilia corresponde a las alcohol deshidrogenasas de cadena corta, ahora designadas como deshidrogenasas/reductasas de cadena corta; esta superfamilia es muy numerosa y se caracteriza por no requerir de ningún cofactor metálico para desarrollar su actividad (Persson et al., 1991; Krozowski, 1994). La tercer superfamilia está compuesta por las ADHs dependientes de Zn, ahora designadas preferentemente como deshidrogenasas/reductasas de cadena media (MDR) (Persson y cols., 1994; Jornvall y cols., 1999; Nordling y cols., 2002). Esta última superfamilia, como ya se mencionó, corresponde a nuestro grupo de estudio.

De esta forma, el análisis *in silico* de la superfamilia de las MDR nos permite obtener información que contribuye a entender algunos de los mecanismos que dirigen la evolución de las proteínas y permiten que un polipéptido, con una estructura y función dadas, adquiera, a través de la sustitución de algunos aminoácidos, una actividad catalítica modificada, pero conservando esencialmente la misma arquitectura.

A este respecto, es importante señalar que existen varios grupos de investigación, que mediante diferentes protocolos de mutagénesis, han tratado de simular *in vitro* la evolución de las proteínas (Hedberg y cols., 1998; Altamirano y cols., 2000; Jurgens y cols., 2000; Tao y Cornish, 2002), tratando de cambiar la actividad catalítica de un polipéptido dado, es decir, lograr que mediante algunos cambios en la secuencia de aminoácidos, la proteína cambie su actividad original por otra. Este tipo de estrategias tienen también como finalidad aportar información que nos ayude a entender la evolución de las proteínas. Sin embargo, soslayan el hecho de que la naturaleza ya efectuó previamente múltiples experimentos de este tipo, y por selección natural ha conservado especialmente los experimentos que resultaron exitosos. El hecho de que dentro de una misma superfamilia de proteínas, las MDR, se haya demostrado la existencia de cerca de 50

subfamilias diferentes, cada una de ellas con una actividad distinta (Riveros-Rosas y cols., 2003) es, a nuestra manera de ver, la prueba más palpable de que la naturaleza continuamente está experimentando con la estructura y la función de las proteínas y que los mecanismos que dieron origen a las proteínas siguen operando y generando hoy nuevas proteínas. Por esta razón, consideramos que el análisis de los experimentos de migración catalítica, desarrollados por la misma naturaleza, poseen una riqueza informativa mayor que los experimentos *in vitro*, cuyo éxito frecuentemente es incierto.

Introducción

A continuación, se presenta la primera publicación de este proyecto de tesis doctoral, en donde se elaboró una revisión sobre los sistemas oxidantes de etanol, poniendo especial énfasis en el sistema de las alcohol deshidrogenasas de cadena media, que conforman la subfamilia de proteínas más representativa de lo que se denominaría posteriormente como superfamilia de deshidrogenasas/reductasas de cadena media. En dicho trabajo se discute también el papel fisiológico desarrollado por este conjunto de enzimas en mamíferos y se propone un mecanismo novedoso de toxicidad del etanol, con base en su capacidad como inhibidor competitivo del catabolismo de diversos metabolitos endógenos responsables de regular el metabolismo, la diferenciación celular, el desarrollo embrionario y las funciones neuroendócrinas. Dicha publicación ha sido citada, hasta mayo del 2004, en 21 ocasiones.[1]

Publicación 1.

Héctor Riveros-Rosas, Adriana Julián-Sánchez y Enrique Piña. (1997). Enzymology of ethanol and acetaldehyde metabolism in mammals. Arch. Med. Res. 28 (4): 453-471.

[1] Este número de citas es sobresaliente considerando que el número promedio de citas de los artículos publicados en el Archives of Medical Research en 1997, es 2.74 ± 3.44 (n=101), y que el factor de impacto de esta revista es de 0.618; de hecho, corresponde a uno de los 3 artículos más citados de esta revista, durante 1997.

Archives of Medical Research

Volume 28, No. 4, pp. 453-471, 1997
Printed in Mexico

Review Article

Enzymology of Ethanol and Acetaldehyde Metabolism in Mammals[1]

HECTOR RIVEROS-ROSAS, ADRIANA JULIAN-SANCHEZ and ENRIQUE PIÑA

Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, México, D.F.

Received for publication February 24, 1997; accepted October 1, 1997 (97/026).

Abstract

Ethanol or ethyl alcohol is a molecule that, in mammals, is naturally present at low concentrations due to its production by gastrointestinal flora fermentation activity. However, it is remarkable that this metabolite, with a clearly minor role in regular vertebrate metabolism, can be oxidized into acetaldehyde through several enzymatic and non-enzymatic mechanisms, which comprise the activity of more than ten enzymes and isozymes, many of them broadly distributed in different species and tissues. In correspondence, acetaldehyde can also be oxidized into acetate through several enzymatic pathways that involve about ten enzymes and isozymes which also have a broad distribution. In this article, a complete review of the aforementioned metabolic pathways is elaborated. From this group, the participation and wide distribution of alcohol dehydrogenase and aldehyde dehydrogenase systems are emphasized. The mechanisms of reaction, kinetic characteristics and physiological relevance are described, and finally, the possible physiological role of these enzymatic systems as responsible to synthesize or catabolize several endogenous metabolites that regulate growth, metabolism, differentiation and neuroendocrine function in mammals are discussed. (*Arch Med Res* 1997; 28:453)

KEY WORDS: Ethanol; Alcohol dehydrogenase; Ethanol metabolism; Acetaldehyde; Aldehyde dehydrogenase; Acetaldehyde metabolism; Cytochrome P450; Physiological role.

Introduction

The manufacture and consumption of alcoholic beverages are ancient practices common to nearly all cultures throughout the world, and their origins are associated in most cases with the occurrence of agriculture (1,2). In this way, the discovery by Batelli and Stern in 1910, that extracts from various animal tissues are able to oxidize ethanol to acetaldehyde and acetic acid, and the subsequent purification of an alcohol and aldehyde dehydrogenase from animal tissues (3-5), was considered to be a normal property inherent to animal metabolism.

Ethyl alcohol is actually consumed in great quantities through the ingestion of diverse types of alcoholic beverages, so that in some countries, the consumption per capita of alcoholic beverages, such as beer, can exceed 120 liters a year, e.g., Germany, the Czech Republic, New Zealand and Belgium (6). However, it must be recognized that except for humans, higher animals neither consume nor produce enough quantities of alcohol to justify the presence of more than twenty enzymes and isozymes with the only apparent finality being that of oxidizing ethanol and acetaldehyde.

In recent years, an impressive load of information concerning ethanol and acetaldehyde metabolism, as well as the metabolic disorders induced by excessive intake of alcohol (e.g., 7,9) has been generated. However, the questions, far from being curtailed, are growing

Correspondence to:
Héctor Riveros-Rosas. Departamento de Bioquímica, Facultad de Medicina, UNAM. Apdo. Postal 70159, Ciudad Universitaria, 04510 México, D.F. E-mail: hriveros@servidor.unam.mx

[1]This work was partially supported by grant No. IN212695 of the Dirección General de Asuntos del Personal Académico, UNAM, México.

in the same proportion, because the number of papers showing novel metabolic processes modified by ethanol increases from year to year. In this manner, a growing number of metabolic aspects must be considered to explain basic questions, such as the toxicity mechanisms of ethanol, or the possible physiological roles for alcohol and aldehyde dehydrogenases.

At present, excellent reviews covering different aspects of alcohol-induced diseases have been written (e.g., 7-10). However, none of them analyzes, in great detail, the entire collection of enzymes employed in ethanol and acetaldehyde metabolism. Several individual studies which take into account particular aspects of each of the metabolic pathways responsible for ethanol oxidation have been elaborated. However, no review gives an overall view, which is the reason that one of goals of this review is to fill this void. This larger vision clearly allows the authors of this review to visualize ethanol and acetaldehyde metabolism from another perspective, where the existence and evolution of this complex collection of enzymes respond to physiological needs other than solely the oxidation of ethanol and acetaldehyde.

Properties of Ethanol

Ethanol, ethyl alcohol or simply alcohol is a colorless volatile liquid of a characteristic odor which is produced principally by means of sugar or carbohydrate fermentation through the action of microorganisms such as, for example, bacteria and yeasts. It is a liquid of polar nature, soluble in water, chloroform and acetone, among other solvents, and is practically insoluble in fats and oils. It is important to emphasize the latter, because ethanol in some publications has been erroneously described as a substance soluble in lipids (e.g., 11-13). This incorrect conception possibly has its origin in the fact that ethanol, because of its small size, is able to pass freely through biological membranes in the same manner as does any liposoluble substance; however, it must be remembered that other low molecular weight polar molecules, such as water, can also easily pass through the double layers of phospholipids without needing to be soluble in lipids.

In mammals, ethanol is a metabolite which is normally present in cells at low concentration, on the order of 0.1 to 1 mM, its presence due to gastrointestinal flora fermentation activity and to endogenous production in different tissues, especially liver (14-16). However, the main source of ethanol in man is not its endogenous production but its direct consumption. Only in rare cases among patients with a blind intestinal loop, and in whom there is also an overproduction of the intestinal flora with *Candida* and other yeasts, can there be observed elevated levels of ethanol in blood, due to the fermentation of carbohydrates in intestine, which may even produce manifest drunkenness.

Ethanol Absorption

Once ethanol is ingested, it begins to be absorbed immediately through the mucous epithelium of the mouth and as vapor in the pulmonary alveoli (17); however, the amount of ethanol which is absorbed through these two pathways is practically negligible, for which reason it can be considered that all the ingested alcohol goes to the stomach, where it is absorbed by simple diffusion through the epithelium of the stomach and intestines in a manner entirely similar to that of water (17,18). Thus, once ethanol arrives at the stomach, it is absorbed in small proportions in the bloodstream through the gastric mucosae (approximately 20 - 30%), while the largest amount is channeled through the same gastric emptying toward the small intestine (17,19,20). It is in the small intestine (principally duodenum and jejunum) that the major part of the absorption of ethanol is performed. Ethyl alcohol is rapidly channeled toward the portal vein and from there, carried directly to liver to be metabolized. Figure 1 exhibits the principal ethanol absorption routes.

Once the ethanol is in blood, it is distributed throughout all of the bodily fluids, crossing freely through the totality of the biological membranes (18); in this way

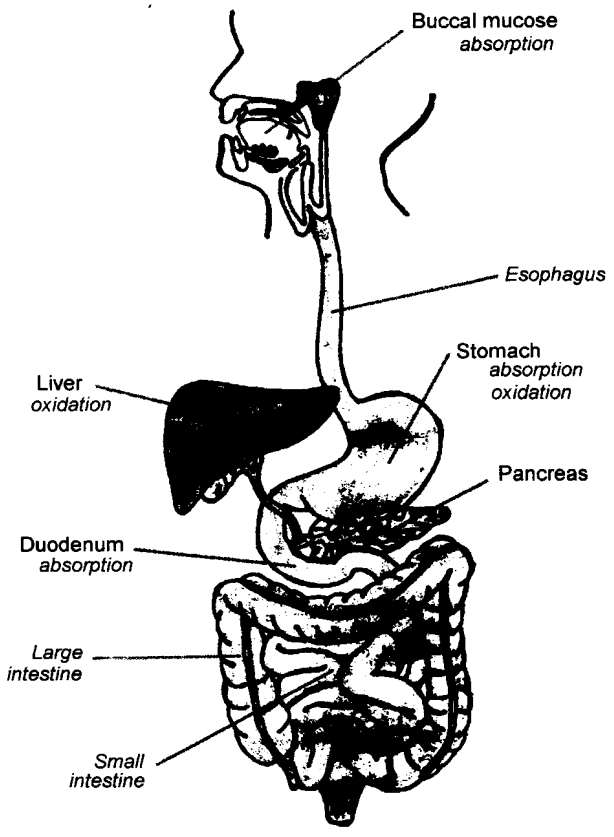


Figure 1. Main organs involved in ethanol absorption and subsequent oxidation when alcohol is administered by orogastric via.

Table 1
Factors Affecting Ethanol Absorption

Stomach emptying
Food in the stomach
Ethanol concentration
Blood flow at the site of absorption
Irritant properties of ethanol
Rate of ingestion
Type of beverage
Protein deficiency
Body temperature
Physical exercise
Menstrual cycle

Modified from Agarwal & Goedde (13).

women, who have a smaller proportion of bodily fluids than men (53% vs. 61.8%), reach higher levels of ethanol in blood even if both men and women ingest equivalent amounts of alcohol (17,18). In fact, ethanol, as has already been pointed out, is practically not distributed in body fat (17).

The absorption rate of ethanol depends mainly on the velocity with which it passes from the stomach to the small intestine, which in turn depends principally on the speed at which gastric emptying is carried out. In general, the presence of solid food retards gastric emptying and therefore, the speed of ethanol absorption (21). On the other hand, the concentration of the ingested ethanol also influences its absorption, so that high concentrations of ethanol retard the emptying of food (21,22); there are, in fact, many other factors which influence to a greater or lesser extent the velocity of ethanol absorption. Table 1 shows some of these factors.

Ethanol Metabolism

Ethanol is absorbed mainly in intestine, where it is channeled through the porta vein directly toward liver before passing through the circulatory system and the rest of the body. The liver is the principal organ responsible for the oxidation and elimination of ingested alcohol, as well as the majority of pharmaceuticals and xenobiotics which may be present in blood.

In liver, there exist three metabolic systems capable of carrying out ethanol oxidation. The first system is made up of a series of specialized enzymes known generically as alcohol dehydrogenases or *ADHs*, which are found in the cytosol of the different tissues which make up the human body, principally hepatic. These enzymes promote the oxidation of ethanol into acetaldehyde, coupling this oxidation with the reduction of a nicotinamide adenine dinucleotide (*NAD^+*). The second system is found in hepatic cell peroxisomes, and in this system, the oxidation of a molecule of ethanol into acetaldehyde is coupled with the simultaneous decomposition of a hydrogen peroxide molecule in a reaction catalyzed by the catalase enzyme. The third and last oxidizing system is called *Microsomal Ethanol Oxidizing System (MEOS),* which is located in the interior of microsomes and requires cytochrome P-450 participation. The cytochromes P-450 couple ethanol and nicotinamide adenine dinucleotide phosphate oxidation (*NADPH*) to the reduction of an oxygen molecule to form hydrogen peroxide. These three systems work simultaneously in the presence of ethanol, although with different activities and affinities. Figure 2 shows the type of transformation catalyzed in each of the ethanol oxidizing systems.

Besides the three enzymatic systems described above for ethanol oxidation, there exists a non-enzymatic oxidation mechanism which would probably be functional *in vivo* and which depends on iron chelate participation in the presence of hydroxyl radicals (23).

Finally, there exists a non-oxidative metabolic pathway (23, 24) where ethanol forms fatty acid ethyl esters by means of a fatty acyl ester synthetase. If it is probable that its contribution to total ethanol metabolism is not significant (23), its participation in the development of pathological alterations in organs that lack ethanol oxidating systems with high activity, such as pancreas, heart and brain, remains to be discussed (25).

A. Alcohol Dehydrogenase System

Alcohol dehydrogenase activity is amply distributed in numerous phyla which include organisms belonging to each of the different kingdoms in which living beings are classified (26). However, this activity is not carried out by only one type of enzyme, but by three families of enzymes, which seem to have arisen independently throughout evolution (26). In this way, at least three protein families of different origins, structures and reaction mechanisms which exhibit this capacity have been described.

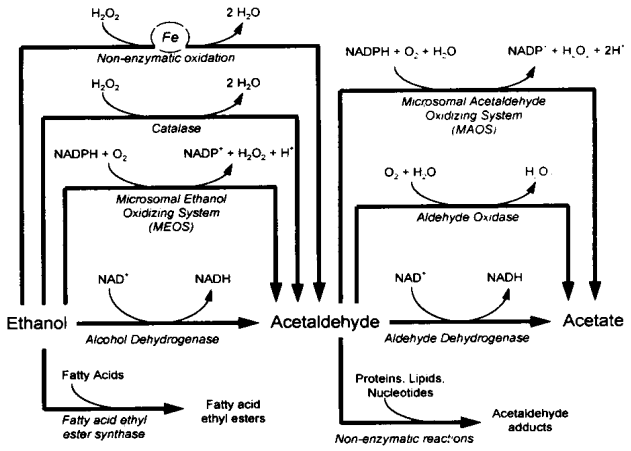


Figure 2. Principal metabolic pathways related to ethanol and acetaldehyde metabolism.

The first family is comprised of "short chain" alcohol dehydrogenases, which do not require a metallic ion as cofactor and which are found characteristically among insects (27,28). This family also includes other related enzymes, such as procariont glucose and ribitol dehydrogenase (29), and some mammal hydroxysteroids- and hydroxyprostaglandins-dehydrogenases (28,30).

The second group is made up of the "medium and long chain" alcohol dehydrogenases, and includes the "classic" mammal alcohol dehydrogenase. These enzymes require zinc atoms as cofactor, and are divided into tetramerics and dimerics (26, 27); the latter conform the most diversified group in mammals, and include about a dozen different isozymes classified into six distinct classes (9,31), one of which, Class III (also denominated glutathione-dependent formaldehyde dehydrogenase), presents an apparently universal distribution (32,33). Included also in this family are other enzymes, such as sorbitol-, xylitol- and treonine-dehydrogenases, among others (26).

Table 2

Alcohol Dehydrogenase Activity on Several Rat Organs with 33 mM Ethanol and 4 mM NAD^+ as Substrates, pH 7.5

	Activity	
Organ	mU/g tissue	Total (mU/organ)
Eyes	5.87 ± 0.7	1.5 ± 0.2
Ear mucosa	8.9 ± 1.3	2.0 ± 0.3
Nasal mucosa	15.3 ± 2.6	1.3 ± 0.2
Trachea	5.3 ± 0.7	1.2 ± 0.1
Lungs	8.1 ± 1.4	13.5 ± 2
Buccal mucosa	3.8 ± 0.8	0.3 ± 0.1
Tongue	5.4 ± 0.5	4.0 ± 0.3
Esophagus	16.1 ± 3.2	3.7 ± 0.7
Stomach	11.8 ± 1.7	19.6 ± 3
Small intestine	19.3 ± 4.5	58.0 ± 14
Colon and cecum	13.5 ± 1.3	30.0 ± 3
Rectum	37.3 ± 7.0	28.5 ± 6
Liver	260.0 ± 50	3500.0 ± 640
Adrenal glands	4.0 ± 0.5	0.8 ± 0.3
Kidneys	16.0 ± 7.0	26.0 ± 13
Urinary bladder	10.0 ± 2	1.8 ± 0.5
Testes	26.2 ± 9.0	60.0 ± 20
Epididymis	10.5 ± 4.1	40.0 ± 16
Penis	17.5 ± 8.09	4.44 ± 1.16
Ovaries	7.0 ± 5.0	0.9 ± 0.7
Uterus	17.4 ± 7.2	9.6 ± 4
Vagina	12.5 ± 4.6	3.0 ± 1
Skin	2.8 ± 0.3	88.0 ± 15

Data reported by Boleda et al. (35). Values are expressed as mean ± standard error.

The third and last family corresponds to the Fe-dependent alcohol dehydrogenases, which are characteristic of unicellular organisms such as *Zymomonas* and *Clostridium* (the latter, *$NAPD^+$* dependent). This group makes up the smallest and least studied family of alcohol dehydrogenases (34).

These families of alcohol dehydrogenases present among themselves a residue identity of around 20% or less, a value which is comparable to other distinct dehydrogenases; therefore, it may be considered that these three families form different enzymatic systems (26).

In vertebrates, just the two first protein families are present, but only the "medium and long chain" family exhibits a significant alcohol dehydrogenase activity in these organisms. Therefore, in this review, the "medium and long chain" family will be referred to as the alcohol dehydrogenase system.

This last system in mammals is divided into six distinct classes which make up, at the least, twenty different isozymes codified by more than seven distinct genes (9,31). These enzymes, as noted in Table 2, are amply distributed in different human body tissues (35), and all of them are characterized by being dimeric proteins with subunits of approximately 40 kDa; each subunit is characterized as containing two atoms of Zn bonded by cysteines that help stabilize the enzyme structure (36), besides the fact that one of them forms part of the active site (9,37,38). Table 3 shows the different classes of *ADH* described above as well as the distinct genes and subunits that they comprise.

All these enzymes use *NAD^+* as electron-accepting coenzyme, to catalyze the oxidation of ethanol into

Table 3

Alcohol Dehydrogenase (ADH) Polymorphisms in Mammals

Class	Gene Locus	Allele[a]	Subunit	Tissue distribution
I	ADH - 1	*ADH1*	α_1	Liver
I	ADH - 2	*ADH2*1*	β_1	Liver
I		*ADH2*2*	β_2	and
I		*ADH2*3*	β_3	lung
I	ADH - 3	*ADH3*1*	γ_1	Liver and
I	ADH - 3	*ADH3*2*	γ_2	stomach
II	ADH - 4	*ADH4*	π	Liver, cornea, kidney and lung
III	ADH - 5	*ADH5*	χ	All tissues
IV	ADH - 7	*ADHσ*	σ	Liver, stomach, skin, cornea
IV	?	*ADHμ*	μ	Liver, stomach, skin
V	ADH - 6	*ADH6*	?	Liver, stomach
VI	?	?	?	Liver

They are the same isozyme.

[a]Each gene locus has one allele, except for ADH-2, which has three alleles, and ADH-3, which has two. Alleles from the same gene locus are indicated with an asterisk.

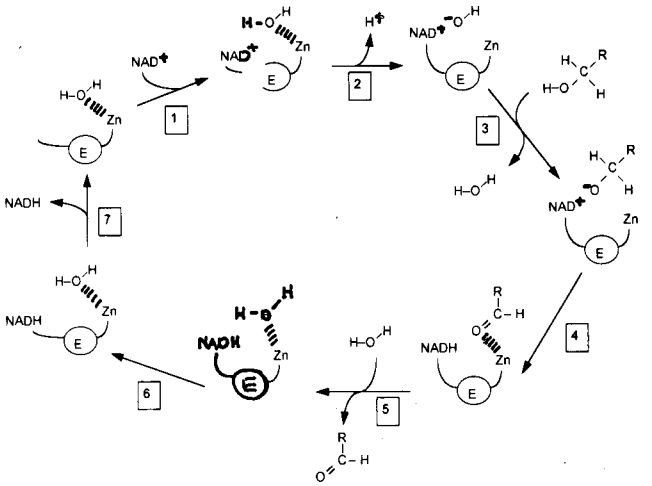


Figure 3. Alcohol dehydrogenase mechanism reaction. ADH possesses a union site for NAD^+ or NADH, and an active site made up of a hydrophobic cavity where one of the Zn molecules, which binds a molecule of water, is found. The reaction starts with the formation of an abortive E-NAD^+ complex (1), where because of a rotation, the active site approximates to the coenzyme union site, thus promoting the liberation of an H^+ (2); later, the molecule of alcohol enters the active site, displacing the molecule of water, occupying its place, now forming a ternary E-NAD^+-alcoholate complex (3); the alcoholate oxidizes, transferring a hydride ion to the coenzyme in order to form NADH and acetaldehyde (4); the latter leaves its active site and is exchanged for a molecule of water (5); the next stage is the limiting step of the reaction, in which the enzyme should rotate again to recuperate its original conformation (6), thus allowing the liberation of NADH (7). Schematic diagram elaborated according to information provided by Brändén et al. (39).

acetaldehyde (EC 1.1.1.1.); its reaction mechanism has been described as bi-bi ordered type (e.g., 39, 40), where the *NAD^+* or *NADH* should penetrate the coenzyme union site before the ethanol or acetaldehyde enters the catalytic domain (Figure 3). All *ADH* isozymes show a Michaelis-Menten kinetics with respect to ethanol (41), and only γγ isozymes seem to exhibit a negative cooperativity for ethanol. The dissociation of *NADH* is the rate-limiting step of these enzymes (41,42).

By means of X-ray crystallography studies, the tridimensional structure of horse Class I *ADH* (37,38,43) and human $\beta_1\beta_1$ *ADH* (44,45) are well known, and both are very similar to each other. Additionally, cod Class I and III *ADH* have been crystallized, and soon their tridimensional structure will be known in detail (46). Figure 4 shows the tridimensional structure of horse Class I alcohol dehydrogenase.

Due to the fact that *ADHs* have their union site for the substrate at the end of a 20 Å hydrophobic pocket (Figure 4), they catalyze not only the oxidation of small molecules such as ethanol or methanol, but also a large variety of hydroxilated hydrophobic molecules, such as long chain alcohols, whether primary, secondary, ramified or cyclical (47), takes part in the dehydrogenation of steroids (48) and in the ω-oxidation of fatty acids (49), and can even catalyze aldehyde oxidation to acids by dismutation (40,50,51).

"Medium and long chain" alcohol dehydrogenases are grouped into classes based on their structural differences; in this way, each *ADH* in distinct classes present among themselves a similarity of approximately 60%, while those that pertain to only one class have a residue identity of no less than 85% (26,52,53). Figure 5 shows the possible phylogenetic relationships that exist between different *ADH* types. Of all the classes heretofore described, the first five have already been found in human tissue (53,55); only Class VI, recently described, has not yet been characterized. Next, the most relevant

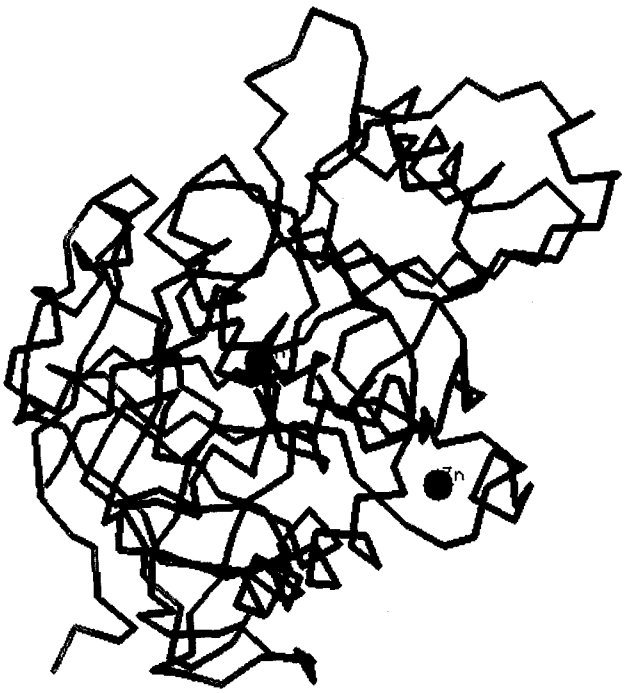


Figure 4. Schematic diagram of the tridimensional structure of Class I horse alcohol dehydrogenase showing the $C_{(\alpha)}$ atom backbone. The enzyme is a dimer with a molecular weight of 80 kDa (only one subunit is illustrated). Each subunit possesses a coenzyme-binding domain and a catalytic domain. The coenzyme-binding domains have a structure similar to that found in other *NAD^+* dependent dehydrogenases, and they form a 12-stranded, β-pleated sheet structure that makes up the central core (37, 42). The active site is situated in a cleft between the coenzyme-binding core and the catalytic domain; the substrate binds in a hydrophobic pocket in the cleft, with its oxygen atoms ligated to the zinc atoms of the catalytic domains (42). When the coenzymes bind, the active site clefts close up by a rigid body rotation of the catalytic domains toward the coenzyme-binding domain (38,42). The Figure was elaborated with the Raswin Ver 2.5 program, and the atom coordinates were obtained from The Genome Database (http://gdbwww.gdb.org/gdb/) on internet page http://pdb.pdb.bnl.gov/cgi-bin/send-short-text?filename=*1DEH* (May 15, 1997).

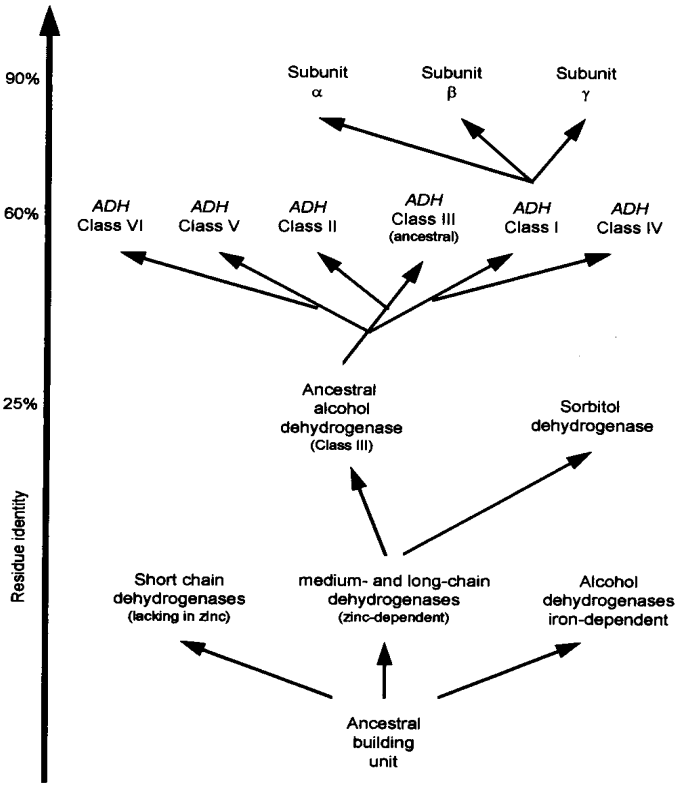


Figure 5. Phylogenetic relations among the different types of alcohol dehydrogenases. The family of "short" chain *ADHs* seem to be present mainly in insects, while the family of "medium and long chain" *ADHs* is much more amply distributed, being found in prokaryotes and eukaryotes, among which mention may be made of mushrooms, plants, cephalopods and vertebrates. The only medium and long chain *ADH* reported in insects corresponds to Class III, and is equivalent to the octanol dehydrogenase of insects previously reported (30). This class of enzymes is highly preserved in the different species, making it probable that it corresponds to an ancestral form which could have originated the different types of *ADHs*. In fact, it is estimated that the divergence among the principal classes of *ADH* must have occurred approximately 600 million years ago (54), a date which also corresponds to the time during which vertebrates arose, giving rise to speculations concerning the existence of the different classes of *ADHs* in all vertebrate groups. In a similar way, the divergence among the different isozymes of Class I *ADH* occurred approximately 80 million years ago; therefore they were presumably present in the different groups of primates (54). Diagram elaborated on basis to information provided by Jörnvall et al. (52), Parés et al. (53) and Danielsson et al. (31).

characteristics of each Class of mammal alcohol dehydrogenase are briefly described.

Class I: This is the most studied class in this enzyme family and comprises the "classic" hepatic enzyme. It includes three distinct gene loci (*ADH-1, ADH-2* and *ADH-3*) found in human chromosome four (27), which encode for three types of subunits (α, β and γ respectively, Table 3). The alleles for the β and γ subunits differ among themselves because of a few aminoacid residues (β_1, β_2, β_3 and γ_1, γ_2), so that all the subunits which make up Class I exhibit among themselves a residue identity of at least 85 - 90%. These subunits can form homodimers or heterodimers indistinctly, contributing then, to the generation of a great polymorphism with many isoforms. Each isoform presents variants as to their kinetic properties (Table 4), which partially explains the great heterogeneity in the capacity for metabolizing ethanol that human populations exhibit. In fact, certain alleles are characteristic of each population; thus, for example, *ADH2*1* is common in Caucasian and Afro-American populations, while *ADH2*2* and *ADH3*1* are predominant in Chinese and Japanese populations (41).

All isozymes of this class are highly sensitive to pyrazole inhibition (31) and show substrate inhibition ([ethanol] >20 mM), because an ethanol excess decreases the speed of *NADH* dissociation of the enzyme (56). They also participate in steroid dehydrogenation (48) and in the ω-oxidation of fatty acids (49).

Class II: This class is made up of homodimeric isozyme ππ, and is found in diverse organs such as liver, cornea, kidney and lung. It participates actively in the degradation of circulating epinephrine and norepinephrine (57), possesses a high affinity for ethanol, and is only slightly sensitive to pyrazole inhibition (58).

Class III: This class is made up of isozyme χχ, which is a homodimer encoded by the ADH5 gene, and probably corresponds to the ancestral form which gave origin to the rest of the *ADHs* (9,31). It possesses a great capacity to metabolize long chain alcohols, while its capacity to oxidize ethanol is, on the other hand, very limited (59). Furthermore, this class is even less sensitive to pyrazole inhibition than is Class II (27). Additionally, due to its broad distribution in the different species

Table 4
Kinetic Properties of Some *ADH* Isozymes, with Ethanol as Substrate

Isoenzyme	Kinetic constant			pH optimum
	NAD⁺ Km (μM)	Ethanol Km (mM)	Vmax (min^{-1})	
αα	13.0	4.2	27.0	10.5
$\beta_1\beta_1$	7.4	0.049	9.2	10.5
$\beta_2\beta_2$	180.0	0.94	400.0	8.5
$\beta_3\beta_3$	712.0	36.0	300.0	7.0
$\gamma_1\gamma_1$	7.9	1.0	87.0	10.5
$\gamma_2\gamma_2$	8.7	0.63	35.0	10.5
ππ	14.0	34.0	20.0	10.5
χχ		≥ 2000	?	

Modified from Burnell & Bosron (41).

and tissues, and also to its weak capacity to oxidize ethanol, it has been suggested that this enzyme carries out special functions different from the other classes of *ADHs* (60). In fact, this enzyme is identical to glutathione-dependent formaldehyde dehydrogenase (EC1.2.1.1) (32,33).

Class IV: This class is expressed preferably in the stomach (31), although it can be found in lesser quantities in other tissues such as liver, skin, and cornea (9). It is made up of isozymes σσ and μμ, which are encoded by the *ADH*σ and *ADH*μ genes, located in the *ADH*-7 locus (9). It presents a very high Km for ethanol, and is thought to participate actively in the first-pass ethanol metabolism in stomach. They are inhibited by the action of cimetidine, ranitidine, nizatidine and aspirin (61-64).

Class V: This class has been found in gastric epithelium, and is made up of a homodimer encoded by the *ADH6* gene (9). It possesses a very high Km for ethanol and is also thought to participate in the gastric first-pass metabolism of ethanol.

Class VI: This is the newest of the *ADH* classes described in mammals (65). It was identified for the first time in *Peromyscus maniculatus* (deer mouse), and is present in *ADH*-positive genotypes as well as in negative *ADH* (which lack the Class I "typical" alcohol dehydrogenase). It is expressed principally in liver and in minimum quantities in kidney, and its role in ethanol metabolism is still uncertain. No reports of its existence have been noted in humans, although it has also been identified in rats (66). It presents its main similarity (67%) with Class V *ADH6*.

B. Microsomal Ethanol Oxidizing System (MEOS).

This system was studied for the first time by Lieber et al. at the end of the 60s (67,68) from an initial report by Orme-Johnson and Ziegler (69), in which they described parallel increments in the capacity to metabolize drugs and ethanol with total activity of cytochrome P-450 in the smooth endoplasmic reticulum of hepatic cells. In this way from these initial reports, the existence of a new pathway for ethanol oxidation, dependent on cytochrome P-450, was rapidly described.

Currently, the *MEOS* presence in the microsomes of different tissues has been described (70), and it is now known that it involves the participation of various enzymes of the present cytochrome P-450 superfamily. This P-450 superfamily is constituted of around 230 genes and pseudogenes distributed among thirty-six distinct families, of which at least ten are present in mammals (70-72); these families are defined basically as to their similarities in terms of aminoacid sequences, so that all members of a family present at least 55% similarity in their sequences, and around 40% similarity with the members of other families (70).

Cytochromes P-450 are divided into two main classes, from the functional point of view: the first class (which is comprised by mammalian classes CYP1, CYP2, CYP3 and CYP4) is involved in the metabolism of xenobiotic compounds, while the second (the CYP7, CYP11, CYP17, CYP19, CYP21 and CYP27 families) is related to the synthesis of steroids and biliary acids (70).

Different types of cytochrome P-450 are capable of catalyzing the ethanol oxidation, although the most efficient is cytochrome P-450 2E1 (23, 73), also termed CYP2E1, or simply 2E1, according to the nomenclature recommended by Nebert et al. (74-76). This cytochrome was originally referred to as "3a" in rabbits and rats, and as "j" in humans, and these names are kept only to be able to refer to papers published previously to the classification proposed by Nebert et al.

Cytochrome 2E1 is the basic constituent of the *MEOS* system (8,77), which possesses activity to oxidize, in addition to ethanol, other alcohols, such as butanol and pentanol (78), as well as other xenobiotic compounds, such as aniline, acetaminophen, carbon tetrachloride, acetone, benzene, phenol and N-nitrosodimethylamine, among others (73,78-82). This system possesses a low affinity for ethanol, with a Km of 8-10 mM (8), for which it is considered that in non-alcoholic individuals, its participation is only important at high tissue concentrations of the substrate.

The *MEOS* system catalyzes the oxidation from ethanol to acetaldehyde, coupling this reaction to the oxidation of an *NADPH* molecule, and the reduction of an oxygen molecule to form hydrogen peroxide (83, 84).

$$CH_3CH_2OH + NADPH + O_2 \rightleftharpoons CH_3CHO + NADP^+ + H_2O_2 + H^+$$

The hydrogen peroxide formed during the *MEOS* activity also stimulates non-enzymatic ethanol oxidation through the hydroxyl radical formation in presence of iron chelates (see corresponding section ahead); consequently, during a long time it was considered that *MEOS* possessed two reaction mechanisms: one dependent on and another independent from the hydroxyl radicals formation. It is now clear that the dependent mechanism in reality corresponds to a non-enzymatic oxidation.

The *MEOS* system is the only one whose activity is significantly induced (5 - 10 times) by the presence of ethanol or other molecules such as pyridine, acetone and pyrazole (8,82). The molecular mechanism underlying 2E1 induction remains disputed (8), although it appears that the enzyme concentration can be regulated by post-translational mechanisms such as mRNA stabilization, increased translation of existing mRNA, and inhibition of protein degradation (82).

C. Catalase

This enzyme (H_2O_2:H_2O_2 oxidoreductase; EC 1.11.1.6) is an oligomeric protein with 4 subunits (of 60 kDa each) arranged in tetrahedric form; each subunit functions in an independent manner, and possesses as prosthetic groups a ferric protoporphyrin IX (85) and a *NADP*$^+$ molecule (86). Catalase was one of the first enzymes to be isolated in a highly purified form, and its crystallization (87) from liver extracts marked a milestone in the history of biochemistry (85).

Other different forms of catalase exist (88), especially in prokaryotes, where hexameric structures can be found (*Escherichia coli* and *Bacillus subtilis*), subunits of greater molecular weight (*E. coli*), with heme d-isomers instead of protoporphyrin IX (*E. coli*, *B. subtilis* and *Neurospora crassa*), and additionally, catalases without heme group that contain manganese (*Lactobacillus plantarum* and *Thermoleophilum album*). In spite of this, in eukaryotes and superior animals, catalase is a highly-preserved enzyme which probably arose approximately 2,000 million years ago, when the atmosphere changed from reducing to oxidizing (88).

Catalase is localized in the peroxisomes, and its main activity is the decomposition of hydrogen peroxide molecules:

$$2\,H_2O_2 \rightleftharpoons 2\,H_2O + O_2$$

a reaction that catalyzes with extraordinary efficiency; however, it is also highly efficient in the peroxide dependent oxidation of nitrous, formic and hydrazoic acids, short chain aliphatic alcohols and hydroxylamine (85). All these reactions (except the oxidation of nitrous and hydrazoic acids) can be represented by a general equation of the following type:

$$ROOH + HXOH \rightleftharpoons XO + ROH + H_2O$$

where:

R = [H,alkyl, acyl]

X = [O, NH, C=O, $(CH_2)_{n=1,2,3}$]

In the particular case of ethanol, its oxidation can be coupled to the decomposition of a molecule of hydrogen peroxide, or even to the decomposition of organic peroxide, where:

$$HOOH + CH_3CH_2OH \rightleftharpoons CH_3CHO + 2H_2O$$
$$ROOH + CH_3CH_2OH \rightleftharpoons CH_3CHO + ROH + H_2O$$

However, the limiting factor in the oxidation of ethanol by catalase is the availability of peroxides, so it is considered that under normal conditions, its participation in ethanol metabolism is minimal (8).

D. Non-Enzymatic Oxidation

It has been demonstrated *in vitro* that ethanol oxidation in microsomes is significantly stimulated in the presence of some ferric chelates, such as, Fe^{3+}-EDTA, without the direct participation of cytochromes P-450 (e.g., 23, 80, 89, 90). This non-enzymatic oxidation depends on the hydroxyl radical (·OH) formation from hydrogen peroxide where, by means of a spontaneous reaction between the hydroxyl radical and ethanol, it gives origin to water and an α-hydroxyethyl radical, which, upon interaction with oxygen, will yield acetaldehyde and a protonated superoxide anion (HO_2^-), in a series of reactions such as those illustrated in Figure 6 (80).

The *in vitro* addition of enzymes such as catalase or peroxidase, or even the presence of scavengers of hydroxyl radicals, e.g., benzoate, mannitol or dimethyl sulfoxide, sensitively diminish the non-enzymatic oxidation of ethanol (90), just as would be expected according to the reaction mechanism described previously (see Figure 6); on the other hand, the presence of the superoxide dismutase enzyme (90) provokes insignificant effects, which makes it probable that Haber-Weiss type reactions have little significance in this oxidation mechanism.

It is important to bear in mind that ethanol by itself is an excellent scavenger of hydroxyl radicals (91), because of which this series of reactions is limited only by the availability of hydroxyl radicals, which in turn depends on: 1) the presence of Fe chelates and a reducing system of the latter, susceptible to catalyze the hydroxyl radicals formation from hydrogen peroxide, and 2) the

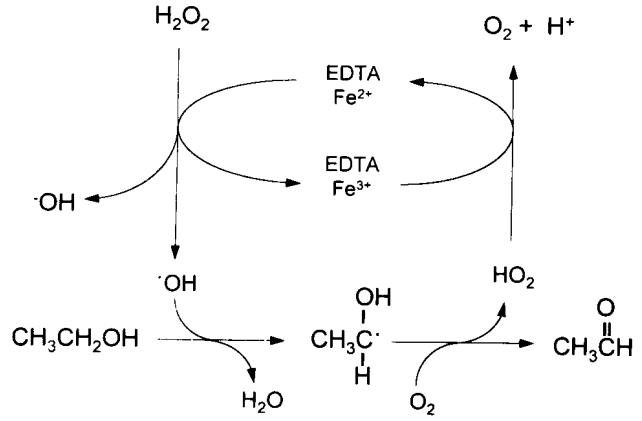


Figure 6. Non-enzymatic mechanism of ethanol oxidation. The hydroxyl radicals (·OH) necessary for ethanol oxidation are formed from hydrogen peroxide by means of the Fenton reaction catalyzed by iron ($M^{n+} + H_2O_2 \rightarrow M^{(n+1)+} + {}^{-}OH + {}^{\cdot}OH$). The hydroxyl radical, on reacting with ethanol, forms transitorily an α-hydroxyethyl radical, which, upon interaction with oxygen, will yield acetaldehyde and a protonated superoxide ion (HO_2^-). These new free radicals may be taken up by the superoxide dismutase enzyme, which catalyzes its dismutation to hydrogen peroxide in a non-illustrated reaction ($2\,HO_2^- \rightarrow O_2 + H_2O_2$); otherwise, this superoxide ion reacts with the Fe-*EDTA* compound, reducing it so that it is ready again in the formation of hydroxyl radicals in a Haber-Weiss type reaction ($Fe^{3+} + O_2^- \rightarrow Fe^{2+} + O_2$ /y/ $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{-}OH + {}^{\cdot}OH$). Modified from Ingelman-Sundberg and Johansson (80).

availability of the hydrogen peroxide itself, necessary for the formation of hydroxyl radicals.

In this way, enzymes and activities that promote H_2O_2 or superoxide anion ($O_2^{\cdot-}$) formation indirectly promote non-enzymatic ethanol oxidation. Among these, enzymes such as xanthine oxidase, dihydro-orotate dehydrogenase, diamine oxidase, tryptophan dehydrogenase and cytochromes P-450, among others, or even cellular activities such as the leukocyte phagocytosis process, or the activity of the mitochondrial electron transport chain itself may be mentioned.

On the other hand, *in vivo,* it is probable that this ethanol oxidation mechanism has an insignificant participation, as the possible organic iron compounds which could be present in the cells (e.g., Fe^{3+}-citrate, Fe^{3+}-ADP, and Fe^{3+}-ATP) are barely efficient for stimulating ethanol oxidation (90, 92). However, it is important to take this mechanism into consideration, because the generation of free radicals during ethanol oxidation can be one of the pathways through which ethanol induces hepatic tissue damage.

It has recently been reported that the heme group of oxyhemoglobin is also able to catalyze ethanol oxidation into acetaldehyde (93), but its contribution to global ethanol metabolism has not yet been evaluated.

E. Non-Oxidative Metabolism

Lange et al. showed, in 1981 (24), that ethanol is capable of reacting with fatty acids to yield ethylic esters from fatty acids by means of an enzymatically catalyzed reaction in which:

$$CH_3CH_2OH + RCOOH \rightleftharpoons H_2O + RCOOCH_2CH_3$$

in this reaction, R corresponds to the hydrocarbon chain of fatty acids such as, for example, to oleic, linoleic and arachidonic acids. The enzyme responsible for this last reaction is the fatty acid ethyl ester synthase, and which is characterized as a 50 KDa dimeric enzyme which does not require ATP or coenzyme A participation to catalyze the reaction (94).

The affinity of this enzyme for ethanol is relatively low, with a Km >0.2 M (94), and its capacity to metabolize ethanol (25) is scarcely on the order of 0.137 nmol g^{-1} h^{-1} in homogenated human liver (see Table 2 for a comparison); therefore, it can be concluded that its participation in global ethanol metabolism is of little significance (23); however, it must be pointed out that organs lacking oxidative ethanol metabolism yet frequently damaged by ethanol abuse have high fatty acid ethyl ester synthetic activities and show substantial transient accumulations of fatty acid ethyl esters in, for example, heart, pancreas and brain, among others (25).

Acetaldehyde Metabolism

Acetaldehyde is the principal metabolite generated in the first stage of hepatic ethanol catabolism; it is a highly reactive molecule and can form adducts with diverse molecules, mainly proteins (95-98). In this way, a great part of the toxic effects associated with ingestion of ethanol is attributed to the formation of acetaldehyde adducts, which inactivate diverse biological molecules (95).

Acetaldehyde, produced mainly in the liver, is metabolized through three metabolic pathways: the first of these is made up of an enzymatic system generically known as aldehyde dehydrogenases or *ALDHs*, which catalyze the oxidation of acetaldehyde to acetate in a reaction that requires *NAD*$^+$ as electron acceptor (9). The second pathway is made up of aldehyde oxidase, a poorly-studied enzyme which catalyzes acetaldehyde oxidation into acetate in an oxygen-dependent reaction. The third pathway, described recently (99), is named the *Microsomal Acetaldehyde Oxidizing System*, also called *MAOS*. This system requires the participation of cytochromes P-450 and carries out the oxidation of acetaldehyde into acetate in a reaction coupled to the oxidation of a reduced nicotinamide adenine dinucleotide phosphate, in a reaction analogous to that achieved by the *Microsomal Ethanol Oxidizing System*.

A. Aldehyde Dehydrogenase System

This system is made up of a series of isozymes denominated aldehyde dehydrogenases *(ALDH*; aldehyde: *NAD*$^+$ oxidoreductase, EC 1.2.1.3), which take charge of oxidizing more than 90% of all acetaldehyde produced from ethanol oxidation, transforming it into acetate in a reaction coupled to *NAD*$^+$ reduction:

$$CH_3CHO + NAD^+ + H_2O \rightleftharpoons CH_3COO^- + NADH + 2\,H^+$$

This reaction, just like *ADH* catalyzed ethanol oxidation, has a bi-bi ordered type mechanism, where the previous union of *NAD*$^+$ is indispensable for the acetaldehyde to penetrate the active site (100-102); however, the rate limiting step for these enzymes is not the dissociation of *NADH*, but the acyl group dissociation of the enzyme (*op cit.*). The reaction mechanism of aldehyde dehydrogenase is shown in Figure 7.

ALDH is responsible not only for acetaldehyde metabolism, but also for the oxidation of other aldehyde groups, such as those present in biogenic monoamines, diamines and polyamines (103), in the metabolism of retinoic acid, or even in the oxidation of aldehydes generated during membrane lipoperoxidation processes (104). Furthermore, *ALDH* possesses an esterase activity, in the absence of *NAD*$^+$, at the same catalytic site that dehydrogenase activity is performed (102,105).

At present, preliminary studies have been accomplished concerning the tridimensional structure of aldehyde dehydrogenase at low resolution (106). However, the assignation and identification of the lateral chains of amino acids in this structure and therefore, the identification of the amino acids which make up the catalytic site (*op cit.*) remain to be elucidated.

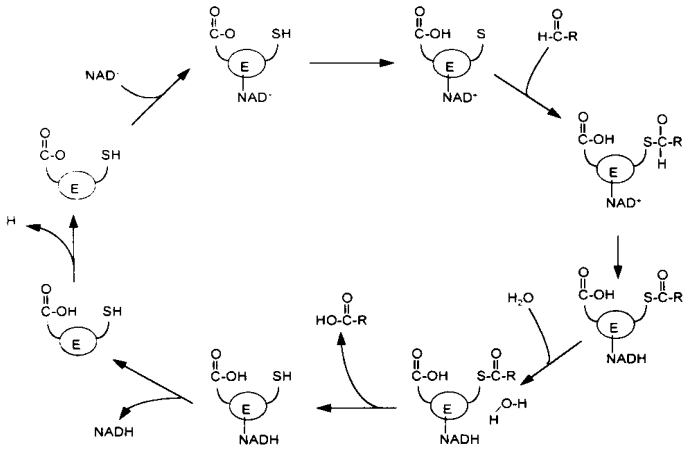


Figure 7. Reaction mechanism of aldehyde dehydrogenase. The reaction starts with the entrance of *NAD*$^+$ to the catalytic site of the enzyme, where the residues of glutamate 268 and cysteine 302 are found (taking as a reference the human Class II enzyme). The role of glutamate 268 is to favor the ionization of the thiol group of cysteine 302, and in this way facilitate the entrance of the aldehyde. This initiates a nucleophylic attack on cysteine 302, forming an intermediate thiohemiacetal, which in turn is dehydrogenated to form an acyl group. Afterward, the entrance of a water molecule allows the liberation of the acyl group as an acid, to finally remove the *NADH* from the enzyme (100-102). Schematic diagram elaborated according to information provided by Brändén et al. (39).

There exist in mammals different isozymes which are grouped in five distinct classes (possibly 8), depending on their structure, their catalytic characteristics and their subcellular localization (9). In general, the degree of similarity within each class is greater than 90%, while the similarity among classes is approximately 70%. Table 5 shows a classification of aldehyde dehydrogenases described up to this date (early 1997). It is interesting to note that each class has a specific cellular localization which prevails even in different species, and this suggests a very early divergence in the evolution of aldehyde dehydrogenases (9).

Class I: This class is composed of homotetrameric enzymes with 54 kDa subunits; these are localized in the cytosolic fraction, and possess a Km for acetaldehyde equal to 30 μM (9, 107). They are encoded by the *ALDH-1* gene, which is found in chromosome 9 (108). *ALDH1* possesses a high affinity for retinal, the active form of Vitamin A, which is oxidized and transformed into retinoic acid, with a Km of 60 nM, at pH 7.5 (109). This piece of data, together with the fact that the gene which encodes for its synthesis is responsive to hormones, suggests that its physiological role is primarily related to the generation of retinoic acid and in the modulation of cellular differentiation in the tissues sensitive to this metabolite (109).

Class II: This class also has a tetrameric structure with 54kDa subunits. It is located principally in the mitochondrial matrix and corresponds to the isozymes of greatest affinity for acetaldehyde, with a Km of 3 μM, because of which it is believed that they are the main enzymes responsible for acetaldehyde oxidation. They are encoded by the *ALDH-2* gene, located in chromosome 12, which possesses two alleles, *ALDH2*1* and *ALDH2*2*. The former manifests an elevated capacity for ethanol oxidation, while the latter is referred to as the "deficient" phenotype because of its low acetaldehyde oxidizing activity. The *ALDH2** 2 "deficient" allele is dominant (110), and is characteristic of Asiatic and Indian populations of South America, with a prevalence of approximately 50%.

The presence of the *ALDH2* 2* allele determines, in those individuals who possess it, a deficient capacity for oxidizing acetaldehyde, which accumulates during the ingestion of alcoholic beverages, producing diverse symptoms such as erythema (vasodilatation and increased blood flow) of the face and upper chest, tachycardia, dizziness, nausea and occasional vomiting, nasal congestion, and pulsating headaches (9). This explains the aversion to the ingestion of alcoholic beverages observed in those individuals who bear this "deficient" phenotype.

Class III: This class is composed of several dimeric isozymes, localized in the cytosol, with 50 kDa subunits. They are encoded by the *ALDH-3* gene (111), and present a relatively high Km for acetaldehyde (in the mM range). They are inducible enzymes whose synthesis is stimulated by the presence of dioxins and polycyclic hydrocarbons (112-116).

Table 5
Aldehyde Dehydrogenase (ALDH) Polymorphisms in Mammals

Class	Gene Locus	Allele[a]	Structure	Distribution
I	*ALDH - 1*	*ALDH1*	Tetramer	Cystol: Most tissues
II	*ALDH - 2*	*ADH2*1*	Tetramer	Mitochondria
II		*ADH2*2*	Tetramer	Most tissues (Liver > kidney > heart)
III	*ALDH - 3*	*ALDH3*	Dimer	Stomach, lung and liver
IV	*ALDH - 4*	*ALDH4*	Dimer	Liver and kidney
?	*ALDH - 5 (ALDH-x)*	?	?	Liver, testes, brain, stomach and others
?	*ALDH - 6*	?	?	Salivary gland
?	*ALDH - 7*	?	?	Kidney and lung
?	*ALDH - 8*	?	?	Parotid gland

Class IV: These enzymes are localized in mitochondria and are encoded by the *ALDH-4* gene, which has been also identified as the γ-semialdehyde glutamic dehydrogenase -EC 1.5.1.12- (117,118). They present, along with the *ALDHs* of Class III, a low affinity for acetaldehyde, while benzaldehyde and medium chain aliphatic aldehydes are its optimal substrates (119).

Lastly, in reference to the *ALDH-5, ALDH-6, ALDH-7* and *ALDH-8* genes, these have been identified by means of reverse transcription through the polymerase chain reaction (108).

The *ALDH-5* gene, first called *ALDH*-x, was identified by Hsu and Chang (120) and is localized in chromosome 9. This gene encodes for a 517 aminoacid polypeptidic chain which shows its greatest similarity to *ALDH1, ALDH2* and *ALDH6* (63-74%). It is found in different tissues, among which are liver, brain, adrenal gland, testicle, stomach and parotid gland (108).

The *ALDH-6* gene is localized in chromosome 15 (15q26) and encodes for a 512 aminoacid polypeptidic chain. Its greatest similarity is to *ALDH1, ALDH2* and *ALDH5* (63-70%). It is found in small amounts in diverse tissues, and only at high levels in salivary glands, stomach and kidney (108).

The *ALDH-7* gene is expressed principally in kidney and lung, and encodes for one 468 aminoacid polypeptidic chain. It is localized in chromosome 11 (11q13) and shares greatest similarity (87%) with the *ALDH-8* gene (108).

The *ALDH-8* gene is expressed only in the parotid gland, and presents the same chromosomal localization as the *ALDH-7* gene, with which it also shares the greatest similarity (108), because of which it is possible for both to belong to the same class of *ALDHs*. This gene encodes for a 466 aminoacid polypeptide.

B. Aldehyde Oxidase

This enzyme (aldehyde: oxygen oxidoreductase; EC 1.2.3.1) is localized in the cytosol of the cell, presents a molecular weight of 300 kDa, and possesses eight atoms of iron, two of molybdenum, two *FAD* molecules and one or two Q_{10} molecules (121).

Aldehyde oxidase catalyzes acetaldehyde oxidation to convert it into acetate in a reaction coupled with oxygen consumption and hydrogen peroxide formation, and possesses a Km for acetaldehyde of 1 mM (122).

$$CH_3CHO + O_2 + H_2O \rightleftharpoons CH_3COO^- + H_2O_2 + H^+$$

In addition to acetaldehyde, this enzyme is also able to catalyze the oxidation of an ample variety of heterocyclic compounds which contain nitrogen such as, for example, purine and pyrimidine derivatives (122-124).

Although the participation of this enzyme is not significant in comparison to the aldehyde dehydrogenase system for global acetaldehyde oxidation, it is nonetheless important to take it into consideration, given that part of the toxic effects of this metabolite has been associated with the generation of free radicals, derived from the production of hydrogen peroxide through aldehyde oxidase reaction (125,126).

C. Microsomal Acetaldehyde Oxidizing System (MAOS)

This system, described in 1996 by Kunitoh et al. (99) in hepatic microsomes, is analogous to the *MEOS* system responsible for ethanol oxidation. It requires the participation of P-450 cytochromes, and especially the CYP2E1 cytochrome, which is the most active. This system couples acetaldehyde oxidation to acetate with the simultaneous oxidation of a molecule of *NADPH* and the reduction of a molecule of oxygen to form hydrogen peroxide.

$$CH_3CHO + NADPH + O_2 + H_2O \rightleftharpoons CH_3COO^- + NADP^+ + H_2O_2 + 2H^+$$

Additionally, the activity of this system is induced 2-fold in animals chronically treated with ethanol. However, this system is as yet slightly characterized, because of which its role and physiological relevance are still to be elucidated.

Discussion

It is unreasonable to expect that these very complex enzymatic systems, with an ample diversity of isozymes and possible substrates, have been formed throughout the lapse of vertebrate evolution with the only finality being of oxidizing a sole exogenous metabolite which is found in minimal amounts under natural conditions. In fact, the alcohol and aldehyde dehydrogenase systems, as was mentioned before, also participate in several physiological functions, catalyzing the oxidation or reduction of diverse endogenous substrates, such as those involved in retinoic acid synthesis (103,127-131), norepinephrine catabolism (57,132), leukotriene catabolism (133), bile acid synthesis (134), serotonine and dopamine catabolism (103, 135-138), or in the detoxification of cytotoxic products of lipoperoxidation such as 4-hydroxinonenal (129). In addition, aldehyde dehydrogenase participates in GABA synthesis (138), and polyamine and histamine catabolism (103,138).

Table 6 shows the reported endogenous substrates for alcohol and aldehyde dehydrogenase isozymes. As can be seen in this Table, most of the endogenous substrates listed for alcohol and aldehyde dehydrogenase comprise important modulators of metabolism, differentiation, and neuroendocrine functions. Furthermore, the *ADH* affinity for these endogenous substrates is higher than ethanol, emphasizing the fact that surely ethanol does not correspond to the natural substrate for this enzyme family. Only Class II *ALDH* possesses a high affinity for

Table 6

Metabolic Pathways, Substrates and Kinetic Parameters on which Alcohol- (*ADH*) and Aldehyde (*ALDH*) Dehydrogenase Participate in Mammals

Metabolic pathway	Substrates	Enzyme / Reaction	Km(μM)	Reference
Norepinephrine catabolism (oxidation / reduction of norepinephrine derivatives)	4-hydroxy-3-methoxy-phenyl glicol	ADH class I (oxidation)	440 - 5000	57, 132
	3,4-dihydroxyphenyl glycol	ADH class I (oxidation)	1100 - 3400	57, 132
	4-hydroxy-3-methoxymandelaldehyde	ADH class II (reduction)	120	57
		ALDH (oxidation)	ND	57
	3,4-dihydroxymandelaldehyde	ADH class II (reduction)	55	57
		ALDH (oxidation)		57
Leukotriene catabolism (ω-oxidation)	20-hydroxy-LTB_4	ADH class III (oxidation)	83	133
	20-oxo-LTB_4	ALDH (oxidation)	ND	133
Retinoic acid metabolism	all-*trans*-retinol	ADH class I (oxidation)	45 - 290	127-129
		ADH class II (oxidation)	14	127
		ADH class IV (oxidation)	20 - 31	127, 129
	13-*cis*-retinol	ADH class I (oxidation)	257	128
	all-*trans*-retinal	ALDH class I (oxidation)	0.06 - 1.1	103,130,131
		ALDH class II (competitive inhibitor)	K_i = 0.043	130
		ALDH class III (oxidation)	8	103
	9-*cis*-retinal	ADH class I (reduction)	13 - 340	127, 129
		ADH class II (reduction)	11	127
		ADH class IV (reduction)	30 - 43	127, 129
		ALDH class I, II and III (oxidation)	ND	103
	13-*cis*-retinal	ADH class I (reduction)	13 - 242	127-129
		ADH class II (reduction)	13	127
		ADH class IV (reduction)	26	127, 129
		ALDH class I (oxidation)	0.37	130
		ALDH class II (competitive inhibitor)	K_i = 5	130
Serotonine metabolism	5-hydroxytryptophol (5-HTOL)	ADH class I (oxidation)	170 - 800	135
		ADH class II (oxidation)	280	135
	5-hydroxyindol-acetaldehyde (5-HIAAL)	ALDH class I (oxidation)	2.4	103
		ALDH class II (oxidation)	0.8	103
Dopamine catabolism	3,4-dihydroxyphenyl-acetaldehyde (DHPAL)	ADH class I (reduction)	8.9 - 720	136
		ALDH class I (oxidation)	0.4	103
		ALDH class II (oxidation)	1.0	103
		ALDH class III (oxidation)	2.6	103, 138
	3,4-dihydroxyphenyl-ethanol (DHPE)	ADH class I (oxidation)	15 - 400	136
	4-hydroxy-3-methoxy-phenyl-acetaldehyde (HMPAL)	ADH class I (reduction)	11 - 240	136
		ALDH (oxidation)	ND	137
	4-hydroxy-3-methoxy-phenyl-ethanol (HMPE)	ADH class I (oxidation)	40 - 230	136
GABA synthesis	γ-aminobutyraldehyde	ALDH class III (oxidation)	46	138
Diamine and polyamine catabolism	aldehyde from spermidine	ALDH class III (oxidation)	71	103, 138
	monoaldehyde from spermine	ALDH class III (oxidation)	164	103, 138
	aldehyde from cadaverine	ALDH class III (oxidation)	185	103

Histamine catabolism	Imidazoleacetaldehyde	ALDH class I (oxidation)	39	103
		ALDH class II (oxidation)	30	103
		ALDH class III (oxidation)	59	103, 138
Bile acid synthesis	5β-cholestane-3α,7α,12α,26-tetrol	ADH class I (oxidation)	32	134
	3α,7α,12α-trihydroxy-5β-cholestan-26-al	ALDH (oxidation)	ND	134
		ADH class I (reduction)	10.9	
Ethanol metabolism	Ethanol	ADH class I (oxidation)	50 - 36000	139
		ADH class II (oxidation)	34000	139
		ADH class III (oxidation)	> 1000000	139
		ADH class IV (oxidation)	3600 - 41000	139
	Acetaldehyde	ALDH class I (oxidation)	50 - 180	103, 130
		ALDH class II (oxidation)	0.2 - 1	103, 130
		ALDH class III (oxidation)	57	103, 138
Formaldehyde detoxification	S-hydroxymethyl glutathione	ADH class III (oxidation)	0.92 - 1.9	33
	Formaldehyde	ALDH class I (oxidation)	330	103
		ALDH class II (oxidation)	320	130
		ALDH class III (oxidation)	410	103
Lipid peroxidation products	4-hydroxynonenal	ADH class I (reduction)	150	129
		ADH class III (reduction)	8300	129
		ADH class IV (reduction)	1300	129
		ALDH (oxidation)	ND	129

Values are for pH 7.0-7.5, 25-30°C, except in Reference 134, which are for pH 9.0, and in Reference 127, and class III ADH in Reference 139, which are for pH 10.0.
ND, not determined.

acetaldehyde, but this is comparable to other important endogenous substrates. On the other hand, it is noteworthy that the majority of *ALDH* substrates are, in turn, *ADH* products, underlining the concerted fashion in which both enzymatic systems work.

At this point, it should be mentioned that in many of the metabolic pathways on which alcohol and aldehyde dehydrogenase participate, cytochromes P-450 also are involved, for instance, in retinoic acid metabolism (140), bile acid synthesis (141), metabolism of prostaglandins derivatives (142), ω—oxidation of long chain fatty acids (70,74-76), and xenobiotic metabolism (70,74-76).

In addition, the expression of both alcohol and aldehyde dehydrogenase genes is regulated by some important metabolites, such as retinoic acid (143-145), thyroid hormone (145-148), glucocorticoids (149-154), androgens (131,145,154-159), estrogens (150,154, 160,161), insulin (145,162,163), and cyclic AMP (145,164), among others (113,116,165-172). It is very important to point out that many of the metabolic modulators mentioned above are also regulators of cytochrome P-450 gene expression (173). Table 7 shows the reported modulators of *ADH*, *ALDH* and P-450 gene expression.

The information presented in this review suggests a much more complex physiological role for alcohol and the aldehyde dehydrogenase system than could have previously been expected, which is in accordance with the hypothesis conceived by Daniel W. Nebert (173) concerning the proposed role of drug-metabolizing enzymes as regulators of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. In this way, these three enzymatic systems, made up of about 7 *ADH*, 8 *ALDH*, and more than 80 P-450 isozymes, possibly evolved in vertebrates in a similar manner and with a common goal. In fact, the complementary reactions developed by these enzymatic systems on diverse hydrophobic substrates (initial hydroxilations by P-450, alcohol oxidation by *ADH* and finally, aldehyde oxidation by *ALDH*), are surely not just coincidences.

It is difficult to establish with certainty the evolutive forces that directed the enzymogenesis of alcohol and aldehyde dehydrogenase systems in vertebrates, but some insights may be mentioned. In the first instance, it should be considered that among all living things, vertebrates (with the mammals at the top) are the organisms with the highest cell-type number (174), and therefore, with the most stringent necessities for development and maintenance of cellular differentiation. This selective pressure might force vertebrates to develop not only a more complex collection of differentiation and growth modulators, but also a complete collection of enzymes qualified to regulate the synthesis and degradation of this kind of metabolites.

Table 7
Modulators of Alcohol Dehydrogenase, Aldehyde Dehydrogenase and Cytochrome P-450 Gene Expression and/or Activity in Mammals

Alcohol dehydrogenase	Aldehyde dehydrogenase	Cytochrome P450
Retinoids	Retinoids	Retinoids
Thyroid hormone	Thyroid hormone	Thyroid hormone
Glucocorticoids	Glucocorticoids	Glucocorticoids
Androgens	Androgens	Androgens
Estrogens	Estrogens	Estrogens
	Progestins	Progestins
Growth hormone	Clofibrate	Clofibrate
Cyclic AMP	Cyclic AMP	Mineralocorticoids
Insulin	Insulin	Bile acids
Epinephrine	Calcium	Cholestereol
Dopamine-derived metabolites	Glutathione	Dopamine
Serotonin-derived metabolites	Non-steroidal, anti-inflammatory drugs	Arachidonic acid
Glucagon	Phenobarbital	Phenobarbital
	Dioxins	Dioxins
	Polycyclic hydrocarbons	Polycyclic hydrocarbons
	Ethanol	Ethanol
		Vitamin D

For references, see text.

The above-mentioned properties and enzymogenesis of alcohol dehydrogenase, aldehyde dehydrogenase, and cytochrome P-450 systems in vertebrates are in accordance with the hypothesis that these protein families developed through vertebrate evolution in response to an evermore complex development and differentiation process.

Thus, the participation of alcohol dehydrogenase, aldehyde dehydrogenase and cytochrome P-450 systems on multiple cellular function related to the anabolism and/or catabolism of metabolites that regulate growth, metabolism, differentiation and neuroendocrine function in mammals, opens a broad stage to analyze the possible consequences of excessive ethanol ingestion. With this novel view in mind, metabolic alterations induced by ethanol reach a new scope, and diverse deleterious effects, such as fetal alcohol syndrome (175), ethanol carcinogenesis (176), conductual alterations (177), or metabolic derangements (10,178,179), just to mention some examples, could be revisited.

Corollary

It is of great significance that, although the systems through which ethanol and acetaldehyde can be metabolized have been thoroughly studied, the mechanism or mechanisms by which ethanol induces its toxic effects are still unclear. In fact, the physiological role of alcohol and aldehyde dehydrogenase systems scarcely begin to be understood. In this sense, there is much which still must be defined, but surely, the understanding of the physiological role of these enzymatic systems will throw an important light on the mechanisms of ethanol toxicity and the current treatments to ameliorate alcohol-induced derangements.

Acknowledgments

The authors thank Dr. Rafael Villalobos-Molina for discussions, suggestions and encouragement, and Dr. Raúl Ondarza for the invaluable advice generously offered at the XXI National Congress of Biochemistry held in Manzanillo, Colima, Mexico (November 3 - 7, 1996). and Dr. Xavier Parés (Universitat Autònoma de Barcelona, Spain).

References

1. Loeb EM. Primitive intoxicants. Q J Stud Alcohol 1943; 4:387.
2. Popham RE. The social history of the tavern. In: Israel Y, Glaser FB, Kalant H, Popham RE, Schmidt W, Smart RG, editors. Research Advances in Alcohol and Drug Problems, Vol 4. New York: Plenum Press; 1978. p. 225.
3. Dixon M, Lutwak-Mann C. Aldehyde mutase. Biochem J 1937; 31:1347.
4. Leloir LF, Muñoz JM. Ethyl alcohol metabolism in animal tissues. Biochem J 1938; 32:299.
5. Lutwak-Mann C. Alcohol dehydrogenase of animal tissues. Biochem J 1938; 32:1364.
6. Pyörälä E. Trends in alcohol consumption in Spain, Portugal, France and Italy from the 1950s until 1980s. Br J Addict 1990; 85:469.

7. Molina-Piñeiro V, Sánchez-Medal L, editors. El alcoholismo en México, Vol 1-4, Second ed. México: Fundación de Investigaciones Sociales; 1985.
8. Lieber CS. Alcohol and the liver: 1994 update. Gastroenterol 1994; 106:1085.
9. Arnon R, Degli Esposti S, Zern MA. Molecular biological aspects of alcohol-induced liver disease. Alcohol Clin Exp Res 1995; 19:247.
10. Lands WEM. Cellular signals in alcohol-induced liver injury: a review. Alcohol Clin Exp Res 1995; 19:928.
11. Erickson CK. Factors affecting the distribution and measurements of ethanol in the body. In: Majchrowicz E, Noble EP, editors. Biochemistry and Pharmacology of ethanol. New York: Plenum Press; 1979. p. 9.
12. Passmore R. The energy value of alcohol. In: Gastineau CF, Darby WJ, Turner TB, editors. Fermented Food Beverages in Nutrition. New York: Academic Press; 1979. p. 213.
13. Agarwal DP, Goedde HW. Enzymology of alcohol degradation. In: Goedde HW, Agarwal DP, editors. Alcoholism: Biomedical and Genetic Aspects. New York: Pergamon Press; 1989. p. 3.
14. McManus IR, Contag AO, Olson RE. Characterization of endogenous ethanol in the mammal. Science 1960; 131:102.
15. Krebs HA, Perkins JR. The physiological role of liver alcohol dehydrogenase. Biochem J 1970; 118:635.
16. Baraona E, Julkunen R, Tannenbaum L, Lieber CS. Role of intestinal bacterial overgrowth in ethanol production and metabolism in rats. Gastroenterol 1986; 90:103.
17. Batt RD. Absorption, distribution, and elimination of alcohol. In: Crow KE, Batt RD, editors. Human Metabolism of Alcohol, Vol. I: Pharmacokinetics, Medicolegal Aspects, and General Interest. Boca Raton, FL: CRC Press; 1989. p. 3.
18. Watson PE. Total body water and blood alcohol levels: updating the fundamentals. In: Crow KE, Batt RD, editors. Human metabolism of alcohol, Vol. I: Pharmacokinetics, Medicolegal Aspects, and General Interest. Boca Raton, FL: CRC Press; 1989. p. 41.
19. Cooke AR, Birchall A. Absorption of ethanol from the stomach. Gastroenterol 1969; 57:269.
20. Smith T, DeMaster EG, Furne JK, Springfield J, Levitt MD. First-pass gastric mucosal metabolism of ethanol is negligible in the rat. J Clin Inv 1992; 89:1801.
21. Horowitz M, Maddox A, Bochner M, Wishart J, Bratasiuk R, Collins P, Shearman D. Relationships between gastric emptying of solid and caloric liquid meals and alcohol absorption. Am J Physiol 1989; 257 (Gastrointest Liver Physiol. 20): G291.
22. Roine RP, Gentry T, Lim Jr RT, Baraona E, Lieber CS. Effect of concentration of ingested ethanol on blood alcohol levels. Alcohol Clin Exp Res 1991; 15:734.
23. Koop DR. Minor pathways of ethanol metabolism. In: Crow KE, Batt RD, editors. Human Metabolism of Alcohol, Vol. II: Regulation, Enzymology, and Metabolites of Ethanol. Boca Raton, FL: CRC Press; 1989. p. 133.
24. Lange LG, Bergman SR, Sobel BE. Identification of fatty acid ethyl esters as products of rabbit myocardial ethanol metabolism. J Biol Chem 1981; 256:12968.
25. Laposata EA, Lange LG. Presence of nonoxidative ethanol metabolism in human organs commonly damaged by ethanol abuse. Science 1986; 231:497.
26. Jörnvall H, Danielsson O, Eklund H, Hjelmqvist L, Höög J-O, Parés X, Shafqat J. Enzyme and isozyme development within the medium-chain alcohol dehydrogenase family. Adv Exp Med Biol 1993; 328:533.
27. Jörnvall H, von Bahr-Lindström H, Höög J-O. Alcohol dehydrogenases-structure. In: Crow KE, Batt RD, editors. Human Metabolism of Alcohol, Vol. II: Regulation, Enzymology, and Metabolites of Ethanol. Boca Raton, FL: CRC Press; 1989. p. 43.
28. Persson B, Krook M, Jörnvall H. Characteristics of short-chain alcohol dehydrogenases and related enzymes. Eur J Biochem 1991; 200:537.
29. Jörnvall H, von Bahr-Lindström H, Jany K-D, Ulmer W, Fröschle M. Extended super family of "short" alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases. FEBS Lett 1984; 165:190.
30. Krozowski Z. The short-chain alcohol dehydrogenase superfamily: Variations on a common theme. J Steroid Biochem Molec Biol 1994; 51:125.
31. Danielsson O, Atrain S, Luque T, Hjelmqvist L, González-Duarte R, Jörnvall H. Fundamental molecular differences between alcohol dehydrogenases classes. Proc Natl Acad Sci USA 1994; 91:4980.
32. Koivusalo M, Baumann M, Uotila L. Evidence for the identity of glutathione-dependent formaldehyde dehydrogenase and class III alcohol dehydrogenase. FEBS Lett 1989; 257:105.
33. Uotila L, Koivusalo M. Glutathione-dependent oxidoreductases: Formaldehyde dehydrogenase. In: Dolphin D, Poulson R, Avramovic O, editors. Coenzymes and Cofactors, Glutathione: Chemical, Biochemical and Medical Aspects. Vol III Part A. New York: Wiley; 1989. p.517.
34. Scopes RK. An iron-activated alcohol dehydrogenase. FEBS Lett 1983; 156:303.
35. Boleda MD, Julià P, Moreno A, Parés X. Role of extrahepatic alcohol dehydrogenase in rat ethanol metabolism. Arch Biochem Biophys 1989; 274:74.
36. Jelokova J, Karlsson C, Estonius M, Jörnvall H, Höög JO. Features of structural zinc in mammalian alcohol dehydrogenase. Site-directed mutagenesis of the zinc ligands. Eur J Biochem 1994; 225:1015.
37. Eklund H, Nordström B, Zeppezauer E, Söderlund G, Ohlsson I, Boiwe T, Söderberg B-O, Tapia A, Brändén C-I, Åkeson Å. Three-dimentional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. J Mol Biol 1976; 102:27.
38. Eklund H, Samama J-P, Wallén L, Brändén C-I, Åkeson Å, Jones TA. Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. J Mol Biol 1981; 146:561.
39. Brändén C-I, Jörnvall H, Eklund H, Furugren B. Alcohol dehydrogenases. In: Boyer PD, editor. The Enzymes, Vol. XI: Oxidation-Reduction, part A: Dehydrogenases (I), Electron Transfer (I). New York: Academic Press; 1975. p. 103.
40. Shearer GL, Kim K, Lee KM, Wang CK, Plapp BV. Alternative pathways and reactions of benzyl alcohol and benzaldehyde with horse liver alcohol dehydrogenase. Biochemistry 1993; 32:11186.
41. Burnell JC, Bosron WF. Genetic polymorphism of human liver alcohol dehydrogenase and kinetic properties of the isoenzymes. In: Crow KE, Batt RD, editors. Human Metabolism of Alcohol, Vol. II: Regulation, Enzymology, and Metabolites of Ethanol. Boca Raton, FL: CRC Press; 1989. p.65.
42. Fan F, Plapp BV. Substitutions of isoleucine residues at the adenine binding site activate horse liver alcohol dehydrogenase. Biochemistry 1995; 34:4709.

43. Ramaswamy S, Eklund H, Plapp BV. Structures of horse liver alcohol dehydrogenase complexed with NAD^+ and substituted benzyl alcohols. Biochemistry 1994; 33:5230.
44. Hurley TD, Bosron WF, Hamilton JA, Amzel LM. Structure of human $\beta_1\beta_1$ alcohol dehydrogenase: Catalytic effects of non-active-site substitutions. Proc Natl Acad Sci USA 1991; 88:8149.
45. Hurley TD, Bosron WF, Stone CL, Amzel LM. Structures of three human beta alcohol dehydrogenase variants. Correlations with the functional differences. J Mol Biol 1994; 239:415.
46. Ramaswamy S, El-Ahmad M, Danielsson O, Jörnvall H, Eklund H. Crystallisation and crystallographic investigations of cod alcohol dehydrogenase class I and class III enzymes. FEBS Lett 1994; 350:122.
47. Pietruszko R. Nonethanol substrates of alcohol dehydrogenase. In: Majchrowicz E, Noble EP, editors. Biochemistry and Pharmacology of Ethanol, Vol. 1. New York: Plenum Press; 1979. p. 87.
48. Waller G, Theorell H, Sjövall J. Liver alcohol dehydrogenase as a 3-β-hydroxy-5-β-cholanic acid dehydrogenase. Arch Biochem Biophys 1965; 111:671.
49. Björkhem I. On the role of alcohol dehydrogenase in ω-oxidation of fatty acids. Eur J Biochem 1972; 30:441.
50. Henehan GTM, Oppenheimer NJ. Horse liver alcohol dehydrogenase-catalyzed oxidation of aldehydes: Dismutation precedes net production of reduced nicotinamide adenine dinucleotide. Biochemistry 1993; 32:735.
51. Oppenheimer NJ, Henehan GTM. Horse liver alcohol dehydrogenase-catalyzed aldehyde oxidation. Adv Exp Med Biol 1995; 372:407.
52. Jörnvall H, Höög J-O, von Bahr-Lindström H, Vallee BL. Mammalian alcohol dehydrogenases of separate classes: Intermediates between different enzymes and intraclass isozymes. Proc Natl Acad Sci USA 1987; 84:2580.
53. Parés X, Cederlund E, Moreno A, Hjelmqvist L, Farrés J, Jörnvall H. Mammalian class IV alcohol dehydrogenase (stomach alcohol dehydrogenase): Structure, origin, and correlation with enzymology. Proc Natl Acad Sci USA 1994; 91:1893.
54. Sun H-W, Plapp BV. Progressive sequence alignment and molecular evolution of the Zn-containing alcohol dehydrogenase family. J Mol Evol 1992; 34:522.
55. Kedishvili NY, Bosron WF, Stone CL, Hurley TD, Peggs CF, Thomasson HR, Popov KM, Carr LG, Edenberg HJ, Li TK. Expression and kinetic characterization of recombinant human stomach alcohol dehydrogenase. Active-site aminoacid sequence explains substrate specificity compared with liver isozymes. J Biol Chem 1995; 270:3625.
56. Kamlay MT, Shore JD. Transient kinetic studies of substrate inhibition in the horse liver alcohol dehydrogenase reaction. Arch Biochem Biophys 1983; 222:59.
57. Mardh G, Dingley AL, Auld DS, Vallee BL. Human class π (pi) alcohol dehydrogenase has a redox-specific function in norepinephrine metabolism. Proc Natl Acad Sci USA 1986; 83:8908.
58. Bosron WF, Li T-K, Dofeldecker W, Vallee BL. Human liver pi-alcohol dehydrogenase: Kinetics and molecular properties. Biochemistry 1979; 18:1101.
59. Bosron WF, Lumeng L, Li T-K. Catalytic properties of human liver dehydrogenase isoenzymes. Mol Aspects Med 1988; 10:147.
60. Beisswenger TB, Holmquist B, Vallee BL. χ-ADH is the sole alcohol dehydrogenase isozyme of mammalian brains: Implications and inferences. Proc Natl Acad Sci USA 1985; 82:8369.
61. Caballería J, Baraona E, Rodamilans M, Lieber CS. Effects of cimetidine on gastric alcohol dehydrogenase activity and blood ethanol levels. Gastroenterology 1989; 96:388.
62. Caballería J, Baraona E, Deulofeu R, Hernández-Muñoz R, Rodes J, Lieber CS. Effects of H_2-receptor antagonists on gastric alcohol dehydrogenase activity. Dig Dis Sci 1991; 36:1673.
63. Roine R, Gentry RT, Hernández-Muñoz R, Baraona E, Lieber CS. Aspirin increases blood alcohol concentrations in humans after ingestion of ethanol. JAMA 1990; 264:2406.
64. DiPadova C, Roine R, Frezza M, Gentry RT, Baraona E, Lieber CS. Effects of ranitidine on blood alcohol levels after ethanol ingestion: Comparison with other H_2-receptor antagonists. JAMA 1992; 267:83.
65. Zheng Y-W, Bey M, Liu H, Felder MR. Molecular basis of the alcohol dehydrogenase-negative deer mouse: Evidence for deletion of the gene for class I enzyme and identification of a possible new enzyme class. J Biol Chem 1993; 268:24933.
66. Höög J-O, Brandt M. Mammalian class VI alcohol dehydrogenase: Novel types of the rodent enzymes. Adv Exp Med Biol 1995; 372:355.
67. Lieber CS, DeCarli LM. Ethanol oxidation by hepatic microsomes: adaptative increase after ethanol feeding. Science 1968; 162:917.
68. Rubin E, Hutterer F, Lieber CS. Ethanol increases hepatic smooth endoplasmic reticulum and drug metabolizing enzymes. Science 1968; 159:1469.
69. Orme-Johnson WH, Ziegler DM. Alcohol mixed function oxidase activity of mammalian liver microsomes. Biochem Biophys Res Commun 1965; 21:78.
70. Gonzalez FJ. Human cytochromes P450: problems and prospects. Trends Pharmacol Sci 1992; 13:346.
71. Degtyarenko KN, Archakov AI. Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS Lett 1993; 332:1.
72. Beaune P. Les cytochromes P450 humains: Applications en pharmacologie. Therapie 1993; 48:521.
73. Koop DR, Casazza JP. Identification of ethanol-inducible P-450 isozyme 3a as the acetone and acetol monooxygenase of rabbit microsomes. J Biol Chem 1985; 260:13607.
74. Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W, Phillips IR, Sato R, Waterman MR. The P450 superfamily: Updated listing of all genes and recommended nomenclature for the chromosomal loci. DNA 1989; 8:1.
75. Nebert DW, Nelson DR, Coon MJ, Estabrook RW, Feyereisen R, Fujii-Kuriyama Y, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Loper JC, Sato R, Waterman MR, Waxman DJ. The P450 superfamily: Update on new sequences, gene mapping, and recommended nomenclature. DNA Cell Biol 1991; 10:1.
76. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW. The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol 1993; 12:1.
77. Lasker JM, Raucy J, Kubota S, Bloswick BP, Black M, Lieber CS. Purification and characterization of human liver cytochrome P-450-ALC. Biochem Biophys Res Commun 1987; 148:232.
78. Morgan ET, Koop DR, Coon MJ. Catalytic activity of cytochrome P-450 isozyme 3a isolated from liver microsomes of ethanol-treated rabbits. J Biol Chem 1982; 257:13951.
79. Morgan ET, Koop DR, Coon MJ. Comparison of six rabbit liver cytochrome P-450 isozymes in formation of a reactive metabo-

lite of acetaminophen. Biochem Biophys Res Commun 1983; 112:8.
80. Ingelman-Sundberg M, Johansson I. Mechanisms of hydroxyl radical formation and ethanol oxidation by ethanol-inducible and other forms of rabbit liver microsomal cytochromes P-450. J Biol Chem 1984; 259:6447.
81. Yang CS, Tu YY, Koop DR, Coon MJ. Metabolism of nitrosamines by purified rabbit liver cytochrome P-450 isozymes. Cancer Res 1985; 45:1140.
82. Koop DR, Tierney DJ. Multiple mechanisms in the regulation of ethanol-inducible cytochrome P450IIE1. BioEssays 1990; 12:429.
83. Teschke R, Gellert J. Hepatic microsomal ethanol-oxidizing system (MEOS): Metabolic aspects and clinical implications. Alcohol Clin Exp Res 1986; 10:20s.
84. Zentella de Piña M, Piña Garza E. Metabolitos del etanol. Mensaje Bioquímico (Mex.) 1987; 10:143. (MARIA - VER APUNTOS A MANO DEL REVISOR. GRACIAS; M.)
85. Schonbaum GR, Chance B. Catalase. In: Boyer PD, editor. The Enzymes. Vol. XIII: Oxidation-Reduction, part C: Dehydrogenases (II), Oxidases (II), Hydrogen Peroxide Cleavage, Third ed. New York: Academic Press; 1976. p. 363.
86. Fita I, Silva AM, Murthy MRN, Rossmann MG. The refined structure of beef liver catalase at 2.5 Å resolution. Acta Crist 1986; B42:497.
87. Sumner JB, Dounce AL. Crystalline catalase. J Biol Chem 1937; 121:417.
88. von Ossowski I, Hausner G, Loewen PC. Molecular evolutionary analysis based on the amino acid sequence. J Molec Evol 1993; 37:71.
89. Cederbaum AI, Dicker E. Inhibition of microsomal oxidation of alcohols and of hydroxyl-radical-scavenging agents by the iron-chelating agent desferrioxamine. Biochem J 1983; 210:107.
90. Dicker E, Cederbaum AI. Generation of reactive oxygen species and reduction of ferric chelates by microsomes in the presence of a reconstituted system containing ethanol, NAD^+ and alcohol dehydrogenase. Alcohol Clin Exp Res 1990; 14:238.
91. Cederbaum AL. Microsomal generation of hydroxyl radicals: Its role in microsomal ethanol oxidizing system (MEOS) activity and requirement for iron. Ann NY Acad Sci 1987; 492:35.
92. Feierman DE, Winston GW, Cederbaum AI. Ethanol oxidation by hydroxyl radicals: role of iron chelates, superoxide, and hydrogen peroxide. Alcohol Clin Exp Res 1985; 9:95.
93. Chen HM, Lin WW, Ferguson KH, Scott BK, Peterson CM. Studies of the oxidation of ethanol to acetaldehyde by oxyhemoglobin using fluorigenic high-performance liquid chromatography. Alcohol Clin Exp Res 1994; 18:1202.
94. Mogelson S, Lange LG. Nonoxidative ethanol metabolism in rabbit myocardium: Purification to homogeneity of fatty acyl ethyl ester synthase. Biochemistry 1984; 23:4075.
95. Sorrell MF, Tuma DJ. Hypothesis: alcohol liver injury and the covalent binding of acetaldehyde. Alcohol Clin Exp Res 1985; 9:306.
96. Lucas D, Ménez JF, Bodénez P, Baccino E, Bardou LG, Floch HH. Acetaldehyde adducts with haemoglobin: determination of acetaldehyde released from haemoglobin by acid hydrolysis. Alcohol Alcohol 1988; 23:23.
97. Hernández-Muñoz R, Baraona E, Blacksberg I, Lieber CS. Characterization of the increased binding of acetaldehyde to red blood cells in alcoholics. Alcohol Clin Exp Res 1989; 13:654.
98. Worral S, De Jersey J, Shanley BC, Wilce PA. Antibodies against acetaldehyde-modified epitopes: Presence in alcoholic, non-alcoholic liver disease and control subjects. Alcohol Alcohol 1990; 25:509.
99. Kunitoh S, Asai H, Imaoka S, Funae Y, Monna T. Metabolism of acetaldehyde to acetate by rat hepatic P-450s: Presence of different metabolic pathway from acetaldehyde dehydrogenase system. Alcohol Clin Exp Res 1996; 20:22A.
100. Farrés J, Wang TTY, Cunningham SJ, Weiner H. Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis. Biochemistry 1995; 34:2592.
101. Wang X, Weiner H. Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis. Biochemistry 1995; 34:237.
102. Pietruszko R, Abriola DP, Blatter EE, Mukerjee N. Aldehyde dehydrogenase: aldehyde dehydrogenation and ester hydrolysis. Adv Exp Med Biol 1993; 328:221.
103. Ambroziak W, Pietruszko R. Human aldehyde dehydrogenase: Activity with aldehyde metabolites of monoamines, diamines, and polyamines. J Biol Chem 1991; 266:13011.
104. Mitchell DY, Petersen DR. Oxidation of aldehydic products of lipid peroxidation by rat liver microsomal aldehyde dehydrogenase. Arch Biochem Biophys 1989; 269:11.
105. Kitson KE, Blythe TJ, Kitson TM. Studies of the esterase activity of cytosolic aldehyde dehydrogenase using sterically hindered and cyclic substrates. Adv Exp Med Biol 1995; 372:45.
106. Sun J, Hempel J, Lindahl R, Perozich J, Rose J, Wang B-C. Progress toward the tertiary structure of (class III) aldehyde dehydrogenase. Adv Exp Med Biol 1995; 372:71.
107. Lundquist F. Acetaldehyde and aldehyde dehydrogenases: Central problems in the study of alcoholism. Eur J Clin Invest 1983; 13:183.
108. Hsu LC, Chang W-C, Lin SW, Yoshida A. Cloning and characterization of genes encoding four additional human aldehyde dehydrogenase isozymes. Adv Exp Med Biol 1995; 372:159.
109. Yoshida A, Hsu LC, Dave V. Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. Enzyme 1992; 46:239.
110. Crabb DW, Edenberg HJ, Bosron WF, Li T-K. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity: The inactive *ALDH2*^{2} allele is dominant. J Clin Invest 1989; 83:314.
111. Hsu LC, Yoshida A. Human stomach aldehyde dehydrogenase, $ALDH_3$. Adv Exp Med Biol 1993; 328:141.
112. Marselos M, Lindahl R. Substrate preference of a cytosolic aldehyde dehydrogenase inducible in rat liver by treatment with 3-methylcholanthrene. Toxicol Appl Pharmacol 1988; 95:339.
113. Hempel J, Harper K, Lindahl R. Inducible (class III) aldehyde dehydrogenase from rat hepatocellular carcinoma and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-treated liver: Distant relationship to the class 1 and 2 enzymes from mammalian liver cytosol/mitochondria. Biochemistry 1989; 28:1160.
114. Koivusalo M, Aarnio M, Baumann M, Rautoma P. NAD(P)-Linked aromatic aldehydes preferring cytoplasmic aldehyde dehydrogenases in the rat. Constitutive and inducible forms in liver, lung, stomach and intestinal mucosa. Prog Clin Biol Res 1989; 290:19.
115. Vasiliou V, Reuter SF, Kozak CA, Nebert DW. Mouse dioxin-inducible cytosolic aldehyde dehydrogenase-3: AHD4 cDNA sequence, genetic mapping, and differences in mRNA levels. Pharmacogenetics 1993; 3:281.

116. Takimoto K, Lindahl R, Dunn TJ, Pitot HC. Structure of the 5' flanking region of class 3 aldehyde dehydrogenase in the rat. Arch Biochem Biophys 1994; 312:539.
117. Forte-McRobbie CM, Pietruszko R. Purification and characterization of human liver "high Km" aldehyde dehydrogenase and its identification as glutamic-semialdehyde dehydrogenase. J Biol Chem 1986; 261:2154.
118. Agarwal DP, Eckey R, Hempel J, Goedde HW. Human liver high Km aldehyde dehydrogenase (ALDH4): Properties and structural relationship to the glutamic γ-semialdehyde dehydrogenase. Adv Exp Med Biol 1993; 328:191.
119. Yin S-J, Vagelopoulos N, Wang S-L, Jörnvall H. Structural features of stomach aldehyde dehydrogenase distinguish dimeric aldehyde dehydrogenase as a "variable" enzyme: "Variable" and "constant" enzymes within the alcohol and aldehyde dehydrogenase families. FEBS Lett 1991; 283:85.
120. Hsu LC, Chang W-C. Cloning and characterization of a new functional human aldehyde dehydrogenase gene. J Biol Chem 1991; 266:12257.
121. Rajagopalan KV, Fridovich I, Handler P. Hepatic aldehyde oxidase: I. Purification and properties. J Biol Chem 1962; 237:922.
122. Rajagopalan KV, Handler P. Hepatic aldehyde oxidase: III. The substrate binding site. J Biol Chem 1964; 239:2027.
123. Krenitsky TA, Neil SM, Elion GB, Hitchings GH. A comparison of the specificities of xanthine oxidase and aldehyde oxidase. Arch Biochem Biophys 1972; 150:585.
124. Hall WW, Krenitsky TA. Aldehyde oxidase from rabbit liver: Specificity toward purines and their analogs. Arch Biochem Biophys 1986; 251:36.
125. Shaw S, Jayatilleke E. The role of aldehyde oxidase in ethanol-induced hepatic lipid peroxidation in the rat. Biochem J 1990; 268:576.
126. Mira L, Maia L, Barreira L, Manso CF. Evidence for free radical generation due to NADH oxidation by aldehyde oxidase during ethanol metabolism. Arch Biochem Biophys 1995; 318:53.
127. Yang Z-N, Davis GJ, Hurley TD, Stone CL, Li T-K, Bosron WF. Catalytic efficiency of human alcohol dehydrogenases for retinol oxidation and retinal reduction. Alcohol Clin Exp Res 1994; 18:587.
128. Pocker Y, Li H. The catalytic specificity of liver alcohol dehydrogenase: vitamin A alcohol and vitamin A aldehyde activities. Adv Exp Med Biol 1993; 328:411.
129. Boleda MD, Saubi N, Farrés J, Parés X. Physiological substrates for rat alcohol dehydrogenase classes: aldehydes of lipid peroxidation, ω-hydroxyfatty acids, and retinoids. Arch Biochem Biophys 1993; 307:85.
130. Klyosov AA. Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes. Biochemistry 1996; 35:4457.
131. Yoshida A, Hsu LC, Yanagawa Y. Biological role of human cytosolic aldehyde dehydrogenase 1: Hormonal response, retinal oxidation and implication in testicular feminization. Adv Exp Med Biol 1993; 328:37.
132. Mårdh G, Luehr CA, Vallee BL. Human class I alcohol dehydrogenase catalize the oxidation of glycols in the metabolism of norepinephrine. Proc Natl Acad Sci USA 1985; 82:4979.
133. Gotoh Y, Sumimoto H, Minakami S. Formation of 20-oxoleukotriene B_4 by an alcohol dehydrogenase isolated from human neutrophils. Biochem Biophys Acta 1990; 1043:52.
134. Okuda A, Okuda K. Physiological function and kinetic mechanism of human liver alcohol dehydrogenase as 5β-cholestane-3α,7α,12α,26-tetrol dehydrogenase. J Biol Chem 1983; 258:2899.
135. Consalvi V, Mårdh G, Vallee BL. Human alcohol dehydrogenases and serotonin metabolism. Biochem Biophys Res Commun 1986; 139:1009.
136. Mårdh G, Vallee BL. Human class I alcohol dehydrogenase catalyze the interconversion of alcohols and aldehydes in the metabolism of dopamine. Biochemistry 1986; 25:7279.
137. Tank AW, Weiner H. Ethanol-induced alteration of dopamine metabolism in rat liver. Biochem Pharmacol 1979; 28:3139.
138. Ambroziak W, Pietruszko R. Metabolic role of aldehyde dehydrogenase. Adv Exp Med Biol 1993; 328:5.
139. Holmes RS. Alcohol dehydrogenases: A family of isozymes with differential functions. Alcohol Alcohol 1994; Suppl 2:127.
140. Napoli JL. Retinoic acid biosynthesis and metabolism. FASEB J 1996; 10:993.
141. Russell DW, Setchell KDR. Bile acid biosynthesis. Biochemistry 1992; 31:4737.
142. Makita K, Falck JR, Capdevila JH. Cytochrome P450, the arachidonic acid cascade, and hypertension: new vistas for an old enzyme system. FASEB J 1996; 10:1456.
143. Duester G, Shean Ml, McBride MS, Stewart MJ. Retinoic acid response element in the human alcohol dehydrogenase gene *ADH3*: Implications for regulation of retinoic acid synthesis. Mol Cell Biol 1991; 11:1638.
144. Ang HL, Deltour L, Žgombić-Knight, M. Wagner MA, Duester G. Expression patterns of class I and class IV alcohol dehydrogenase genes in developing epithelia suggest a role for alcohol dehydrogenase in local retinoic acid synthesis. Alcohol Clin Exp Res1996; 20:1050.
145. Crabb DW, Stewart MJ, Xiao Q. Hormonal and chemical influences on the expression of class 2 aldehyde dehydrogenases in rat H4IIEC3 and human HuH/ hepatoma cells. Alcohol Clin Exp Res 1995; 19:1414.
146. Mezey E, Potter JJ. Effects of thyroidectomy and triiodothyronine administration on rat liver alcohol dehydrogenase. Gastroenterology 1981; 80:566.
147. Crabb DW, Bosron WF, Li T-K. Role of the pituitary and neonatal androgenic imprinting in the hormonal regulation of liver alcohol dehydrogenase activity. Biochem Pharmacol 1986; 9:1527.
148. ŽgombićM, Duester G. DNA elements mediating retinoid and thyroid hormone regulation of alcohol dehydrogenase gene expression. Adv Exp Med Biol 1993; 328:571.
149. Korri U-M. The effect of glucocorticoids, beta-2-adrenoceptor agonist, theophylline and propranolol on the rate of ethanol elimination and blood acetate concentration in humans. Alcohol Alcohol 1990; 25:519.
150. Qulali M, Dipple KM, Crabb DW. Modulation of hepatic and renal alcohol dehydrogenase activity and mRNA by steroid hormones in vivo. Adv Exp Med Biol 1993; 328:581.
151. Majewski JL, Yang VW. The class I alcohol dehydrogenase gene is glucocorticoid-responsive in the rat hepatoma microcell hybrid cell line, 11-3. Alcohol Clin Exp Res 1995; 19:1430.
152. Winter LA, Stewart MJ, Shean ML, Dong Y, Poellinger L, Okret S, Gustafsson JÅ, Duester G. A hormone response element upstream from the human alcohol dehydrogenase gene ADH2

consists of three tandem glucocorticoid receptor binding sites. Gene 1990; 91:233.
153. Cortese JF, Majewski JL, Crabb DW, Edenberg HJ, Yang VW. Characterization of the 5'-flanking sequence of rat class I alcohol dehydrogenase gene. J Biol Chem 1994; 269:21898.
154. Pappas P, Vasiliou V, Karageorgou M, Stefanou P, Marselos M. Studies on the induction of rat class 3 aldehyde dehydrogenase. Adv Exp Med Biol 1995; 372:143.
155. Ceci JD, Lawther R, Duester G, Hatfield GW, Smith M, O'Malley MP, Felder MR. Androgen induction of alcohol dehydrogenase in mouse kidney. Studies with a cDNA probe confirmed by nucleotide sequence analysis. Gene 1986; 41:217.
156. Mårdh G, Falchuk KH, Auld DS, Vallee BL. Testosterone allosterically regulates ethanol oxidation by homo- and heterodimeric gamma-subunit-containing isozymes of human alcohol dehydrogenase. Proc Natl Acad Sci USA 1986; 83:2836.
157. Pereira FA, Rosenmann E, Nylen EG, Wrogemann K. Human cytosolic aldehyde dehydrogenase in androgen insensitivity syndrome. Adv Exp Med Biol 1993; 328:45.
158. Yanagawa Y, Chen JC, Hsu LC, Yoshida A. The transcriptional regulation of human aldehyde dehydrogenase I gene: The structural and functional analysis of the promoter. J Biol Chem 1995; 270:17521.
159. Karageorgou M, Papadimitriou C, Marselos M. Sexual differentiation in the induction of the class 3 aldehyde dehydrogenase. Adv Exp Med Biol 1993; 328:123.
160. Dembic Z, Sabolic I. Alcohol dehydrogenase activity in rat kidney cortex stimulated by oestradiol. Biochim Biophys Acta 1982; 714:331.
161. Qulali M, Ross RA, Crabb DW. Estradiol induces class I alcohol dehydrogenase activity and mRNA in kidney of female rats. Arch Biochem Biophys 1991; 288:406.
162. Crabb DW. Hormonal regulation of alcohol dehydrogenase activity. In: Crow KE, Batt RD, editors. Human Metabolism of Alcohol, Vol II: Regulation, Enzymology, and Metabolites of Ethanol. Boca Raton, FL: CRC Press; 1989. p. 31.
163. Lakshman MR, Chambers LL, Chirtel SJ, Ekarohita N. Roles of hormonal and nutritional factors in the regulation of rat liver alcohol dehydrogenase activity and ethanol elimination rate in vivo. Alcohol Clin Exp Res 1988; 12:407.
164. Potter JJ, MacDougald OA, Mezey E. Regulation of rat alcohol dehydrogenase by cyclic AMP in primary hepatocyte culture. Arch Biochem Biophys 1995; 321:329.
165. Potter JJ, Yang VW, Mezey E. Influence of growth hormone on the synthesis of rat liver alcohol dehydrogenase in primary hepatocyte culture. Arch Biochem Biophys 1989; 274:548.
166. Mezey E, Potter JJ, Litt MR, Rhodes D. Influence of epinephrine on alcohol dehydrogenase activity in rat hepatocyte culture. Biochem Pharmacol 1988; 37:2993.
167. Mezey E, Potter JJ, Rhodes DL. Effect of glucagon on alcohol dehydrogenase activity in rat hepatocyte culture. Gastroenterology 1986; 91:1271.
168. Messiha FS. Alcohol and aldehyde-dehydrogenase: modulation by biogenic amine metabolites, neuropeptides and psychoactive agents. Adv Exp Med Biol 1993; 328:591.
169. Vasiliou V, Kozak CA, Lindahl R, Nebert DW. Mouse microsomal class 3 aldehyde dehydrogenase: AHD3 cDNA sequence, inducibility by dioxin and clofibrate, and genetic mapping. DNA Cell Biol 1996; 15:235.
170. Marselos M, Strom SC, Michalopoulus G. Effect of phenobarbital and 3-methylcholanthrene on aldehyde dehydrogenase activity in cultures of HepG2 cells and normal human hepatocytes. Chem Biol Interact 1987; 62:75.
171. Dunn TJ, Koleske AJ, Lindahl R, Pitot HC. Phenobarbital-inducible aldehyde dehydrogenase in the rat. cDNA sequence and regulation of the mRNA by phenobarbital in responsive rats. J Biol Chem 1989; 264:13057.
172. Chang C, Yoshida A. Cloning and characterization of the gene encoding mouse mitochondrial aldehyde dehydrogenase. Gene 1994; 148:331.
173. Nebert DW. Proposed role of drug-metabolizing enzymes: regulation of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. Mol Endocrinol 1991; 5:1203.
174. Valentine JW. Late precambrian bilaterians: grades and clades. Proc Natl Acad Sci USA 1994; 91:6751.
175. Deltour L, Ang HL, Duester G. Ethanol inhibition of retinoic acid synthesis as a potential mechanism for fetal alcohol syndrome. FASEB J 1996; 10:1050.
176. Garro AJ, Lieber CS. Alcohol and cancer. Annu Rev Pharmacol Toxicol 1990; 30:219.
177. Sellers EM, Higgins GA, Sobell MB. 5-HT and alcohol abuse. Trends Pharmacol Sci 1992; 13:69.
178. Ondarza RN. Alteraciones bioquímicas producidas por el alcoholismo. In: Molina-Piñeiro V, Sánchez-Medal L, editors. El alcoholismo en México, Vol. I: Patología. 2nd edition. México: Fundación de Investigaciones Sociales; 1985. p.83.
179. Pérez-Tamayo R. Patología del alcoholismo. In: Molina-Piñeiro V, Sánchez-Medal L, editors. El alcoholismo en México, Vol. I: Patología. 2nd edition. México: Fundación de Investigaciones Sociales; 1985. p.49.

Antecedentes

Evolución y diversidad de las alcohol deshidrogenasas.

La elaboración y consumo de bebidas alcohólicas es una práctica humana muy antigua, común a casi todas las culturas del mundo y cuyo origen esta asociado, la mayor de las veces, al desarrollo de la agricultura (Loeb, 1943; Popham, 1978). De este modo, el descubrimiento por Batelli y Stern en 1910, de que los extractos de diversos tejidos animales son capaces de oxidar el etanol, y la subsecuente purificación de proteínas con actividad de alcohol y aldehído deshidrogenasa (Dixon y cols, 1937; Leloir y Muñoz, 1938; Lutwak-Mann, 1938), fue considerado como una propiedad normal inherente al metabolismo de los animales.

En condiciones normales, el etanol en mamíferos es un metabolito que se encuentra en bajas concentraciones dentro de las células, en el intervalo de 0.1 a 1.0 mM. Este etanol se origina principalmente de la actividad fermentativa de la flora intestinal, y en segundo lugar, de la producción endógena de algunos tejidos como por ejemplo, el hígado (McManus y cols., 1960; Krebs y Perkins, 1970; Baraona y cols., 1986). Sin embargo, estas bajas cantidades no parecen justificar la existencia, en los mamíferos, de un complejo sistema enzimático conformado por hasta siete isoenzimas diferentes, cuyo papel aparente es la oxidación del etanol (Riveros-Rosas y cols., 1997).

La actividad de alcohol deshidrogenasa (*ADH*) y la capacidad de oxidar etanol, están ampliamente distribuidas en numerosos phyla, mostrando una distribución universal (Jörnvall y cols., 1993). Esta actividad se encuentra en al menos 4 familias de proteínas de diferente origen, de las cuales la de las alcohol deshidrogenasas de cadena media (que incluyen a la clásica *ADH* de mamíferos), es la que presenta el mayor número de isoenzimas y la de más amplia distribución (Jörnvall y cols., 1984, 1987). Esta familia de enzimas deriva de una forma ancestral (alcohol deshidrogenasa de clase III o formaldehído deshidrogenasa dependiente de glutatión), la cual presenta una distribución universal (Uotila y Koivusalo, 1989). De hecho, no se ha descrito hasta ahora, algún vertebrado que carezca de este grupo de enzimas.

La evolución de este grupo de enzimas en vertebrados se desarrolló en forma paralela a la evolución de los tetrápodos en el medio terrestre (Sun y Plapp, 1992). De esta manera, en peces por ejemplo, se encuentra la *ADH* de clase III típica y otra enzima con propiedades intermedias entre las clases I y III; en contraste, en anfibios ya existen las clases I y III bien

definidas, y en mamíferos se encuentran ya 6 clases diferentes de *ADHs* con casi una decena de isoenzimas (Danielsson y cols., 1994; Arnon y cols., 1995).

En este punto, es difícil entender la razón por la cuál los mamíferos presentan un sistema enzimático tan complejo para la oxidación del etanol, con diferentes isoenzimas expresadas en diferentes tejidos, siendo que éste es un metabolito que se produce en cantidades mínimas dentro de los vertebrados y que no es necesario para el metabolismo normal de estos organismos, razón por la cual se ha especulado mucho sobre el posible papel fisiológico de este sistema de enzimas en mamíferos.

De esta manera, se han elaborado diversas propuestas para explicar el papel fisiológico de las *ADHs* en vertebrados. Inicialmente, Krebs sugirió que esta actividad es necesaria para metabolizar la producción endógena de etanol, formado principalmente por la actividad de la flora intestinal, aunque más tarde se reconoció su participación en el metabolismo de los retinoides, en la deshidrogenación de esteroides y la síntesis de ácidos biliares (Waller y cols., 1965), e inclusive en la ω-oxidación de los ácidos grasos (Björken, 1972). Además, por su amplio espectro de posibles sustratos, alcoholes de cadena larga, ya sean primarios, secundarios, ramificados o cíclicos (Pietruszko, 1979), las *ADHs* se han relacionado con funciones de detoxificación de xenobióticos (Nebert, 1991). Todo este amplio espectro de funciones ha llevado a considerar al sistema de las alcohol deshidrogenasas de cadena media, como un sistema multifuncional que participa en numerosas actividades necesarias para el desarrollo normal de estos organismos (Riveros-Rosas et al., 1997). Es importante señalar que las ADHs de vertebrados muestran una alta identidad con otras oxidorreductasas, como por ejemplo, las sorbitol deshidrogenasas, las enoil reductasas, las quinona oxidorreductasas o las treonina deshidrogenasas. Todas ellas conforman lo que se denomina actualmente como la superfamilia de las deshidrogenasas/reductasas de cadena media (MDR).

A la fecha, se han identificado más de 500 proteínas como miembros de las MDR. Algunos trabajos recientes enfatizan la gran diversidad de proteínas que conforman las MDR; sin embargo, los análisis más recientes sobre la evolución de las MDR se publicaron entre 1992 y 1993, utilizando menos de 50 secuencias (Sun y Plapp, 1992; Yokoyama y Harry, 1993). En 2002, mientras este trabajo se encontraba en revisión, Hans Jornvall y colaboradores publicaron un trabajo en el que analizaron todas las secuencias pertenecientes a las MDR de 5 genomas eucariontes completos, incluyendo el de la bacteria *Escherichia coli* (Nordling y cols., 2002). Sus resultados, utilizando métodos filogenéticos, les permitió identificar 8 familias de proteínas distintas. Sin embargo, por el pequeño número de proteínas utilizadas en su análisis, no es posible identificar las subfamilias de proteínas existentes, amén de quedar abierta la posibilidad de descubrir la existencia de otras familias de proteínas, al utilizar el conjunto completo de proteínas pertenecientes a las MDR.

Por esta razón, en este trabajo se analizan la evolución y la taxonomía de las MDR, utilizando el conjunto completo de las proteínas pertenecientes a las MDR, con la idea de completar el trabajo iniciado por Hans Jornvall y colaboradores, identificando las subfamilias de proteínas presentes en cada una de las familias de las MDR. Para ello fue necesario establecer criterios para identificar grupos ortólogos, y se propuso un sistema de clasificación de proteínas dentro de las MDR, que puede ser aplicable a otras superfamilias de proteínas.

Material y métodos

Obtención y depuración de secuencias:

Se realizó una busqueda exhaustiva de las secuencias de aminoácidos de las proteínas correspondientes a: alcohol deshidrogenasas dependientes de Zn, sorbitol, manitol y treonina deshidrogenasas, cinamil ADH, quinona oxidorreductasas y enoil reductasas, en las bases de datos públicas SWISS-PROT + TrEMBL (Bairoch y Apweiler, 2000), y en la base de datos no redundante de secuencias de proteínas del GenBank, en el National Center for Biotechnology Information (NCBI) (Benson y cols., 2000). El acceso a las bases de datos del NCBI se efectuó a través del sistema integrado de recuperación de datos ENTREZ (Benson y cols., 2000). De esta busqueda, se seleccionaron secuencias de proteínas representativas que se encuentran listadas en la base de datos SWISS-PROT + TrEMBL (Bairoch y Apweiler, 2000), como miembros de las MDR. Posteriormente, se procedió a buscar mediante BLASTP (Altschul y cols, 1997), proteínas homólogas a cada una de las secuencias seleccionadas, con el fin de identificar nuevos miembros de las MDR. El programa BLASTP se utilizó empleando la penalización sugerida por defecto y la matriz de sustitución BLOSUM62. Siempre que una nueva secuencia fue identificada ($p<0.00001$), la búsqueda se repetía con el fin de identificar sus homólogos más cercanos. El procedimiento se repitió iterativamente hasta que no se identificaron nuevos miembros de las MDR.

Posteriormente, se identificaron y se excluyeron del análisis todos los registros duplicados, conservándose un total de 656 secuencias no redundantes. De este total se excluyeron también: a) secuencias alélicas; b) secuencias con menos de 75 aminoácidos, que por su pequeño tamaño suelen alinearse mal; c) isoenzimas con 100% de identidad; d) múltiples secuencias que corresponden a genes ortólogos identificados en diferentes especies pertenecientes a un mismo género, y que fueron consideradas como redundantes para el análisis filogenético; y e) fragmentos duplicados, esto es, fragmentos correspondientes al extremo amino y carboxilo terminal de una misma proteína, y que en las bases de datos se reportan como proteínas independientes (ver resultados).

Adicionalmente, se efectuó también una búsqueda exhaustiva (database mining) de las publicaciones de todos los autores que reportaron secuencias de aminoácidos pertenecientes a la superfamilia de las MDR. Esta búsqueda permitió incrementar notablemente la proporción de proteínas cuya función está determinada experimentalmente, en comparación con las registradas originalmente en las bases de datos, incluyendo bases de datos cuidadosamente curadas, como es el caso del SWISS-PROT. El problema de la falta de información acerca de la caracterización

experimental de las proteínas en las bases de datos se da por el hecho de que los autores deben registrar en estas bases de datos las nuevas secuencias de proteínas, antes de enviar a publicación sus resultados. Este procedimiento es impecable, porque asegura el libre acceso a la información. Sin embargo, con gran frecuencia los autores, después de publicar sus resultados, no incluyen en el registro la cita de la publicación en donde se caracteriza la proteína, por lo que el registro se conserva, pero sin información experimental sobre la misma.

Alineamientos:

El alineamiento progresivo múltiple de las secuencias de proteínas se realizó con el programa CLUSTAL-X (Thompson y cols, 1997) y se corrigió de acuerdo con los resultados de los alineamientos generados por BLASTP. Los dendrogramas de las secuencias de proteínas se calcularon utilizando el programa Clustal_X y se visualizaron con el programa Treeview (Page, 1996).

Análisis filogenético:

El análisis filogenético se realizó con el programa MEGA 2.1 (Kumar y cols, 2001), utilizando tanto métodos de máxima parsimonia (MP), como los métodos basados en distancias: UPGMA[1], vecino más próximo (NJ) y mínima evolución (ME). El método de Poisson se utilizó para corregir distancias y las interrupciones de la secuencia (gaps) se trataron como deleciones pareadas. Los límites de confiabilidad del análisis se calcularon por el método de bootstrap, utilizando 1000 repeticiones.

Vale la pena comentar que los métodos y herramientas de cómputo aquí incluidos son de uso corriente en nuestro laboratorio, y esta misma metodología se ha empleado en colaboración con otros grupos de trabajo, como el del Dr. Diego González Halphen (Atteia y cols., 2003), y el del Dr. Jesús Aguirre (Lara-Ortíz y cols, 2003), ambos del Instituto de Fisiología Celular de la UNAM.

[1] del ingles, unweighted pair-group method using arithmetic averages.

Resultados

A continuación se presenta la publicación principal de este proyecto de tesis doctoral, que analiza la diversidad, taxonomía y evolución de las proteínas que pertenecen a la superfamilia de las alcohol deshidrogenasas de cadena media, o dependientes de Zn, y actualmente denominadas como deshidrogenasas/reductasas de cadena media (MDR).

Al final de la sección de resultados se reproducen las figuras a color incluidas en este artículo (por cortesía del editor), y en la sección de anexos, se incluye la correspondencia mantenida con los árbitros del European Journal of Biochemistry. En la edición electrónica de ésta tesis doctoral, se incluyen las dos tablas con material suplementario que acompañan a este artículo.

Publicación 2.

Héctor Riveros-Rosas, Adriana Julián-Sánchez, Rafael Villalobos-Molina, Juan Pablo Pardo y Enrique Piña. (2003). Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. Eur. J. Biochem. 270: 3309-3334.

Eur. J. Biochem. **270**, 3309–3334 (2003) © FEBS 2003 doi:10.1046/j.1432-1033.2003.03704.x

Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily

Héctor Riveros-Rosas[1], Adriana Julián-Sánchez[1], Rafael Villalobos-Molina[2], Juan Pablo Pardo[1] and Enrique Piña[1]

[1]*Depto. Bioquímica, Fac. Medicina, UNAM, Cd. Universitaria, México D.F., México;* [2]*Depto. Farmacobiología, CINVESTAV-Sede Sur, México D.F., México*

A comprehensive, structural and functional, in silico analysis of the medium-chain dehydrogenase/reductase (MDR) superfamily, including 583 proteins, was carried out by use of extensive database mining and the BLASTP program in an iterative manner to identify all known members of the superfamily. Based on phylogenetic, sequence, and functional similarities, the protein members of the MDR superfamily were classified into three different taxonomic categories: (a) subfamilies, consisting of a closed group containing a set of ideally orthologous proteins that perform the same function; (b) families, each comprising a cluster of monophyletic subfamilies that possess significant sequence identity among them and might share or not common substrates or mechanisms of reaction; and (c) macrofamilies, each comprising a cluster of monophyletic protein families with protein members from the three domains of life, which includes at least one subfamily member that displays activity related to a very ancient metabolic pathway. In this context, a superfamily is a group of homologous protein families (and/or macrofamilies) with monophyletic origin that shares at least a barely detectable sequence similarity, but showing the same 3D fold.

The MDR superfamily encloses three macrofamilies, with eight families and 49 subfamilies. These subfamilies exhibit great functional diversity including noncatalytic members with different subcellular, phylogenetic, and species distributions. This results from constant enzymogenesis and proteinogenesis within each kingdom, and highlights the huge plasticity that MDR superfamily members possess. Thus, through evolution a great number of taxa-specific new functions were acquired by MDRs. The generation of new functions fulfilled by proteins, can be considered as the essence of protein evolution. The mechanisms of protein evolution inside MDR are not constrained to conserve substrate specificity and/or chemistry of catalysis. In consequence, MDR functional diversity is more complex than sequence diversity.

MDR is a very ancient protein superfamily that existed in the last universal common ancestor. It had at least two (and probably three) different ancestral activities related to formaldehyde metabolism and alcoholic fermentation. Eukaryotic members of this superfamily are more related to bacterial than to archaeal members; horizontal gene transfer among the domains of life appears to be a rare event in modern organisms.

Keywords: protein taxonomy; protein evolution; medium-chain alcohol dehydrogenase; enoyl reductase; formaldehyde dehydrogenase.

Correspondence to H. Riveros-Rosas, Depto. Bioquímica, Fac. Medicina, UNAM, Apdo. Postal 70–159, Cd. Universitaria, México, 04510, D.F., México. Fax: + 52 55 5616 2419, Tel.: + 52 55 5622 0829, E-mail: hriveros@servidor.unam.mx

Abbreviations: AADH, allyl alcohol dehydrogenase; ACR, acyl-CoA reductase; ADH, alcohol dehydrogenase; AL, alginate lyase; ARP, auxin-regulated protein; AST, membrane traffic protein; BCHC, 2-desacetyl-2-hydroxyethyl bacteriochlorophyllide-a dehydrogenase; BDH, 2,3-butanediol dehydrogenase; BDOR, bi-domain oxidoreductase; BRP, bacteriocin-related protein; CADH, cinnamyl alcohol dehydrogenase; CCAR, crotonyl-CoA reductase; COG, cluster of orthologous groups of proteins; DHSO, sorbitol dehydrogenase; DINAP, dinoflagellate nuclear-associated protein; DI-QOR, dark induced-quinone oxidoreductase; ELI3, elicitor-inducible defense-related proteins; ER, enoyl reductase; FADH, formaldehyde dehydrogenase; FAS, fatty acid synthase; FDEH, 5-exo-hydroxycamphor dehydrogenase; GATD, galactitol 1-phosphate dehydrogenase; GDH, glucose dehydrogenase; GSH, glutathione; HNL, hydroxynitrile lyase; LTD, leukotriene B_4 12-dehydrogenase; MDR, medium-chain dehydrogenases/reductases; MP, maximum parsimony; MRF, mitochondrial respiratory function protein; MSH, mycothiol; MTD, mannitol-1-phosphate dehydrogenase; NCBI, National Center for Biotechnology Information; NJ, neighbour-joining; NRBP, nuclear receptor binding protein; PDH, polyol dehydrogenase; pER, probable enoyl reductase; PGR, 15-oxoprostaglandin 13-reductase; PIG3, animal P53-induced gen. 3; PKS, polyketide synthase; PKS-IAP, polyketide synthase-independent associated protein; QOR, quinone oxidoreductase; QORL-1, quinone oxidoreductase-like 1; SORE, L-sorbose-1-phosphate dehydrogenase; SSP, sensing starvation protein; TDH, threonine dehydrogenase; TED2, quinone oxidoreductase involved in tracheary element differentiation in plants; UPGMA, unweighted pair-group method using arithmetic averages; Y-ADH, yeast alcohol dehydrogenase.

Note: a web site is available at http://laguna·fmedic.unam.mx/%7Eadh/

(Received 2 April 2003, revised 27 May 2003, accepted 5 June 2003)

NAD(P)-dependent alcohol dehydrogenase (ADH) activity is widely distributed in nature and is carried out by three main superfamilies of enzymes that arose independently throughout evolution [1]. Their amino acid identity is 20% or less and they exhibit different structures and reaction mechanisms. The first superfamily corresponds to the Fe-dependent ADHs and makes up the smallest and least studied family of alcohol dehydrogenases [2–4]. The second group includes the short-chain dehydrogenase/reductase superfamily; this large family of enzymes do not require a metallic ion as cofactor [5,6]. The third superfamily is composed of zinc-dependent ADHs, and is named preferentially medium-chain dehydrogenases/reductases (MDRs) [7,8]. These enzymes usually require zinc atom(s) as cofactor and the family includes the classical horse liver ADH. In addition to these three NAD(P)-dependent ADH families, other minor families of ADH exist, which use different cofactors such as FAD, and pyrroquinoline quinone, among others; however, the distribution of these minor families is limited to some bacterial groups [1].

To date, nearly 1000 protein sequences have been identified as MDR superfamily members [8–10]. Identification of new members of the MDR superfamily is performed with high statistical significance using tools such as BLASTP [11] or FASTA [12,13]. However, efforts to assign proteins to families and/or subfamilies within the MDR superfamily have not been equally successful. Public proteins databases use different criteria to classify proteins, and therefore, several inconsistencies in the identification of protein subfamilies and families have been observed. Recently, Nordling *et al.* [14], based on analysis of five complete eukaryotic genomes, and *Escherichia coli*, constructed an evolutionary tree of the MDR in which at least eight families can be distinguished: dimeric ADHs in animals and plants; tetrameric ADHs in fungi (Y-ADHs), polyol dehydrogenases (PDHs), quinone oxidoreductases (QORs), cinnamyl alcohol dehydrogenases (CADHs), leukotriene B4 dehydrogenases (LTDs), enoyl reductases (ERs), and nuclear receptor binding protein (NRBPs). ERs and NRBPs were originally described [14] as acyl-CoA reductases (ACRs) and mitochondrial respiratory function proteins (MRFs), respectively; the Results section discusses why the names of these enzymes are described differently here.

Because the MDR protein families proposed by Nordling *et al.* [14] were identified considering only a few genomes, it is possible that other protein families of the MDR may be identified if complete sets of their protein sequences are used. Furthermore, a larger set of MDRs will allow us to make a more detailed taxonomic analysis. Therefore, in this report we analysed MDR taxonomy on the basis of the entire set of currently known MDR members, and completed the work initiated by Nordling *et al.* with identification of further protein subfamilies that comprise each protein family within the MDR superfamily. To contribute to validation of the eight protein families previously identified, we grouped protein sequences employing a different method from that used by Nordling *et al.* [14]. Indeed, the limited number of protein sequences employed by Nordling *et al.* [14], precluded them from identifying protein subfamilies.

Finally, we analysed evolution of the MDR superfamily and identified some putative selective forces that directed their enzymogenesis. This analysis is valuable as a paradigm of protein evolution and provides information to understand previously defined concepts such as protein family, subfamily, and superfamily, and their relationships to several protein classification efforts. Furthermore, recruitment of selected members of this superfamily may offer clues about the evolution of some metabolic pathways, and show the evolutionary history of different organisms: for example, ER was recruited from MDR and incorporated into the multifunctional enzyme fatty acid synthase from animals (not fungi or plants); additionally, the capacity for retinoic acid synthesis, a powerful regulator of genetic expression active only in vertebrates, evolved in parallel to evolution of animal ADHs; and animal ADHs are involved in the synthetic or catabolic route of paramount modulators such as epinephrine, serotonin, and dopamine [15].

Materials and methods

Extensive database searches for zinc-dependent ADH, sorbitol dehydrogenase, threonine dehydrogenase, CADH, mannitol dehydrogenase, ER, and QOR were performed. Protein sequence data were taken from SWISS-PROT + TrEMBL protein databases [16] and the GenBank nonredundant protein sequence database at the National Center for Biotechnology Information (NCBI) [17]. Access to NCBI databases was achieved by means of the integrated database retrieval system ENTREZ [17]. Gapped BLASTP program with default gap penalties and BLOSUM62 substitution matrix was employed [11]. Thus, based on selected protein sequences that belong to each of the subfamilies that compose the MDR superfamily, a search for homologous sequences was performed through BLASTP for each selected sequence to identify new members of MDRs not yet recognized. Whenever a new sequence was identified ($P < 0.00001$), the BLASTP search was repeated, seeking closer relative sequences. The procedure was repeated iteratively until no new members of MDRs were recognized.

Progressive multiple protein sequence alignment was calculated with the CLUSTAL_X package [18] using secondary structure-based penalties and corrected according to results of gapped BLASTP [11]. Dendrograms were calculated using CLUSTAL_X [18] and displayed with TREEVIEW [19]. Phylogenetic analyses were performed with MEGA2 software [20], using both maximum parsimony (MP) and distance-based methods [UPGMA, and neighbour-joining (NJ)], with the Poisson correction distance method, and gaps treated by pairwise deletion. Confidence limits of branch points were estimated by 1000 bootstrap replications.

The procedure to define protein subfamilies and families is explained with detail in the Results section.

Results

A total of 656 nonredundant sequences (allelic forms excluded) were identified as members of MDR superfamily. Of this total, 73 sequences were excluded from final analysis for one of the following reasons: (a) sequences with less than 75 amino acids; (b) isozymes with 100% identity; (c) multiple sequences corresponding to orthologous genes identified in several species from the same genera, because they were considered redundant for the phylogenetic analysis; and

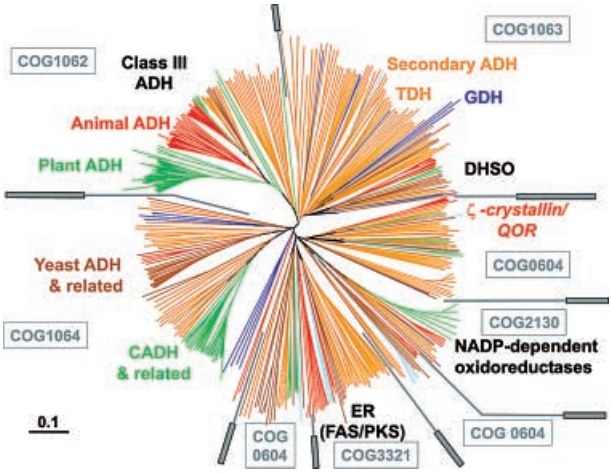


Fig. 1. Unrooted tree constructed with identified 583 nonredundant protein sequences that belong to the MDR superfamily. Each sequence is coloured as follows: red, animals; green, plants; brown, fungi; light blue, protista; orange, bacteria; dark blue, archaea. Protein sequences were ascribed to different subfamilies, as indicated in the SWISSPROT database [16]. As a guide, the protein families considered by COG Database [30–32] are displayed (Table 1); grey pins mark the boundaries of clusters of orthologous groups of proteins (COGs).They do not correspond to the protein families and subfamilies proposed in this work.

(d) duplicity in information, for example, two fragments of proteins in *Streptomyces coelicolor* (CAB53403 and CAB55521), were identified as the N- and C-terminus, respectively, of the same protein (kindly confirmed by S. Bentley, Sanger Institute, Hinxton, Cambridge, UK; personal communication). Thus, 583 nonredundant protein sequences were considered for phylogenetic analysis; of these, 21 proteins belong to archaea, 234 to bacteria, 11 to protista, 62 to fungi, 148 to plants, and 107 to animals.

The 583 sequences permitted construction of the unrooted tree shown in Fig. 1. Protein sequences were ascribed to different subfamilies, as indicated in the SWISSPROT database. Conserved groups with high degree of identity can be identified easily (e.g. class III ADH, plant ADHs, animal ADHs), as well as poorly conserved subfamilies, such as sorbitol dehydrogenase, ER, or QOR. Conserved protein subfamilies are identified because distances between their members are short, and appear as a group of branches that join among themselves far from the centre of the tree. In comparison, poorly conserved subfamilies with low identity among themselves, resemble groups of long branches that depart close to the centre of the tree. However, the latter, more than being an inherent property of these subfamilies, might be due to problems concerning particular aspects with regard to reliability of database information, because a significant fraction of functional annotations in databases is dubious or even incorrect [21,22]. This problem arises because there are many noncharacterized sequences. An especially illustrative example is the case of the QOR/ζ-crystallin subfamily, in which many protein sequences are assumed to be QOR only by sequence similarities with the well-characterized animal QOR/ζ-crystallins. Thus, other noncharacterized distantly related sequences are assumed to be also QOR only by similarity to the second group of QOR-related sequences.

In summary, GenBank reports might be produced before characterization is completed and/or published; usually, authors do not update the original GenBank report after publication. Therefore, many proteins would already have been characterized, but this information is not quoted in the GenBank and other protein databases. Thus, to record reliable functional identification for most proteins, an extensive search for published papers by authors who made contributions to GenBank for each of the MDRs was carried out. This functional identification plus statistically significant degree of similarities calculated with BLAST (*E*-value), allowed us to identify many additional small subfamilies as members of MDR superfamily. *E*-value represents the number of alignments with an equivalent or greater score, that would be expected to occur purely by chance [23].

Table 1 lists the main protein families that are found with the MDR superfamily, as stated by several public protein databases. Several inconsistencies in the nomenclature for protein subfamilies, families and superfamilies are observed: for example, Pfam [24] does not attempt to identify families or subfamilies in the MDR superfamily; PROSITE [25] uses motifs to identify two protein families in the MDR superfamily; PIR [26,27] uses distance-based criteria to identify 119 families in MDR; CATH [28,29] uses structural data to identify six superfamilies in MDR; COG [30–32] uses phylogenetic criteria to identify six families; and SYSTERS uses a non-distance-based method to identify 80 families. This discrepancy is due to the different criteria used for defining each of these terms.

To clarify this, we have defined a protein subfamily as a set of homologous (ideally orthologous) protein sequences that (a) performs the same function and (b) forms a closed group in which identity, similarity, and statistical significance between any two members of the closed group are higher than to any other protein sequence outside the subfamily, i.e. clusters of proteins with BLAST reciprocal best hits. Often, members of protein subfamilies share more than 30% sequence identity, and *E*-value of approximately 10–30 or less. It should be mentioned that all-vs.-all BLAST-based searches have recently been used to find orthologs [33–36], and that these methods bypass multiple alignments and construction of phylogenetic trees, which can be slow and error-prone steps in classical ortholog detection [37].

The previously mentioned definition of subfamily is nearly identical to the approach employed in the SYSTERS database to define protein families or clusters of protein sequences [38–40], but with the additional condition that all sequences in a cluster must (ideally) share the same function. This functional criterion is necessary because true orthologous proteins must perform the same function; if this last condition is not true, then the proteins are paralogous. In contrast, paralogous proteins do not necessarily possess different functions, in that by definition, two proteins are said to be paralogous if they are derived from a duplication event, but orthologous if they are derived from a speciation event [41–44]. Therefore, initially a duplication event will produce two proteins possessing identical properties, and only after evolution might they acquire different functions.

Table 1. Protein families/subfamilies within medium-chain dehydrogenase/reductase superfamily (MDR) as it is indicated on several public databases.

Database	Protein families/subfamilies considered within MDR
Pfam [24]	PF00107 adh_zinc (consider only one superfamily)
PROSITE [25]	PDOC00058 Zinc-containing alcohol dehydrogenases Considers two patterns or signatures: PS00059 ADH-ZINC PS01162 QOR_ZETA_CRYSTAL.
SCOP [147]	Family: alcohol dehydrogenase-like, N-terminal domain Family: alcohol/glucose dehydrogenases, C-terminal domain Considers two similar families and both contain the same five domains: Sorbitol dehydrogenase/secondary ADH/Glucose dehydrogenase/Alcohol dehydrogenase/Quinone oxidoreductase
InterPro [148]	IPR002085 Zinc-containing alcohol dehydrogenase superfamily. Considers two families: IPR002364 Quinone oxidoreductase/zeta-crystallin IPR002328 Zinc-containing alcohol dehydrogenase Considers one subfamily: IPR004627 L-threonine 3-dehydrogenase
CATH [28,29]	Considers six homologous superfamilies based on structural data. Two of them are domains contained inside the other four multidomain superfamilies Homologous superfamily 3.40.50.720 NAD(P)-binding Rossmann-like domain Homologous superfamily 3.90.180.10 Medium-chain alcohol dehydrogenases, catalytic domain Homologous superfamily 5.1.120.1 Oxidoreductase (NAD(A)-CHOH(D)); include animal ADH, class III ADH Homologous superfamily 5.1.2796.1 Oxidoreductase; include secondary ADH Homologous superfamily 5.1.1670.1 Oxidoreductase: include quinone oxidoreductase Homologous superfamily 7.1.147.10 Oxidoreductase; include sorbitol dehydrogenase
PIR-PSD (MIPS/IESA) [26,27]	SF000091 alcohol dehydrogenase superfamily. Considers 119 protein families, the main protein families are: Fam000150 (94 sequences: includes animal ADH, plant ADH, class III ADH) Fam000152 (18 sequences: includes fungi ADH) Fam007438 (31 sequences: includes CADH) Considers two motifs: PCM00059 zinc-containing ADH PCM0162 Quinone oxidoreductase/zeta crystalline
COG [30–32]	Considers six families or Clusters of Orthologous Groups of proteins (COGs): COG 1063: Threonine dehydrogenase and related Zinc-dependent dehydrogenases COG 1062: Zinc-dependent alcohol dehydrogenases, class III (and related) COG 1064: Zinc-dependent alcohol dehydrogenases (include CADH and fungi ADH) COG 0604: NADPH: quinone oxidoreductase and related Zinc-dependent oxidoreductases COG 3321: Polyketide synthase (PKS) modules and related proteins (enoyl reductase from PKS and FAS) COG 2130: Putative NADP-dependent oxidoreductases AADH/LHD (and related)
SYSTERS [38–40]	adh_zinc Include 80 clusters (families), organized into superfamilies; the main superfamilies are: Superfamily of cluster O60787: includes six aditional clusters with sequences from animal ADH, plant ADH, class III ADH (equivalent to COG1062) Superfamily of cluster N60795; includes 13 aditional clusters with sequences from CADH, fungi ADH, DHSO, TDH, secondary ADH among others (equivalent to COG1063 plus COG1064) Superfamily of cluster N60499: includes five aditional clusters with sequences from QOR/ζ-crystallin and related (equivalent to COG0604) Superfamily of cluster O59495 and O59531: includes other nonrelated clusters (equivalent to COG3321).

This explanation is obligatory because some papers provide inexact definitions [45–47].

This non-distance-based method allows us to sort MDR sequences into nonoverlapping clusters (subfamilies), in which the granularity of this clustering is determined by data and not by a user-supplied data-dependent cut-off [38]. Identification of closed groups of protein sequences, or perfect clusters (in agreement with SYSTERS nomenclature), is advantageous over distance-based clustering methods because it is not necessary to set an arbitrary identity cutoff value to define a subfamily (or families in the SYSTERS database), and permits identification of both highly and poorly conserved groups of orthologous proteins. Furthermore, Krause & Vignron [39] showed that this

method is highly conservative, as the probability of obtaining a false positive is extremely low, i.e. we almost never observe sequences that do not belong to a cluster being included.

On the other hand, this subfamily definition fits with the widely used nomenclature proposed by Persson *et al.* [7] for the MDR superfamily. Thus, only closed groups with at least one characterized protein were listed as true protein subfamilies in this work. This criterion excluded some minor clusters without characterized proteins, or protein sequences located in the twilight zone, which can not be assigned with certainty to a protein subfamily. Furthermore, there is always the possibility that best match in a database hit is solely a well-conserved paralog [22] that in reality belongs to a related, but different, protein subfamily.

As a consequence of application of these criteria, subfamilies identified in this work are equivalent to a carefully crafted, manual-curated version from clusters of proteins proposed in the SYSTERS database. Figure 2 shows an unrooted tree constructed with all the MDR protein sequences identified in bacteria and archaea, with recognized protein subfamilies indicated. Figure 3 shows an equivalent unrooted tree constructed with protein sequences identified in eukaryota. In both trees, the main subfamilies of the MDR superfamily are easily visualized. Comparison of Figs 2 and 3 clearly shows that in addition to the well-characterized protein subfamilies that exist simultaneously in several phylogenetic lineages, there are additional

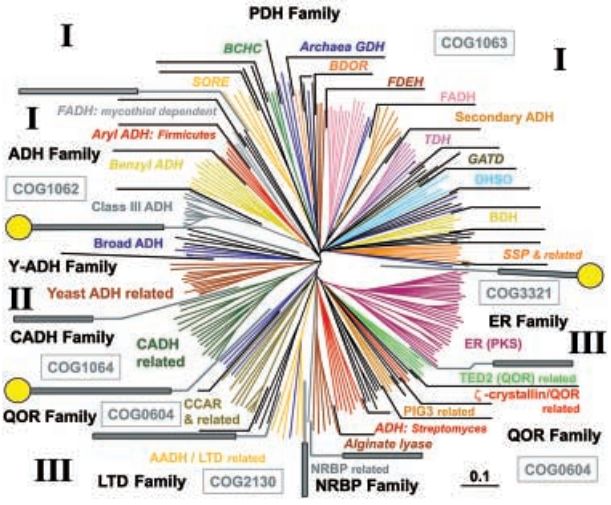


Fig. 2. Unrooted tree constructed with identified protein sequences that belong to MDR in bacteria and archaea. Subfamilies were identified based on statistical identity and similarity calculated with BLAST. Only subfamilies with at least one functionally characterized protein received a name. The three main clusters of subfamilies (macrofamilies) are indicated with roman numerals and the name of each family and subfamily is abbreviated. Grey pins mark the boundaries of protein families; yellow-capped pins mark the boundaries of protein macrofamilies. COGs are also indicated in boxes. The complete names of the protein subfamilies are indicated in Tables 3–8, according to the protein family to which they belong. Subfamilies present only in one kingdom are indicated in italics: bacteria or archaea; normal type indicates subfamilies present in two or more kingdoms. All archaea sequences are coloured in blue, for clarity, bacterial sequences are coloured in the font colour selected to name each subfamily.

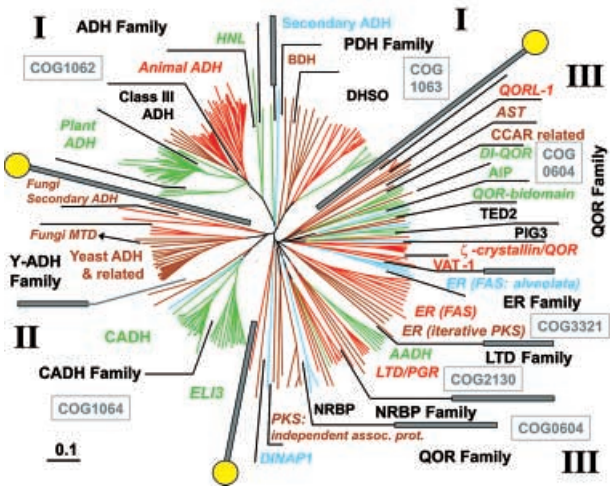


Fig. 3. Unrooted tree constructed with 328 protein sequences that belong to MDR in eukaryota. Each sequence is coloured as follows: red, animals; green, plants; brown, fungi; light blue, protista. The three main clusters of subfamilies (macrofamilies) are indicated with roman numerals and the name of each family and subfamily is abbreviated. Grey pins mark the boundaries of protein families; yellow-capped pins mark the boundaries of protein macrofamilies. COGs are also indicated in boxes. The complete names of the protein subfamilies are indicated in Tables 3–8, according to the protein family to which they belong. Subfamilies with restricted distribution are shown in italics, with subfamilies with broad distribution shown in normal font.

subfamilies associated with only one phylogenetic lineage, suggesting a more recent evolutionary origin.

It can also be observed that several protein subfamilies are formed by clusters of related subfamilies (Figs 2 and 3). According to the previous proposal for protein subfamilies, we define a protein family as a set of protein subfamilies in which identity and/or similarity of proteins in the family is higher among them than when compared with other proteins belonging to a different family. Therefore, a family is composed of a closed group of subfamilies in which the closest relative of one subfamily is always another subfamily member from the same family. However, although protein subfamily definition used in this work comprises (ideally) a natural unit (orthologous proteins with the same function), the protein family is not a straightforward concept, as it is necessary to set author cutoff criteria to identify it. In fact, with tools such as BLASTP, identification of the protein superfamily to which one new protein belongs is easy and accurate. An additional functional analysis of the new protein permits recognition of the orthologous group (subfamily) to which this protein belongs. Nonetheless, at present there are no universal criteria to classify proteins into intermediate categories located between subfamily and superfamily. Indeed, a universally accepted protein family definition, does not exist; thus, different authors use different concepts with a different emphasis, e.g. homology in sequence, structure, and/or function.

Therefore, using BLAST to compare *E*-values and identity/similarity values among different protein subfamilies, we can identify several clusters of protein subfamilies in the MDR superfamily. In this way, at the highest level of

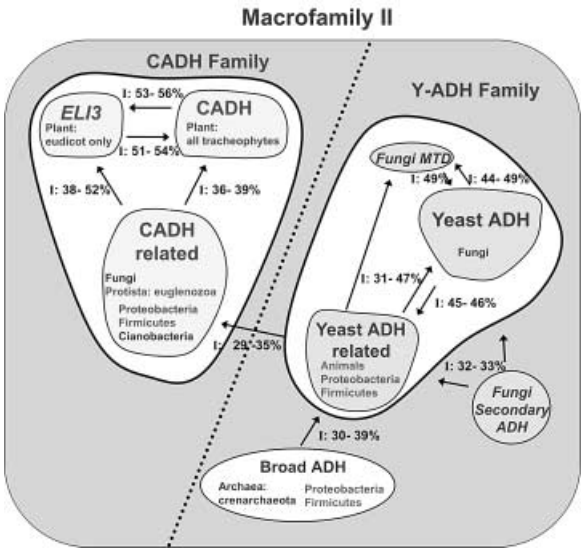


Fig. 4. Schematic diagram showing the main relationships between different protein subfamily members of macrofamily II (COG1064), listed in Table 4. The arrows point toward subfamilies with the highest statistical significance (*E*-value); not all possible relationships are displayed. Two clusters of closely related subfamilies (CADH family, and Y-ADH family) are seen, but all are interrelated among themselves, forming a closed group. The relationships between subfamilies are not necessarily symmetric; nonsymmetric relationships can be observed in amino acid sequences [39]. Inside each subfamily, taxa, where found, are indicated. Identity (I), indicated as percentage is showed for illustrative purpose only. The dotted line separates the CADH and Y-ADH families.

integration, we herein identify three great clusters or macrofamilies in the MDR superfamily (see Figs 2 and 3). At lower levels of integration, we identify six clusters of orthologous groups of proteins (COGs), that comprise the MDR superfamily (according to the COG database proposed by Koonin & Tatusov (see Table 1) [30–32]), or the eight protein families recently proposed by Nordling *et al.* [14]. To illustrate the criteria used to identify clusters of protein subfamilies, Fig. 4 illustrates schematically the main relationships among the different subfamily members that comprise macrofamily II in Figs 3 and 4 (this big cluster is equivalent to COG1064, and comprises the Y-ADH and CADH families from Nordling *et al.* [14]). Similar data were obtained with the other protein subfamilies (not shown).

Additionally, the proposed taxonomic categories (subfamilies, families, and macrofamilies) were validated by bootstrap analysis with conventional phylogenetic methods, using both distance-based methods (neighbour-joining and UPGMA), and character-based methods (maximum parsimony). To perform this phylogenetic analysis, only subsets of the MDR superfamily were utilized (the complete set demands excessive resources of computing power). Initial subsets employed for phylogenetic analysis included protein sequences that belong to only one kingdom (archaea, bacteria, animals, plants, or fungi). These kingdom-specific subsets were used to validate by bootstrap analysis the proposed taxonomic categories: macrofamilies and families. Later, subsets of proteins that belong to each of the proposed three macrofamilies, or eight families, were used to validate by bootstrap analyses, the proposed 49 protein subfamilies. Figure 5 shows a phylogenetic tree constructed with protein sequences belonging to macrofamily II of MDR superfamily. The additional phylogenetic trees constructed with protein sequences pertaining to macrofamilies I and III, and to each of the kingdoms to which belong the MDR proteins (archaea, bacteria, fungi, animals or plants) are not shown.

Table 2 shows a comparison of the proposed protein families that comprise MDR superfamily, according to COG database, the Nordling *et al.* paper [14], and the three macrofamilies or main clusters identified in this work. It is clear that information in addition to sequence data is needed to define the true protein families comprising the MDR superfamily. Consensus agreements among protein taxonomists must be reached before setting up intermediate categories between ideally true orthologous clusters (subfamilies in this paper) and superfamilies. Sequence data alone are not enough to set up true protein families with a real biological sense. It is important to point out that the intermediate categories proposed in COG database, the Nordling *et al.* paper [14], and in this work create a congruent pattern despite the different criteria used to define them in each study.

Tables 3–8 present lists of subfamilies in the eight families of the MDRs, and their distribution into the different kingdoms, with a brief summary for each subfamily (a complete list with all protein sequences and consulted references was included as supplementary material and can be requested from the publisher or the authors).

Interestingly, archaea protein sequences appear to be concentrated in only two families (macrofamily I: PDH family, COG1063, and macrofamily II: Y-ADH family, COG1064), suggesting that these two families, with a universal distribution, are the probable ancestral protein families in the MDR superfamily. However, in macrofamily III, a small uncharacterized cluster related to crotonyl-CoA reductase (CCAR) subfamily also possesses archaea members, also suggesting an ancient group.

In bacterial phyla, the taxa with sequences most related to eukaryota are firmicutes (Gram-positive) and proteobacteria (γ subdivision), see Tables 3–8. However, this proximity could simply be due to the fact that these bacterial clades possess the greatest number of completely sequenced genomes. Table 9 shows the number of identified genes that belong to the MDR in completely sequenced species. There is great variability with respect to total number of genes identified in each organism, even whitin the same taxonomic category, as well as variability with respect to the number of genes identified in MDR superfamily.

Macrofamily I: PDH family (COG1063): DHSO, TDH, and related subfamilies

This family was formerly denominated by Nordling *et al.* [14] as PDH (polyol dehydrogenase) family; however, after including bacteria and archaea members, it is clear that less than half of their subfamily members possess an activity related to polyol metabolism. The PDH family is

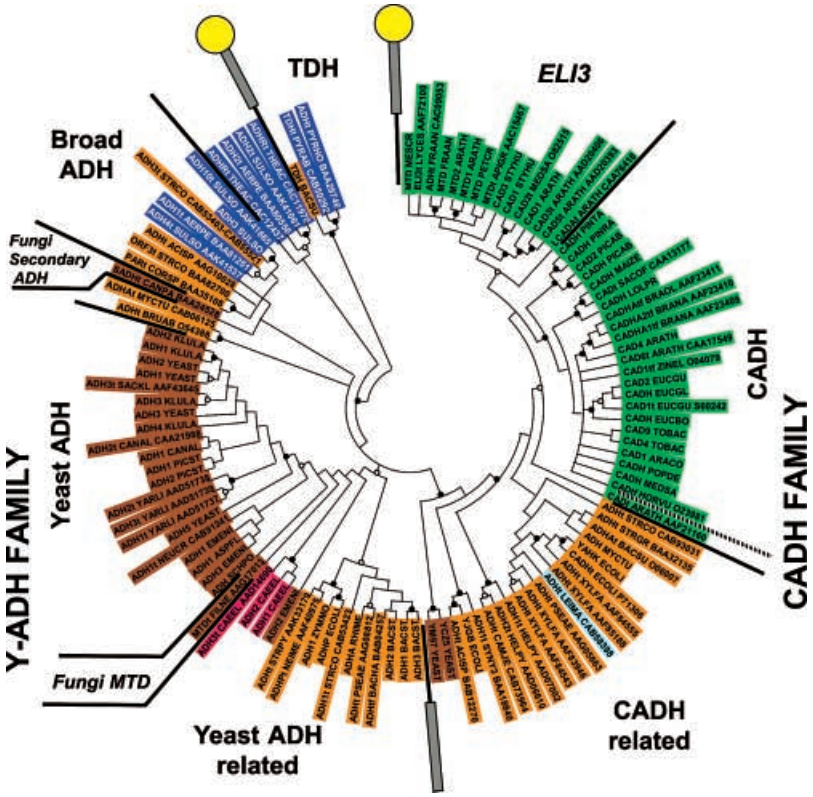


Fig. 5. Phylogenetic tree constructed with the protein sequences that belong to macrofamily II within MDR superfamily. Shown is the consensus UPGMA tree which was constructed with the computer software MEGA v. 2.1 [20], using the 50% majority-rule. Sequence names are shaded as follows: red, animals; green, plants; brown, fungi; light blue, protista; orange, bacteria; dark blue, archaea. The circles indicate those nodes supperted in > 70% (open), > 80% (grey) or > 90% (closed) of 1000 random bootstrap replicates of all NJ, UPGMA and MP. Resultant trees were rooted with threonine dehydrogenase protein sequences (macrofamily I). Grey pins mark the boundaries of protein families (Y-ADH family and CADH family); yellow-capped pins mark the boundaries of protein macrofamilies. Sequence names are indicated with a SwissProt-like identifier (Gene_organism), followed by the accession number assignated by the database (GenBank, PIR, TrEMBL, etc.; only sequence names reported by the nonredundant SWISSPROT database were used directly).

composed of 12 subfamilies (Table 3). Their characterized members contain zinc, show dehydrogenase or reductase activities, bind NAD(H), except secondary ADHs that use NADP(H), and are cytosolic proteins, with the exception of the bi-domain oxidoreductase subfamily (BDOR), which appears to be represented by transmembrane proteins. They are organized as homotetramers or homodimers that are involved in several metabolic roles, but only two correspond to anabolic activities: BDOR, involved in exopolysaccharide biosynthesis, and 2-desacetyl-2-hydroxyethyl bacteriochlorophyllide-a dehydrogenase subfamily (BCHC), in bacteriochlorophyll-a biosynthesis in proteobacteria. Remaining enzymes in PDH family show catabolic activities related either to aryl/alkyl metabolism (FDEH, secondary ADH, and BDH), formaldehyde metabolism (FADH, formaldehyde dismutase), carbohydrate catabolism (DHSO, SORE, GATD, and archaea GDH), and threonine and derivative compound catabolism (TDH and SSP). Five subfamilies have polyphyletic distribution and simultaneously exist in at least two domains (eukaryota and bacteria, or archaea and bacteria). Of these five subfamilies, four include tetrameric proteins and three are present in archaea.

Macrofamily I: ADH family (COG1062): class III ADH and related subfamilies

This family includes classical ADHs from animals and plants. ADH family comprises seven subfamilies absent in archaea (Table 4). Only one subfamily has a broad distribution: class III ADH, which is present in animals, plants, fungi and bacteria (cyanobacteria and proteobacteria). Proteins belonging to these subfamilies are cytoplasmic, although class III ADHs in animals are also nuclear [48]. They contain zinc, bind NAD(H), except animal ADH8 from *Rana perezi* that uses NADP(H) [49,50], and show dehydrogenase or reductase activities, with the exception of hydroxynitrile lyase (HNL) in plants. They are homodimers and only mycothiol-dependent formaldehyde dehydrogenase is atypically reported as a homotrimer [51–53].

With the exception of HNL, involved in cyanogenesis in plants, all enzymatic activities fulfilled by the MDR subfamilies in the ADH family are catabolic activities related either to aryl/alkyl metabolism (benzyl ADH, firmicute aryl/alkyl ADH), or formaldehyde metabolism (class III ADH, mycothiol-dependent FADH). It is likely

that the function of plant and animal ADHs, although typically associated with ethanol metabolism, is more complex, in that these comprise an intricate system with a broad diversity of enzymatic forms. The animal ADH subfamily, in addition to ethanol oxidation, participates in oxidation or reduction of diverse endogenous substrates involved in retinoic acid and bile acid synthesis, norepinephrine, leukotriene, serotonin, and dopamine catabolism, or in detoxification of cytotoxic products of lipoperoxidation such as 4-hydroxynonenal (reviewed in [15]). Thus, it is difficult to accept that this complex enzymatic system with its broad diversity of enzymatic forms and substrates (up to eight ADH classes in vertebrates) [49,54] was produced in the course of vertebrate evolution with the sole purpose of oxidizing ethanol, an exogenous metabolite found in minimal quantities under regular conditions: in fact, there are several endogenous substrates metabolized by this complex of enzymatic forms with an efficiency at least one thousand times higher than that of ethanol [15]. A similar history probably occurred in plants. Plant ADHs comprise a complex subfamily with numerous enzymatic forms expressed in a developmental and tissue-specific manner; it was suggested recently that these participate in flooding tolerance, anther development, fruit ripening, disease resistance, and stress response (reviewed in [55]).

Macrofamily II: CADH family (COG1064): ELI3, CADH and related subfamilies

The CADH family comprises two subfamilies; only one shows a broad distribution (Table 5). Their members are oxidoreductases and use zinc. All are dimeric proteins and bind NADP(H), except ELI3 in celery. Enzymes in the

Table 2. Comparison of the protein families included within MDR superfamily according to COG database, Nordling *et al.* [14], and the three macrofamilies or main clusters of protein subfamilies identified in this work. The distribution of MDR subfamilies inside each protein family is indicated, as well as their distribution into eukaryota, bacteria, and archaea domain.

Intermediate Taxonomic Categories inside MDR Superfamily			Subfamilies (true orthologous clusters)		
			Eukaryota	Bacteria	Archaea
Macrofamily I	COG 1063	PDH Family	--	TDH	
			DHSO		--
			BDH		--
			Secondary ADH		--
			--	--	Archaea GDH
			--	BCHC	--
			--	SORE	--
			--	GATD	--
			--	SSP & related	--
			--	FDEH	--
			--	BDOR	--
			--	FADH	
	COG 1062	ADH Family	--	Benzyl ADH	--
			--	Aryl/Alkyl ADH	--
			HNL	--	--
			--	FADH: mycothiol-dependent	
			Class III ADH		--
			Animal ADH	--	--
			Plant ADH	--	--
Macrofamily II	COG 1064	CADH Family	ELI3	--	--
			CADH and related		--
		Y-ADH Family	Fungi MTD	--	--
			Fungi secondary ADH	--	--
			Yeast ADH and related		--
			--	Broad ADH	

Table 2. (Continued).

Intermediate Taxonomic Categories inside MDR Superfamily			Subfamilies (true orthologous clusters)		
			Eukaryota	Bacteria	Archaea
Macrofamily III	COG 1064	QOR Family	ζ-crystallin/QOR		--
			PIG3 and related		--
			TED2 and related		--
			Bifunctional QOR and related	--	--
			VAT1	--	--
			--	pER	--
			PKS-IAP	--	--
			QORL1	--	--
			DINAP	--	--
			ARP	--	--
			DI-QOR	--	--
			DI-QOR/ARP related		
			AST	--	--
			--	AL	--
			--	BRP	--
			CCAR and related		
		NRBP Family [1]	NRBP and related		--
	COG 2130	LTD Family	LTD/PGR	--	--
			AADH	--	--
			LTD/AADH related [2]		
	COG 3321	ER Family [3]	ER (FAS)	--	--
			ER (iterative PKS)	--	--
			ER - FAS (alveolata)	--	--
			--	ER (modular PKS)	--

[1] This family was formerly denominated by Nordling *et al.* [14] as the mitochondrial respiratory function proteins (MRF) family. [2] This subfamily is probably comprised by two or more paralogous related groups. [3] Nordling *et al.* [14] named inappropriately this family as acyl-CoA reductase (ACR).

CADH subfamily perform anabolic functions and participate in biosynthesis of cinnamyl alcohols, the monomeric precursors of lignin in plants. In bacteria, in which lignin is absent, CADH-related proteins participate in biosynthesis of the lipids composing the bacterial cell envelope; in fungi, they could participate in ligninolysis and fusel alcohol synthesis pathways [56,57].

Elicitor-inducible defense-related proteins (ELI3) are present only in eudicot plants, and show different, but related, defense activities: CADH, benzyl alcohol dehydrogenase, or mannitol dehydrogenase. ELI3 expression is elicited by fungal pathogens [58], wounds [59], salicylic acid [60], and leaf senescence [61]. In celery, there is down-regulation by sugars or salt stress [62–64].

Macrofamily II: Y-ADH family (COG1064): yeast ADH, and related subfamilies

The Y-ADH family comprises four subfamilies; two show broad distribution (Table 5). Their members are oxidoreductases and use zinc. This family contains tetrameric proteins that use NAD(H) and have catabolic functions, involved mainly in metabolism of ethanol or short-chain alcohols (typical yeast ADH, broad ADH, and fungal-secondary ADH), or metabolism of mannitol (fungal MTD). The most ancient subfamily is probably the broad ADH; it is present in archaea and bacteria, and its members exhibit broad substrate specificity.

Table 3. Main subfamilies that comprise the PDH family of MDR (COG1063) and their occurrence in eukaryota, archaea and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria [d]
DHSO (sorbitol dehydrogenase) [a]		
Homotetramer	Animals	Firmicutes
NAD^+/NADH	Plants	Proteobacteria (γ subdivision)
1 Zn^{2+}/subunit	Fungi	Proteobacteria (α subdivision)
Cytoplasm		
BDH (2,3-butanediol dehydrogenase)		
Homodimer	Fungi	Firmicutes
NAD^+/NADH		Proteobacteria (γ subdivision)
2 Zn^{2+}/subunit (putative)		Proteobacteria (β subdivision)
Cytoplasm		
TDH (threonine dehydrogenase)		
Homotetramer	–	Euryarchaeota
1 Zn^{2+}/subunit (2 Zn^{2+}/subunit?)		Firmicutes
NAD^+/NADH		Proteobacteria (γ subdivision)
Cytoplasm		Proteobacteria (α subdivision)
		Thermus/Deinococcus group
BCHC (2-desacetyl-2-hydroxyethyl bacteriochlorophyllide *a* dehydrogenase)		
Unpurified protein, characterized by genetic analysis only	–	Proteobacteria (α subdivision)
		Proteobacteria (β subdivision)
SORE (L-sorbose-1-phosphate reductase)		
Homodimer	–	Proteobacteria (γ subdivision)
Use both NAD^+/NADH and $NADP^+$/NADPH		
Requires an activating divalent metal (Zn^{2+})		
Secondary ADH		
Homotetramer	Protista:	Firmicutes
NADP/NADPH	Entamobidae	Proteobacteria (γ subdivision)
1 Zn^{2+}/subunit (only catalytic)		Proteobacteria (β subdivision)
Cytoplasm		
GATD (galactitol 1-phosphate dehydrogenase)		
Homodimer	–	Proteobacteria (γ subdivision)
NAD^+/NADH		
Require divalent cations for activity and stability		
Cytoplasm		
SSP and related (sensing starvation protein)		
Unpurified protein		Firmicutes
Catabolic enzyme that suppress induction of rpoS expression at starvation or stationary phase		Proteobacteria (γ subdivision)
		Thermotogales
FDEH (5-exo-hydroxycamphor dehydrogenase)		
Homodimer		
NAD/NADH	–	Proteobacteria (γ subdivision)
2 Zn^{2+} (putative)		Thermotogales
BDOR (bi-domain oxidoreductase) [b]		
Unpurified protein		Firmicutes
Probable transmembrane protein		Proteobacteria (β subdivision)
		Proteobacteria (γ subdivision)
Archaea GDH (glucose dehydrogenase)		
Homotetramer (*Sulfolobus*: crenarchaeota)		Euryarchaeota
Homodimer (*Haloferax*: euryarchaeota)		Crenarchaeota
Both NAD^+/NADH and $NADP^+$/NADPH		
2 Zn^{2+}/subunit		

Table 3. (Continued).

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria [d]
FADH (formaldehyde dehydrogenase-independent of cofactor-/formaldehyde dismutase)		
Homotetramer	–	Euryarchaeota
NAD^+/NADH		Firmicutes
2 Zn^{2+}/subunit		Proteobacteria (γ subdivision)
		Proteobacteria (β subdivision)
		Thermus/Deinococcus group

[a] The members of this subfamily receive the official name of L-iditol 2-dehydrogenase, and possess alternative names as glucitol dehydrogenase, xylitol dehydrogenase or polyol dehydrogenase, in addition to sorbitol dehydrogenase. This subfamily catalyzes the reversible oxidation of D-sorbitol and other polyalcohols, like xylitol and L-iditol, to the corresponding keto-sugars [149–152]. [b] N-terminus is similar to diverse DHSO; C-terminus is probably an NAD(P)H oxidoreductase, which belongs to the GFO_IDH_MocA family. It is related to synthesis of exopolysaccharides. [c] Two enzymes have been purified, and characterized: formaldehyde dehydrogenase from *Pseudomonas putida*, and formaldehyde dismutase also from *Pseudomonas putida*. However, recently Oppenheimer *et al*., demonstrate that formaldehyde dehydrogenase from *P. putida* is a functional alcohol dehydrogenase that conducts the efficient dismutation of wide range of aldehydes (including formaldehyde), where NADH production represents a pH-dependent burst. Thus, both enzymes can be considerated as formaldehyde dismutases. [d] For bacteria and archaea, only sequences that can be unambiguously assigned to one subfamily are considered in the table. References are included on Table S2 of supplementary material.

Macrofamily III: QOR family (COG0604): QOR and related subfamilies

Members of this family lack zinc and use mainly NADP(H) as cofactor. It is the most complex and divergent family, with 16 subfamilies (Table 6). Twelve subfamilies are found in only one taxon, suggesting intensive and recent enzymogenesis. In functional and structural terms, this is a highly divergent family and their members, in addition to oxidoreductase activity, act as lyases, nuclear-associated proteins, membrane traffic proteins (that participate in subcellular protein distribution), and integral membrane proteins with ATPase activity and calcium-binding capacity. This family is nearly absent in archaea; only *Halobacterium* sp. and *Sulfolobus sulfataricus* have proteins related to CCARs. It is likely that CCAR and related proteins are the most ancient subfamily of macrofamily III, because they have the widest distribution (archaea, bacteria, and eukaryota) and because it is the only subfamily with a physiologic role related to primary metabolic pathways.

Macrofamily III: NRBP family (COG0604): NRBP1 subfamily and related

This small family comprises only nuclear receptor binding protein 1 (NRBP1) and related subfamily (Table 6). It has broad distribution, and is present in animals, plants, fungi and bacteria. Their members are homodimers, with both nuclear and cytosolic location. This family was formerly designated by Nordling *et al*. [14] as the mitochondrial respiratory function proteins (MRF) family; however, this name is unfortunate in that members of this family probably do not have enzymatic activity. In animals these proteins are nuclear receptor co-operators; in the cytosol, in presence of the appropriate ligand, they interact with several nuclear hormone receptors, such as peroxisome proliferator-activated receptor α, thyroid hormone receptor, retinoic acid receptor, retinoid-X receptor, and hepatocyte nuclear factor-4 [65]. Later, NRBP1-activated nuclear receptor complex is translocated to the nucleus by a piggyback mechanism, where they act as transcription factors. Although fungi and bacteria lack nuclear receptors, in *Saccharomyces cerevisiae*, MRF1_YEAST (P38071), a single-stranded DNA-binding protein, has acquired the activity of a transcription factor [66,67]. Indeed, it is a transcriptional regulatory protein of certain genes whose products are necessary for the functional assembly of mitochondrial respiratory proteins. In bacteria, uncharacterized related proteins are reported in *Corynebacterium glutamicum* and *Xanthomonas campestris*. Thus, it is likely that in the course of evolution, NRBP1 acquired a new function to work with nuclear receptors. This family appears to be evolved from members of QOR family (COG 0604).

Macrofamily III: LTD family (COG2130): LTD/AADH and related subfamilies

This is a small family with only three subfamilies (Table 7). Members lack zinc and have a preference for NADP(H) over NAD(H). Two subfamilies are found in only one taxon: leukotriene B_4 12-hydroxydehydrogenase (LTD)/15-oxoprostaglandin 13-reductase (PGR), found in animals and allyl alcohol dehydrogenase (AADH), found in plants. Both subfamilies clearly have their origin in an uncharacterized protein subfamily (LTD/AADH related) with broad distribution. This protein family is closely related to QOR Family COG0604 (Figs 2 and 3).

Macrofamily III: ER family (COG3321): enoyl reductases

This family contains four related subfamilies comprising multifunctional polypeptides that enclose a MDR domain with ER activity (Table 8). ER domains in MDR enzymes use NADP(H) and lack zinc. These subfamilies show limited distribution and are involved in biosynthesis of fatty acids and polyketides. Nordling *et al*. [14] inappropriately named this family as acyl-CoA reductase (ACR). As they

Table 4. Main subfamilies that comprise the ADH family of MDR (COG1062) and their occurrence in eukaryota, archaea and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
Aryl/Alkyl ADH: Firmicutes [a]		
Unpurified protein; characterized by genetic analysis only	–	Firmicutes
Benzyl ADH [b]		
Homodimer (*Pseudomonas putida*)	–	Proteobacteria (γ subdivision)
Homotetramer (*Acinetobacter calcoaceticus*)		Proteobacteria (α subdivision)
2 Zn^{2+}/subunit		Firmicutes
NAD^+/NADH		
Cytoplasm		
HNL (Hydroxynitrile lyase: acetone cyanohydrin lyase)		
Homodimer	Plants (derived from	–
(not an oxidoreductase)	plant-/class III ADH)	
2 Zn^{2+}/subunit		
Cytoplasm		
FADH: mycothiol-dependent (formaldehyde dehydrogenase dependent on mycothiol)		
Homotrimer	–	Firmicutes
NAD^+/NADH		
2 Zn^{2+}/subunit		
Cytoplasm		
Class III ADH (formaldehyde dehydrogenase dependent on glutathione)		
Homodimer (Eukaryota; Cyanobacteria	Animals	Cyanobacteria
and Proteobacteria)	Fungi	Proteobacteria (γ subdivision)
Homotetramer (*Paracoccus*: Proteobacteria α)	Plants	Proteobacteria (β subdivision)
NAD^+/NADH		Proteobacteria (α subdivision)
2 Zn^{2+}/subunit		
Cytoplasm (all) and nucleus (animals)		
Animal ADH [c]		
Homodimer [d]	Animals	–
NAD^+/NADH [e]	(derived from animal class III)	
2 Zn^{2+}/subunit		
Cytoplasm		
Plant ADH		
Dimer	Plants	–
NAD^+/NADH	(derived from plant class III)	
2 Zn^{2+}/subunit		
Cytoplasm		

[a] This belongs to a highly conserved gene cluster encoding haloalkane catabolism on the plasmid Prtl1. [b] This shows affinity for a wide range of (substituted) aromatic alcohols, but are not capable of oxidizing aliphatic alcohols. [c] This subfamily comprises eight different classes involved besides ethanol metabolism, on the synthesis and catabolism of several endogenous metabolites that regulate growth, metabolism, differentiation, and neuroendocrine functions [15,50,54]. [d] Some animal ADH are also heterodimers (e.g., isozymes from human class I ADH). [e] Only class VIII ADH from *Rana perezi* uses NADP(H) rather than NAD(H) [49,50]. See final note (d) in Table 3.

identified correctly the enoyl-acyl carrier protein (ACP) reductase domain contained in multifunctional fatty acid synthase from animals, or enoyl-ACP reductase domain from iterative polyketide synthase in fungi, the generic name enoyl reductase is preferable. The enzyme ACR is absent in fatty acid synthase; this latter multidomain enzyme uses ACP as carrier for intermediates, not coenzyme A. ACR is usually a membrane-bound enzyme involved in the biosynthesis of fatty alcohols and waxes, and it is clearly a different enzyme that does not belong to the MDR superfamily [68,69].

Animal fatty acid synthases are closer to fungal iterative polyketide synthases than to any other fatty acid synthases from fungi, plant, or bacteria. The latter kingdoms possess one ER that does not belong to the MDRs. As can be seen in Figs 2 and 3, this protein family is also closely related to QOR Family (COG0604).

Discussion

We will focus our discussion on five topics: criteria used to define a protein family; mechanisms of evolution in MDR;

Table 5. Main subfamilies that comprise the CADH family and Y-ADH family of MDR (COG1064) and their occurrence in eukaryota, archaea, and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
CADH FAMILY		
CADH and related (cinnamyl alcohol dehydrogenase) [a]		
Homodimer	Plants (tracheophytes)	Firmicutes
$NADP^+$/NADPH	Fungi	Proteobacteria (γ subdivision)
	Protista: Euglenozoa	Proteobacteria (ε subdivision)
		Cyanobacteria
ELI3 (elicitor-inducible defense-related proteins) [b]		
Homodimer	Plants	–
Monomer (celery)	(Eudicots: derived from CADH)	
$NADP^+$/NADPH		
NAD^+/NADH (in celery)		
Y-ADH FAMILY		
Yeast ADH and related		
Homotetramer	Fungi	Proteobacteria (γ subdivision)
NAD^+/NADH	Animals	Proteobacteria (α subdivision)
2 Zn^{2+}/subunit		Proteobacteria (β subdivision)
Cytoplasm and mitochondria		Firmicutes
Fungi MTD (mannitol-1-phosphate dehydrogenase)		
Homotetramer	Fungi	–
NAD^+/NADH	(derived from yeast ADH)	
2 Zn^{2+}/subunit		
Cytosol		
Fungi secondary ADH		
Homotetramer	Fungi	–
NAD^+/NADH	(derived from yeast ADH)	
2 Zn^{2+}/subunit (putative)		
Cytosol		
Broad ADH (broad substrate specificity ADH) [c]		
Homotetramer	–	Crenarchaeota
NAD/NADH		Firmicutes
2 Zn^{2+}/subunit		Proteobacteria (γ subdivision)
Cytosol		

See final note (d) in Table 3. [a] Induced by several elicitors, such as pathogens, ozone, and wounding. [b] Proteins described with different activities: cinnamyl alcohol dehydrogenase, benzyl alcohol dehydrogenase, or mannitol dehydrogenase. Induced by fungal pathogens, wound, salicylic acid, and leaf senescence; shows a down-regulation by sugar or salt stress. [c] Shows broad substrate specificity; carbon source stimulated.

whether eukaryota inherited their enzymatic machinery mainly from bacteria; ancestral activities of MDR; and taxonomy within MDR superfamily.

Criteria used to define a protein family: sequence over functional similarities

Generally, the term protein family describes 'a group of homologous (frequently orthologous) enzymes that catalyse the same reaction (mechanism and substrate specificity)' [47]. However, in addition to their primary activities, enzymes often have other secondary activities with lower efficiency and different substrates and mechanism of reaction [70]. For example, horse ADH also exhibits aldehyde dismutase [71,72] and esterase activities [73]; yeast ADH additionally shows methylformate synthase activity [74]. Therefore, it is clear that through evolution several proteins acquired, with only a few point mutations, activities that differed from the primary activity [46]. This implies the existence of several structurally related proteins with high identity or similarity, but different functional roles [75]. These proteins (closely related paralogous, but with a different mechanism of reaction and/or substrates) might even show higher similarity than the most distant phylogenetic derivatives in the same protein family (true orthologous) with the same activity, substrates, and mechanism of reaction. For example, identity and similarity between plant ADHs and class III ADHs from plants (paralogous proteins with different substrates) are higher than identity and similarity between class III ADHs from plant and bacteria; albeit both orthologous proteins have the same activity, substrates, and mechanism of reaction [indeed, identity between ADH1_MAIZE (P00333) and ADHX_MAIZE (P93629) (paralogous proteins) is 59%, but identity between ADHX_MAIZE (P93629) and FADH_PARDE (P45382) (orthologous proteins) is 55%]. Based on this type of data, it is clear that several proteins exhibit significant similarity (> 30–40% identity), but have

Table 6. Main subfamilies that comprise the QOR family and NRBP family of MDR (COG0604) and their occurrence in eukaryota, archaea and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
NRBP FAMILY		
NRBP1 (nuclear receptor binding protein/transcription factor) [a]		
Homodimer	Animals	Firmicutes
Transcription factor (yeast)	Plants	
Nuclear receptor co-operator (animals)	Fungi	
Nucleus (fungi and animals) and cytosol (animals)		
QOR FAMILY		
ζ-crystallin/QOR (quinone oxidoreductase) [b]		
Taxon-specific lens crystallin	Animals	Firmicutes
Homotetramer		
$NADP^+$/NADPH		
Lack Zn^{2+}		
PIG3 and related (animal P53 Induced Gen 3: putative quinone oxidoreductase) [c]		
Unpurified protein; characterized by genetic analysis only.	Animals	Firmicutes
Cytoplasm	Plants	Proteobacteria (α subdivision)
	Protozoa: Euglenozoa	
TED2 and related (quinone oxidoreductase involved in Tracheary Element Differentiation in plants)		
Homodimer (γ Proteobacteria: *E. coli*)	Plants	Proteobacteria (γ subdivision)
Both NAD^+/NADH and $NADP^+$/NADPH (*E. coli*)	Protozoa: Euglenozoa	Proteobacteria (α subdivision)
Lack Zn^{2+}	Fungi	Firmicutes
Cytoplasm		
Bifunctional QOR and related [d]		
Monomer (Euglenozoa)	Plants	–
$NADP^+$/NADPH	Protozoa: Euglenozoa	
Lack Zn^{2+}		
Cytoplasm		
VAT1 [e]		
Localized in the synaptic membranes, as an integral membrane protein	Animals	–
pER in actinomycetes (probable enoyl reductase in actinomycetes) [f]		
Unpurified protein; characterized by genetic analysis only.	–	Firmicutes
PKS-IAP (polyketide synthase-independent asociated proteins) [g]		
Unpurified protein; characterized by genetic analysis only.	Fungi	–
Heterodimers?		
QORL-1 (quinone oxidoreductase-like 1) [h]		
Unpurified proteins.	Animals	–
DINAP (dinoflagellate nuclear associated protein) [i]		
Unpurified protein	Protozoa:	–
Nucleus	Alveolata, dinophyceae	
ARP (auxin regulated protein) [j]		
Unpurified protein; characterized by genetic analysis only	Plants	–
DI-QOR (dark induced-quinone oxidoreductase) [k]		
Unpurified protein; characterized by genetic analysis only	Plants	–
DI-QOR/ARP related		
Unpurified protein; uncharacterized	Fungi	–
AL (alginate lyase)		
This protein is not an oxidoreductase	–	Proteobacteria (γ subdivision)
Does not require either NAD^+/NADH or $NADP^+$/NADPH.		
Cytosol		
AST (membrane traffic protein)		
Unpurified protein; characterized by genetic analysis only	Fungi	–
Plasma membrane-associated		

Table 6. (Continued).

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
BRP (bacteriocin-related proteins) [l]		
Unpurified proteins	–	Firmicutes
CCAR (crotonyl-CoA reductase) and related		
Homodimer	Fungi	Euryarchaeota
$NADP^+$/NADPH		Firmicutes
		Proteobacteria (γ subdivision)
		Proteobacteria (α subdivision)

See final note (d) in Table 3. [a] In animals, NRBP1 is translocated to the nucleus by a piggyback mechanism. In rat, it interacts with peroxisome proliferator-activated receptor α, PPARα; thyroid hormone receptor, TR; retinoic acid receptor, RAR; retinoid-X receptor, RXR, and hepatocyte nuclear factor-4, HNF-4. Fungi lack nuclear receptors; in yeast, it is a single-stranded DNA-binding protein that fulfills a role as transcription factor. [b] Several activities for ζ-crystallin/QOR have been reported, however, the relative importance of any remains an enigma. Nevertheless, all ζ-crystallin retain NADPH binding capacity as a common character. [c] PIG3 in humans seems to be a redox-related protein involved in the formation of reactive oxygen species in response to p53-induced apoptosis. [d] Bifunctional protein in plants; monofunctional protein in Euglenozoa. In plants, it is a defense protein whose synthesis is activated as response to pathogen-inoculation. In Euglenozoa, its functional role is not resolved. [e] VAT-1 forms a high-molecular-mass complex within the synaptic vesicle membrane, and is composed of three or four VAT-1 subunits, displays an ATPase activity, and binds calcium with low affinity. [f] Probable monofunctional enoyl reductase involved in biosynthesis of actinomycete aromatic polyketides in a multicomponent (type II) polyketide synthase complex. [g] Monofunctional enoyl reductase associated to iterative multidomain type I polyketide synthase. [h] Expressed mainly in heart, brain, and skeletal muscle, and moderately expressed in placenta, kidney, and pancreas. [i] Dinap1 protein is one of the quantitatively major nuclear proteins in the dinoflagellate *Crypthecodinium cohnii*. Although Dinap1 did not bind directly to DNA, it activated basal transcription activity. [j] Protein highly expressed during fruit-ripening, or induced in response to auxin treatment. [k] These proteins are expressed in plant roots, where light-induced a negative regulation. They are involved in biosynthesis of antimicrobial or allelopathic quinines. [l] They are included inside plasmids that contain a bacteriocin production region.

Table 7. Main subfamilies that comprise the LTD family of MDR (COG2130) and their occurrence in eukaryota, archaea and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
LTD (Leukotriene B_4 12-hydroxydehydrogenase)/PGR (15 oxoprostaglandin 13-reductase) [a]		
Monomer	Animals	–
Preference for $NADP^+$/NADPH over NAD^+/NADH		
Cytoplasm		
AADH (allyl alcohol dehydrogenase) [b]		
Homodimer	Plants	–
$NADP^+$/NADPH		
Cytoplasm (probably)		
LTD/AADH related [c]		
Uncharacterized proteins	Fungi	Euryarchaeota
	Animal?	Firmicutes
		Proteobacteria (γ subdivision)

See final note (d) in Table 3. [a] This subfamily in animals corresponds to proteins with two different activities, indicating that enzymes are capable of carrying out reduction of a double bond, as well as oxidation of a hydroxy group. [b] Enzymes efficient for dehydrogenation of secondary allylic alcohols and reduction of azodicarbonyl compounds and quinones. Induced by various oxidative-stress treatments. [c] Bacterial and archaea proteins show 40.2 ± 2.5% (SD, $n = 36$) average identity with animal LHD family, and a 39.6 ± 2.4% (SD, $n = 36$) with plant AADH family.

different functional roles. Therefore, sequence data alone cannot be used as sole criterium to define protein families, because without functional data, orthologous and paralogous groups cannot be accurately identified.

On the other hand, the protein function cannot be the main criterium used to define a protein family because one domain might have several catalytic activities. In fact, LTD subfamily shows two different and equally efficient catalytic activities: leukotriene B_4 12-hydroxydehydrogenase, which catalyses oxidation of a hydroxyl group, and 15-oxoprostaglandin 13-reductase, which carries out reduction of a double bond [76]. In contrast, there are several examples where the same function can be fulfilled by several nonrelated proteins with distinct domains, conforming analogous enzymes [75,77,78]. The MDR and the short-chain dehydrogenase/reductase (SDR) superfamilies contain several analogous enzymes. Thus, the SDR superfamily contains an analogous alcohol dehydrogenase found in

Table 8. Main subfamilies that comprise the ER family of MDR (COG3321) and their occurrence in eukaryota, archaea and bacteria.

Subfamily/main characteristics	Eukaryota	Archaea/Bacteria
Enoyl reductase (Fatty acid synthase -*FAS*-) [a]		
Homodimer	Animal	–
$NADP^+$/NADPH		
Cytoplasm		
Enoyl reductase (modular polyketide synthase -*PKS*-) [b]		
Homodimer		Firmicutes
$NADP^+$/NADPH		Proteobacteria (δ subdivision)
Cytoplasm		Proteobacteria (γ subdivision)
		Proteobacteria (α subdivision)
Enoyl reductase (iterative polyketide synthase -*PKS*-) [c]		
Heterodimer	Fungi (PKS)	–
$NADP^+$/NADPH (by similarity to modular *PKS* and *FAS*)		
Cytoplasm		
ER-FAS: alveolata (enoyl reductase from type I fatty acid synthase in alveolata) [d]		
Homodimer?	Protozoa: alveolata	–
$NADP^+$/NADPH (by similarity to modular *PKS* and *FAS*)		
Cytoplasm		

See final note (d) in Table 3. [a] This enoyl reductase domain belongs to a multifunctional polypeptide of approximately 2500 aa that contains seven enzymatic domains. [b] This enoyl reductase domain belongs to a multifunctional polypeptide with modular organization where each module designates a repeated unit whose functional domains resemble a single type I fatty acid synthase. [c] This enoyl reductase domain belongs to a multifunctional polypeptide whose functional domains resemble a single type I fatty acid synthase. In fungi, *PKS* is involved in mycotoxin biosynthesis. [d] This enoyl reductase domain belongs to a multifunctional polypeptide of 8243 aa that contains 21 enzymatic domains in *Cryptosporidium parvum*. Three ER domains are organized inside three modules, each containing a complete set of six enzymes for elongation of fatty acid C2-units (i.e., one ER/module).

Drosophila [79], a glucose dehydrogenase from *Bacillus* [80,81], an ER from bacteria and plants [82–84], a sorbitol dehydrogenase from *Klebsiella* [85], and a threonine dehydrogenase in animals [86]. These enzymes represent different protein structure solutions to the same activities observed in MDRs.

In summary, phylogenetic data can not be overlooked as a criterium for identification of a protein family. All families recognized inside the MDR superfamily are made up of clusters of phylogenetically related paralogous proteins, which may or may not conserve their original substrates or mechanisms of reaction. All paralogous proteins are generated by duplication events, and initially possess the same function; selective pressures and evolutive forces shape the functional role that duplicated proteins will perform. A change in the functional role of a protein is not necessarily related to a change in substrates or mechanism of reaction. Recruitment of a duplicated protein into a different metabolic pathway, a different physiological role, or even a change in the spatiotemporal pattern of expression, expressing a protein in novel tissues and/or developmental stages [87], could be a good evolutionary reason to conserve the duplicated protein, and result in a novel paralogous protein with a different functional role.

Therefore, we propose that the condition of performing the same function (with one, two, or more catalytic activities) must be assigned solely at a more specific (or restricted) taxonomic level, such as at the subfamily level (employed in this work). A protein family must be defined based mainly on sequence similarities, but in conjunction with other biological criteria different from function, such as phylogenetic data, since minor changes in amino acid sequence may induce changes of function.

Mechanisms of evolution in MDR superfamily

Enzymogenesis. Currently, two different evolutionary scenarios are envisioned for enzyme evolution [88]. New catalytic functions of enzymes can evolve by: (a) changing the chemistry of catalysis, while retaining the binding capacity for a common ligand (hypothesis initially proposed by Horowitz [89]) or (b) retaining the chemistry of catalysis while changing the substrate specificity. Interestingly, we found several enzymes of the MDR superfamily that conserved their chemistry of catalysis, but changed their substrate specificity, e.g. plant ADH and animal ADH subfamilies that evolved both from class III ADH subfamily; or secondary ADH from fungi and mannitol-1-phosphate dehydrogenase from fungi (Fungi MTD), that evolved both from yeast ADH subfamily. In contrast, we could not find two related enzymes of MDR superfamily that maintained their binding capacity for a common ligand, but with modification in their chemistry of catalysis. This possibility, described as retrograde evolution or substrate-driven evolution, suggests that metabolic pathways evolved in a backward manner, i.e. divergent members of the same protein family catalyse successive reactions inside a metabolic pathway. To our knowledge, only a few examples have been reliably identified to date: two pairs of enzymes in tryptophan and histidine biosynthesis [47,88].

Table 9. Number of MDR members in organisms with complete genome sequences. Number of protein coding genes in each genome were taken from NCBI (http://www.ncbi.nlm.nih.gov), except human [153,154], and fruitfly (http://www.fruitfly.org).

Organism	Number of protein coding genes	PDH Family [COG 1063]	ADH Family [COG 1062]	CADH & Y-ADH Family [COG 1064]	QOR & NRBP Family [COG 0604]	LTD Family [COG 2130]	ER Family [COG 3321]
Archaea							
Euryarchaeota							
Archaeoglobus fulgidus	2407	1	–	–	–	–	–
Methanobacterium thermoautotrophicum	1869	–	–	–	–	–	–
Methanococcus jannaschii	1715	–	–	–	–	–	–
Pyrococcus abyssi	1765	1	–	–	–	–	–
Pyrococcus horikoshii	2064	1	–	–	–	–	–
Halobacterium sp. NRC-1	2630	3	–	–	1	1	–
Crenarchaeota							
Aeropyrum pernix	2694	1	–	2	–	–	–
Bacteria							
Thermotogales							
Thermotoga maritima	1846	3	–	–	–	–	–
Spirochaetales							
Borrelia burgdorferi	850	–	–	–	–	–	–
Treponema pallidum	1031	–	–	–	–	–	–
Thermus/Deinococcus group							
Deinococcus radiodurans	2937	3	–	–	2	–	–
Chlamydiales							
Chlamydia muridarum	818	–	–	–	–	–	–
Chlamydia trachomatis	894	–	–	–	–	–	–
Chlamydia pneumoniae	1052–1110	–	–	–	–	–	–
Proteobacteria; gamma subdivision							
Buchnera sp.	564	–	–	–	–	–	–
Vibrio cholerae	3828	1	–	–	–	–	–
Escherichia coli	4289	11	2	4	2	1	–
Haemophilus influenzae	1709	1	1	–	–	–	–
Pseudomonas aeruginosa	5565	5	1	2	4	2	–
Xylella fastidiosa	2766	1	–	4	1	–	–
Proteobacteria; alpha subdivision							
Rickettsia prowazekii	834	–	–	–	–	–	–
Proteobacteria; beta subdivision							
Neisseria meningitidis	2025–2121	2	1	1	–	–	–
Proteobacteria; epsilon subdivision							
Campylobacter jejuni	1654	–	–	1	–	–	–
Helicobacter pylori	1491–1553	–	–	1–2	–	–	–
Firmicutes (Gram positives)							
Bacillus subtilis	4100	6	–	1	2	1	–
Bacillus halodurans	4066	4	–	1	3	–	–
Mycoplasma genitalium	480	–	–	–	–	–	–
Mycoplasma pneumoniae	677	1	–	–	–	–	–
Ureaplasma urealyticum	611	–	–	–	–	–	–
Actinobacteria							
Mycobacterium tuberculosis	3918	3	4	2	5	–	10
Cyanobacteria							
Synechocystis sp	3169	–	1	1	–	–	–
Aquificales							
Aquifex aeolicus	1522	–	–	–	1	–	–

Table 9. (Continued).

Organism	Number of protein coding genes	PDH Family [COG 1063]	ADH Family [COG 1062]	CADH & Y-ADH Family [COG 1064]	QOR & NRBP Family [COG 0604]	LTD Family [COG 2130]	ER Family [COG 3321]
Eukaryota							
Fungi							
Saccharomyces cerevisiae	6297	5	1	6	8	1	–
Plant							
Arabidopsis thaliana	27707	1	4	9	5	5	–
Animal							
Drosophila melanogaster	13601	3	1	1	1	–	3
Caenorhabditis elegans	20238	2	1	4	3	1	1
Homo sapiens	42 000–48 000	1	7	–	8	1	1

The data presented in our manuscript enlarge perspectives on protein evolution, because in addition to the previously mentioned mechanism of enzyme evolution, we showed that preexisting enzymes can be recruited to form novel pathways in which proteins acquire new activities by changing both their binding capacity and their chemistry of catalysis. This last possibility is in concordance with a novel third hypothesis, recently proposed by Gerlt & Babbit [47], which does not require conservation of either substrate specificity or chemical mechanisms; instead, they proposed that an active site is able to support an alternate reaction that may use some functional groups of the active site in a different mechanistic and metabolic context; in this proposal, only active site architecture is conserved. We discuss below one interesting example to support this third hypothesis. A divergent plant ADH with an acetone cyanohydrin lyase activity (P93243) has been described in flax (*Linum usitatissimum*) [90–93]. This protein belongs to a novel class of hydroxynitrile lyases (HNLs), and its amino acid sequence shows no overall homology to any cloned HNLs. Indeed, HNLs from plants form a heterogenous group of proteins differing in molecular mass, quaternary structure, presence or absence of flavin adenine dinucleotide, as well as glycosylation. They have convergently evolved from FAD-dependent oxidoreductases, α/β hydrolases, and MDRs [94]. Interestingly, HNL from flax, is a zinc-containing protein and conserves all amino acid residues important for structural integrity or coordinating zinc [91,92]; however, flax HNL neither displays ADH activity nor is inhibited by reagents interfering with zinc coordination [91]. This information, together with the fact that flax HNL is more related to plant-, animal- and class III ADH [93], suggest that flax HNL evolved late from a plant-/class III ADH, which was recruited for cyanogenesis in plants, a recent secondary pathway used as a defence mechanism against herbivorous [95]. Existence of multiple phylogenetically independent HNLs in plants supports this proposal. Therefore, this novel activity within MDR superfamily was acquired without conservation of the original binding capacity and the chemistry of catalysis. In conclusion, proteins exhibit a huge unrecognized plasticity.

Another and different alternative mechanism for enzyme evolution, also observed in members of MDR superfamily corresponds to modular construction or gene fusion, in which separate gene products join together and generate new genes containing two or more domains with novel activities [75,96]. Examples of this modular construction within the MDR superfamily are as follows: bi-domain oxidoreductase (BDOR) involved in biosynthesis of exopolysaccharides [97]; bifunctional QOR in plants, with an N-terminal domain related to short-chain dehydrogenase/reductase superfamily [98,99]; fatty acid synthase (FAS), a multifunctional polypeptide with seven enzymatic domains from animals [100] or alveolata (protozoa) [101]; modular polyketide synthase from bacteria [100], and the iterative polyketide synthase from fungi [102,103]. All of them possess modular architecture. In this sense, it is important to mention that oligomerization is not conserved among members of MDR superfamily. For example, monomers, homodimers, homotrimers, homotetramers and heterodimers, are present in this superfamily, and it has been proposed that degree of oligomerization might be involved with changes in the functional role developed by proteins [75,96].

Taken together, we conclude that the deep-rooted statements 'one enzyme, one function' and 'one protein family, one function' are not accurate for many enzymes. Several secondary activities might exist in one protein, as in the previously mentioned animal ADH or yeast ADH subfamilies (see the first topic in the Discussion section), and this can be the point of departure to gain novel and completely different functions. Indeed, we point out the fact that two different and equally efficient catalytic activities can be a feature of a single protein, as described for LTD/PGR subfamily. This catalytic promiscuity has been recognized as a vital springboard from which new catalytic activities can emerge from existing folds and active sites [70,104].

Data presented in this paper reinforce the idea that a protein can gain or lose a function through a limited number of amino acid changes, and several such examples from natural protein evolution are shown. MDR belongs to the limited number of protein superfamilies that posses both different mechanisms of reaction and substrate specificity [47,75]. Indeed, several laboratories [45,88,105] have mimicked the evolution of paralog proteins *in vitro*, showing generation of new catalytic or binding properties by modifications of a preexisting protein scaffold, and forget that evolution has carried out many such successful experiments.

Proteinogenesis vs. enzymogenesis. Several subfamilies within the MDR superfamily evolved as nonenzyme homologs, i.e. novel proteins that have lost their original catalytic activity. ζ-Crystallin/QOR is probably the most well-investigated example. This protein is expressed in a taxon-specific fashion in the lens of the phylogenetically distant guinea pig, camel, and Japanese tree frog (*Hyla japonica*) [106–109], and constitutes approximately 10% of total water-soluble proteins of the lens. Other examples of nonenzymes within the MDR superfamily are: (a) NRBP1 that functions as a transcription factor in yeast [66,67], or nuclear receptor co-operator in animals [65]; (b) dinoflagellate nuclear-associated protein (DINAP) that corresponds to the quantitatively major nuclear protein in *Crypthecodinium cohnii*, and although DINAP did not bind directly to DNA, it activated basal transcription activity [110,111]; and (c) the membrane traffic protein (AST) in fungi [112].

On the other hand, subcellular location is not conserved across members of the MDR superfamily. Although the great majority are soluble cytoplasmic proteins, some of them are located in mitochondria (yeast ADH), and nuclei (DINAP; NRBP1; class III ADH in animals), and others have a membrane location (VAT-1, and probably BDOR), or function as a structural protein (ζ-crystallin/QOR).

All these examples serve as a cogent reminder that Nature is not restricted to chemically or substrate- conserved strategies for divergent evolution; instead, divergent evolution is opportunistic and one active site architecture, can be used to develop mechanistically distinct catalytic [47] or noncatalytic functions. In other words, inside one protein superfamily (e.g. MDR), functional diversity is more complex than sequence diversity.

Eukaryota inherited MDR from bacteria

Our analysis of MDR superfamily shows that most MDR subfamilies in eukaryota are more closely related to their counterparts in bacteria than in archaea. This supports the idea that in eukaryota, although the machinery for DNA duplication, transcription, and protein synthesis is more related to archaea (informational genes), the enzymatic machinery is more related to bacteria (operational genes) [113]. This agrees with the generally accepted notion that eukaryotic cells are the symbiotic result of bacteria (the symbiont) and archaea (the host). Therefore, horizontal gene transfer of operational genes had a significant role in development of metabolic pathways in eukaryotes. In bacterial taxa, phylogenetic relationships that can be established within each protein subfamily suggest a significant horizontal gene transfer. In fact, it is calculated that nearly 20% of *Escherichia coli* genes were acquired by lateral transfer events in the last 100 million years [114]. This contrasts with the nearly complete absence of recent examples of horizontal gene transfer between species that belong to different domains of life (eukaryota, bacteria, and archaea) in MDRs. Thus, although horizontal gene transfer among bacterial taxa appears to be a recurrent event, horizontal gene transfer between bacteria and eukaryota or between bacteria and archaea is a rare event (at least in MDRs). Only two clear-cut examples were identified: the first corresponds to the previously reported horizontal gene transfer of a secondary ADH from anaerobic bacteria to the protist *Entamoeba histolytica* [115], and the second, not previously reported, corresponds to horizontal gene transfer of an LTD/AADH-related protein from firmicutes (Gram-positive bacteria) to the archaea *Halobacterium* sp. NRC-1 (NCBI accession no. AAG19273). This latter example is shown in Fig. 2, where the LTD/AADH subfamily contains some bacterial sequences that are more related to the archaea sequence (coloured in dark blue) than to other bacterial sequences within the same subfamily, obtaining a phylogenetically discordant pattern that displays a distribution compatible with horizontal gene transfer. Furthermore, this archaea sequence is the only sequence in which its branch departs far from the centre of the unrooted tree (see Fig. 2).

Is there a MDR ancestral activity?

A preliminary answer to this question can be approached from several directions, but it is clear that ancestral activity (within a protein subfamily) should be related to a primary (also ancient) metabolic pathway with (an ideally) broad phylogenetic distribution. Thus, protein subfamilies with restricted phylogenetic distribution involved in secondary metabolic pathways cannot be considered as ancestral subfamilies.

Glutathione-dependent formaldehyde dehydrogenase activity of class III ADH in ADH family (COG1062). This has been proposed as the ancient activity from which both animal and plant ADHs are derived [116]. However, this activity cannot be the ancestral function for the remaining subfamilies within the MDR superfamily, as shown by several pieces of evidence. First, glutathione (GSH) does not show the universal distribution observed for MDRs, inasmuch as GSH is restricted to proteobacteria, cyanobacteria, and eukaryotes [117,118]. Second, in organisms in which the mycothiol (MSH) molecule fulfils the functions of GSH, as in firmicutes, formaldehyde dehydrogenase activity exists in any event, but now as a mycothiol-dependent activity. A third cofactor-independent formaldehyde dehydrogenase subfamily (FADH) exists, present either in proteobacteria (with GSH), firmicutes (with MSH), and archaea (without GSH or MSH). Overall, data suggest that formaldehyde dehydrogenase activity in MDRs is very ancient and predates the origin of GSH or MSH. This is reasonable if we consider that formaldehyde reacts spontaneously with GSH or MSH to form *S*-hydroxymethylglutathione or *S*-hydroxymethyl-mycothiol, the true substrates for glutathione-dependent formaldehyde dehydrogenase (class III ADH) or mycothiol-dependent formaldehyde dehydrogenase, respectively. Furthermore, the FADH subfamily also shows formaldehyde dismutase activity and the capacity to catalyse a dismutation reaction has been conserved in animal ADH, a subfamily derived from class III ADH. Consequently, it is probable that ADH family (COG1062), absent in archaea, forms a paralogous group derived from FADH subfamily, which in turn exhibits more ample distribution than ADH family (COG1062).

Another interesting option for ancestral activity within MDR superfamily is ER; it is necessary in one of the primary (and ancient?) anabolic pathways, i.e. synthesis of

fatty acids. However, little evidence supports this proposal. First, archaea contain membranes with isoprenoid-based ether lipids, lacking fatty acids. Furthermore, gene(s) for fatty acid synthase complex (FAS), as occurs in both bacteria and eukaryotes, is (are) absent in *Methanococcus jannaschii* [119], as well as in other completely sequenced archaea genomes such as *Aeropyrum pernix* K1, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, *Pyrococcus abyssi*, and *P. horikoshii* (in agreement with our BLAST results). Thus, although archaea possess some members of MDR superfamily, ER activity probably cannot be the ancestral activity of this superfamily because archaea lacks known FAS, as well as medium-chain ER. Second, different types of FASs exist and each possesses different and unrelated ER. Thus, the ER member of the MDR superfamily is one of seven activities that comprise type I multifunctional fatty acid synthase in animals [100]. ER present in type II fatty acid synthase characteristic of bacteria and plants belongs to short-chain dehydrogenase/reductase (SDR) superfamily, not to MDR as occurs in type I animal fatty acid synthase and some bacterial polyketide synthases. Additionally, ER present in fungi (type I fatty acid synthase $\alpha 6\beta 6$ complex) does not show significant homology either to medium-chain ER or to short-chain ER (calculated with BLAST), suggesting the existence of a third class of ER. Indeed, the finding that multifunctional FAS protein exists in two distinct architectural forms, the $\alpha 2$ animal FAS and the $\alpha 6\beta 6$ yeast FAS, with protein domains arranged in a different order, is compatible with the idea that FAS complexes evolved independently several times and that they are a late acquisition in metabolic evolution of organisms, subsequent to the split of major kingdoms. Thus, both arguments strengthen the idea that ER is not an ancestral activity of the MDR superfamily. Furthermore, extensive similarity between each domain in FAS and polyketide synthase (PKS), the presence of medium-chain ER, and the order in which the domains are arranged in these multifunctional complexes [100] suggest that animal FAS is more closely related to PKS than to any other FAS from fungi, plants, or bacteria. In conclusion, there is no one member in ER family (COG3321) that can be considered as an ancestral group.

According to heterotrophic theory, the only theory with experimental support to substantiate the origin of the first metabolic pathways [120], the most ancient catabolic activities should be semienzymatic fermentative routes fed by stable and available prebiotic compounds. Thus, glycolysis, proposed as the first catabolic route [121], should have been preceded by simpler versions. The upper part of glycolysis, from hexoses to trioses, appeared as a late adaptation because glucose 6-phosphate and aldopentoses are unlikely prebiotic compounds due to rapid decomposition on a geological timescale [122]. Additionally, the step from glucose to glyceraldehyde 3-phosphate is not a universal pathway; it is absent in archaea, while there are other alternatives to transform glucose into triose derivatives [123–125]. On the other hand, the lower part of glycolysis, from glyceraldehyde 3-phosphate to pyruvate is universally conserved, and glyceraldehyde is one of the most attractive intermediates as an energy source for primitive organisms provided with nascent glycolysis. Some advantages of glyceraldehyde are: (a) it can be produced from formaldehyde under plausible prebiotic conditions [126–128]; (b) through glycolysis, it is an energy source for living purposes; (c) it is an important metabolite in photosynthesis; (d) it can be used in prebiotic condensation reactions [129,130]; and (e) it is a source of glycerol, necessary for synthesis of glycerolipids, the precursors of biomembranes.

Furthermore, results of Fukuchi & Otsuka [131] suggest that the glycolytic stage from glyceraldehydes 3-phosphate to pyruvate corresponds to one of the most ancient catabolic pathways, because genes involved in this stage of glycolysis exhibit the highest similarity to nucleotide sequences of ribosomal RNA and/or transfer RNA gene clusters, clearly predating the origin of proteic enzymes in the ancient RNA world and strongly suggesting that these metabolic pathways were developed by chance assembly of enzyme proteins generated from pre-existing genes. If this is true, it is clear that fermentative activity should be an early metabolic development to sustain activity of the ancient stage of glycolytic pathway to dispose of generated NAD(P)H. Alcoholic fermentation has been suggested as an early pathway, considering that ethanol permeates the membrane and is easily eliminated by the cell. Lactic acid fermentation should be a later development, in that lactate is a nonpenetrant product, hence retained inside the cell to be utilized to regenerate carbohydrates when autotrophic pathways became available [132]. Therefore, one ancestral activity of the MDR superfamily is probably related to an ancient alcoholic fermentative activity, such as actually observed in some subfamilies like broad ADH (from the Y-ADH family), present in eukaryota, bacteria, and archaea [133,134]; these enzymes catalyse oxidation of a broad variety of substrates, which includes primary and secondary, linear- and branched-chain, aliphatic and aromatic alcohols, in addition to several of their corresponding aldehydes and ketones. Moreover, theoretical studies predict that primordial enzymes were nonspecific, with broad substrate specificity, and showing different activities characterized by slow reaction rates [120,135]. Indeed, some MDRs fulfil all these requirements (e.g. broad ADH subfamily [133,136,137], or animal ADH subfamily [15,138]).

Finally, we cannot disregard other activities, such as threonine dehydrogenase (TDH) or crotonyl CoA-reductase (CCAR), present both in archaea and bacteria. These activities are also probably ancient. TDH is involved in amino acid metabolism, and CCAR in benzoate catabolism, acetate assimilation, and interestingly, in the supply of precursors for polyketides biosynthesis [139]. In animals, TDH initiates a minor degradative pathway [140], and the enzyme does not belong to the MDR superfamily. It is a small subfamily whose distribution is restricted to animals, and was recruited from short-chain dehydrogenase/reductase superfamily (bacterial UDP-glucose 4-epimerase, according to our BLAST analysis). On the other hand, the supply of precursors for fatty acid synthesis in bacteria and eukaryota is provided by acetyl-CoA carboxylase, an ancient enzyme also present in archaea. This suggests that the origin of acetyl-CoA carboxylase predates that of fatty acid synthesis, because fatty acids are absent in archaea. Apparently, the role of acetyl-CoA carboxylase in the supply of precursors for fatty acid synthesis is a later recruitment in the evolution of this enzyme. Thus, TDH and

CCAR probably belong to ancient metabolic pathways subsequently substituted by other metabolic pathways.

Taxonomy within the MDR superfamily

Use of the complete set of known MDR proteins, together with criteria and procedures described under the Results section, has allowed us to identify within the MDR superfamily, 49 subfamilies, and two additional taxonomic levels containing eight families and three macrofamilies. From these three taxonomic levels, only the subfamily level, as defined by us, comprises a natural unit that can be used to sort protein members of a protein superfamily with clear-cut rules. Thus, each subfamily encloses a set of ideally orthologous proteins that perform the same function, and delineate a closed group (see Results).

Two specific examples of subfamilies containing highly related paralogous rather than orthologous proteins, are the animal ADH and plant ADH subfamilies. Both subfamilies originated by successive gene duplications from an ancient class III ADH. Animal ADH evolved only in vertebrates and plant ADH, only in tracheophytas. Within the former subfamily, fishes possess one animal ADH, while amphibia, reptiles, and birds, appear to have at least two enzymes and mammals, up to six. It seems that animal ADH enzymogenesis developed in parallel to vertebrate evolution. Animal ADHs conserved the same mechanism of reaction, and share the same substrates; their main differences occur in their pattern of expression. Today, the functional roles developed by the different animal ADHs overlap, and this functional redundancy allows the individual to tolerate mutational or environmental perturbations [141]. Absence of one ADH can be overcome by the existence of other members of the animal ADH subfamily [142]. This partial functional redundancy contributes to a more general phenomenon designated 'canalization', which is the genetic capacity to buffer developmental pathways against deleterious perturbations [141]; similar advantages can be described in plant ADHs. Therefore, these singular subfamilies comprise clusters of highly related paralogous proteins that share functional roles.

A protein family, as discussed previously, must comprise a cluster of monophyletic subfamilies, i.e. highly related paralogous proteins, that all derive from a common ancestor. They possess significative sequence identity and/or similarity, and may or may not share common substrates or mechanisms of reaction.

In contrast, a protein macrofamily within MDR comprises a cluster of related protein families with broad phylogenetic distribution, i.e. with protein members from the three domains of life, and that originate from a common ancestor (monophyletics). Furthermore, within each macrofamily at least one subfamily possesses a physiological role related to primary metabolic pathways (with a probable ancient origin). Thus, the advantage of clustering protein families into macrofamilies lies in the fact that not all families are equally related, and this is probably due to the fact that some protein families are more ancient than others. Indeed, within each MDR macrofamily, there is a probable ancestral group (see the previous section), that might be tracked to the last universal common ancestor. If the latter is true, the number of macrofamilies within the MDR superfamily, reflects the original number of MDR proteins that existed in the last universal common ancestor. It is important to mention that Castresana [143], after analysing the phylogenetic distribution and evolution of bioenergetic pathways, concluded that the last universal common ancestor contained several members of each gene family. This agrees with the idea that the last universal common ancestor was a metabolically sophisticated organism.

Finally, it is interesting to point out that in comparison with the other taxonomic categories, the superfamily concept is not the focus of extensive discussion and there is a near consensus agreement that in addition to sequence similarities, and a common evolutionary origin, 3D structure data should be taken into consideration. Thus, a superfamily can be considered as groups of homologous protein families (and/or macrofamilies) with a monophyletic origin, that share, at least, a barely detectable sequence similarity, but showing similar 3D structure [144,145].

Inclusion of phylogenetic criteria to define subfamilies, families, macrofamilies, and superfamilies can be subscribed to the present tendency to construct a natural taxonomy of proteins and protein families. Figure 6 illustrates the relationships among the different taxonomic categories defined in this work.

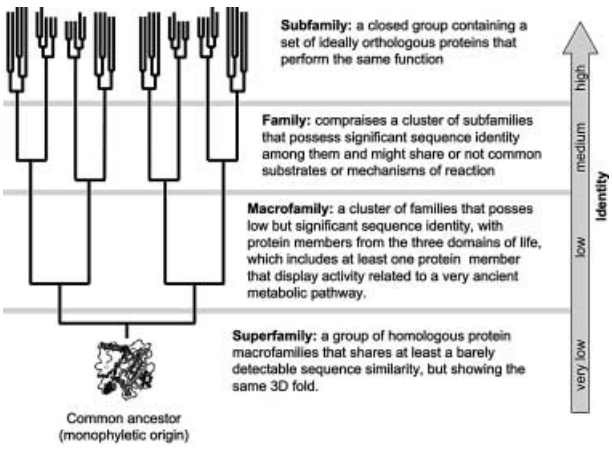


Fig. 6. Schematic display showing the main relationships among the different taxonomic categories inside a protein superfamily. Although the definition of homology has remained elusive and is the subject of intense debates [146], in this work, the concept of homologous proteins essentially refers to proteins derived from a common ancestor (phylogenetic homology). Therefore, all the taxonomic ranks comprise monophyletic groups. Identification of protein subfamilies as non-overlapping clusters (closed groups) is advantageous over distance-based clustering methods because it is not necessary to set an arbitrary identity cutoff value, and permits the identification of both highly and poorly conserved groups of orthologous proteins. Because of the huge protein plasticity, families cannot be defined by taking the function as a criterion, as only inside subfamilies (orthologous groups) is the function conserved. Macrofamilies represent probable ancestral groups that might be tracked to the last universal common ancestor; in addition, they show a wide phylogenetic range, with protein members in archaea, bacteria and eukarya.

Final consideration

After development of MDR molecular taxonomy, we propose application of the methodology employed in this paper to other protein superfamilies for several reasons. First, use of the BLASTP program in an iterative manner allows for identification of all members of any protein superfamily. Second, use of all-vs.-all BLAST-based searches within one protein superfamily together with extensive database mining, allow to sort members of any protein superfamily in subfamilies, i.e. closed groups of orthologous proteins with BLASTP reciprocal best hits. This procedure provides an advantage over classical methods for ortholog detection because it permits use of all available protein sequence members of one superfamily, bypassing global multiple alignments and construction of phylogenetic trees, which can contain slow and error-prone steps. Thus, one can benefit from all the available information without the need of selecting representative proteins and/or genomes by means of employing this faster and clear-cut procedure. In addition, the different taxonomic categories proposed in this work: subfamily, family and macrofamily, can be applied to other protein superfamilies, once formal definitions for each taxonomic rank are provided.

Acknowledgements

We thank to R.N. Ondarza (Instituto Nacional de Salud Pública, México), H. Weiner (Purdue University, USA), S. Bentley (Sanger Institute, UK), R.F. Doolittle (University of California, La Jolla, USA), A. Steinbüchel (Wilhelms-Universität, Münster, Germany), M. Pharr (North Carolina State University, USA), A. Gómez-Puyou and M. Tuena de Gómez-Puyou (Instituto de Fisiología Celular-UNAM, México), K. Yazaki (Kyoto University, Japan), A. Sosa-Peinado (Facultad de Medicina-UNAM, México), and X. Parés, J. Farrés, J. A. Biosca and their collaborators (Universitat Autònoma de Barcelona, Spain), and three anonymous referees for helpful critical review of this manuscript and/or discussions. This work was supported by grants 34823-M from CONACyT, México, and IN214101 from DGAPA-UNAM, México. H. R. R. has been supported by a graduate fellowship from DGEP-UNAM and CONACyT, México.

References

1. Reid, M.F. & Fewson, C.A. (1994) Molecular characterization of microbial alcohol dehydrogenases. *Crit. Rev. Microbiol.* **20**, 13–56.
2. Conway, T. & Ingram, L.O. (1989) Similarity of *Escherichia coli* propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *J. Bacteriol.* **171**, 3754–3759.
3. Scopes, R.K. (1983) An iron-activated alcohol dehydrogenase. *FEBS Lett.* **156**, 303–306.
4. Williamson, V.M. & Paquin, C.E. (1987) Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. *Mol. General Genet.* **209**, 374–381.
5. Krozowski, Z. (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *J. Steroid Biochem. Mol. Biol.* **51**, 125–130.
6. Persson, B., Krook, M. & Jornvall, H. (1991) Characteristics of short-chain alcohol dehydrogenases and related enzymes. *Eur. J. Biochem.* **200**, 537–543.
7. Persson, B., Zigler, J.S.J. & Jornvall, H. (1994) A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur. J. Biochem.* **226**, 15–22.
8. Jornvall, H., Hoog, J.O. & Persson, B. (1999) SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* **445**, 261–264.
9. Jornvall, H. (1999) Multiplicity and complexity of SDR and MDR enzymes. *Adv. Exp. Med. Biol.* **463**, 359–364.
10. Jornvall, H., Shafqat, J. & Persson, B. (2001) Variations and constant patterns in eukaryotic MDR enzymes. Conclusions from novel structures and characterized genomes. *Chem. Biol. Interact.* **130–132**, 491–498.
11. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
12. Pearson, W.R. & Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448.
13. Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183**, 63–98.
14. Nordling, E., Jornvall, H. & Persson, B. (2002) Medium-chain dehydrogenases/reductases (MDR). *Eur. J. Biochem.* **269**, 4267–4276.
15. Riveros-Rosas, H., Julián-Sánchez, A. & Piña, E. (1997) Enzymology of ethanol and acetaldehyde metabolism in mammals. *Arch. Med. Res.* **28**, 453–471.
16. Bairoch, A. & Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48.
17. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. & Wheeler, D.L. (2000) GenBank. *Nucleic Acids Res.* **28**, 15–18.
18. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.
19. Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358.
20. Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
21. Pennisi, E. (1999) Keeping genome databases clean and up to date. *Science* **286**, 447–450.
22. Chen, R. & Jeong, S.S. (2000) Functional prediction: identification of protein orthologs and paralogs. *Protein Sci.* **9**, 2344–2353.
23. Altschul, S.F. & Koonin, E.V. (1998) Iterated profile searches with PSI-BLAST – a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447.
24. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. & Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280.
25. Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. & Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* **30**, 235–238.
26. Barker, W.C., Garavelli, J.S., Hou, Z., Huang, H., Ledley, R.S., McGarvey, P.B., Mewes, H.W., Orcutt, B.C., Pfeiffer, F., Tsugita, A., Vinayaka, C.R., Xiao, C., Yeh, L.S. & Wu, C. (2001) Protein Information Resource: a community resource for expert annotation of protein data. *Nucleic Acids Res.* **29**, 29–32.

27. Wu, C.H., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.Z., Ledley, R.S., Lewis, K.C., Mewes, H.W., Orcutt, B.C., Suzek, B.E., Tsugita, A., Vinayaka, C.R., Yeh, L.S., Zhang, J. & Barker, W.C. (2002) The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.* **30**, 35–37.
28. Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. & Thornton, J.M. (1997) CATH – a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1108.
29. Orengo, C.A., Bray, J.E., Buchan, D.W.A., Harrison, A., Lee, D., Pearl, F.M.G., Sillitoe, I., Todd, A.E. & Thornton, J.M. (2002) The CATH protein family database: a resource for structural and functional annotation of genomes. *Proteomics* **2**, 11–21.
30. Tatusov, R.L., Koonin, E.V. & Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**, 631–637.
31. Tatusov, R.L., Galperin, M.Y., Natale, D.A. & Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36.
32. Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. & Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–28.
33. Mushegian, A.R., Garey, J.R., Martin, J. & Liu, L.X. (1998) Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* **8**, 590–598.
34. Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T., Weng, S., Cherry, J.M. & Botstein, D. (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**, 2022–2028.
35. Wheelan, S.J., Boguski, M.S., Duret, L. & Makalowski, W. (1999) Human and nematode orthologs – lessons from the analysis of 1800 human genes and the proteome of Caenorhabditis elegans. *Gene* **238**, 163–170.
36. Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor, M.G., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S.E., Chervitz, S.A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R.F., Gelbart, W.M., George, R.A., Goldstein, L.S., Gong, F., Guan, P., Harris, N.L., Hay, B.A., Hoskins, R.A., Li, J., Li, Z., Hynes, R.O., Jones, S.J., Kuehl, P.M., Lemaitre, B., Littleton, J.T., Morrison, D.K., Mungall, C., O'Farrell, P.H., Pickeral, O.K., Shue, C., Vosshall, L.B., Zhang, J., Zhao, Q., Zheng, X.H. & Lewis, S. (2000) Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215.
37. Remm, M., Storm, C.E. & Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052.
38. Krause, A. & Vingron, M. (1998) A set-theoretic approach to database searching and clustering. *Bioinformatics* **14**, 430–438.
39. Krause, A., Stoye, J. & Vingron, M. (2000) The SYSTERS protein sequence cluster set. *Nucleic Acids Res.* **28**, 270–272.
40. Krause, A., Haas, S.A., Coward, E. & Vingron, M. (2002) SYSTERS, GeneNest, SpliceNest: exploring sequence space from genome to protein. *Nucleic Acids Res.* **30**, 299–300.
41. Li, W.-H. (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA, USA.
42. Page, R.D. & Holmes, E.C. (1999) *Molecular Evolution: A. Phylogenetic Approach*. Blackwell Science, Oxford, UK.
43. Graur, D. & Li, W.-H. (1999) *Fundamentals of Molecular Evolution*. 2nd edn. Sinauer, Sunderland, MA, USA.
44. Sonnhammer, E.L.L. & Koonin, E.V. (2002) Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* **18**, 619–620.
45. Altamirano, M.M., Blackburn, J.M., Aguayo, C. & Fersht, A.R. (2000) Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* **403**, 617–622.
46. Gerlt, J.A. & Babbit, P.C. (2000) Can sequence determine function? *Genome Biol.* **1**, REVIEWS0005.1–000.10.
47. Gerlt, J.A. & Babbit, P.C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246.
48. Iborra, F.J., Renau-Piqueras, J., Portoles, M., Boleda, M.D., Guerri, C. & Pares, X. (1992) Immunocytochemical and biochemical demonstration of formaldehyde dehydrogenase (class III alcohol dehydrogenase) in the nucleus. *J. Histochem. Cytochem.* **40**, 1865–1878.
49. Peralba, J.M., Cederlund, E., Crosas, B., Moreno, A., Julia, P., Martinez, S.E., Persson, B., Farres, J., Pares, X. & Jornvall, H. (1999) Structural and enzymatic properties of a gastric NADP(H)-dependent and retinal-active alcohol dehydrogenase. *J. Biol. Chem.* **274**, 26021–26026.
50. Valencia, E., Rosell, A., Larroy, C., Farres, J., Biosca, J.A., Fita, I., Pares, X. & Ochoa, W.F. (2003) Crystallization and preliminary X-ray analysis of NADP(H)-dependent alcohol dehydrogenases from *Saccharomyces cerevisiae* and *Rana perezi*. *Acta Crystallogr. D-Biol. Cryst.* **59**, 334–337.
51. Eggeling, L. & Sahm, H. (1985) The formaldehyde dehydrogenase of *Rhodococcus erythropolis*, a trimeric enzyme requiring a cofactor and active with alcohols. *Eur. J. Biochem.* **150**, 129–134.
52. Norin, A., Van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A. & Jornvall, H. (1997) Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaroytic alcohol dehydrogenases – primary structure, conformational modelling and functional correlations. *Eur. J. Biochem.* **248**, 282–289.
53. Van Ophem, P.W., Van Beeumen, J. & Duine, J.A. (1992) NAD-linked, factor-dependent formaldehyde dehydrogenase or trimeric, zinc-containing, long-chain alcohol dehydrogenase from Amycolatopsis methanolica. *Eur. J. Biochem.* **206**, 511–518.
54. Duester, G., Farres, J., Felder, M.R., Holmes, R.S., Hoog, J.O., Pares, X., Plapp, B.V., Yin, S.J. & Jornvall, H. (1999) Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem. Pharmacol.* **58**, 389–395.
55. Tadege, M., Dupuis, I. & Kuhlemeier, C. (1999) Ethanolic fermentation: new functions for an old pathway. *Trends Plant Sci.* **4**, 320–325.
56. Larroy, C., Pares, X. & Biosca, J.A. (2002) Characterization of a *Saccharomyces cerevisiae* NADP(H)-dependent alcohol dehydrogenase (ADHVII), a member of the cinnamyl alcohol dehydrogenase family. *Eur. J. Biochem.* **269**, 5738–5745.
57. Larroy, C., Fernandez, M.R., Gonzalez, E., Pares, X. & Biosca, J.A. (2002) Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction. *Biochem. J.* **361**, 163–172.
58. Logemann, E., Reinold, S., Somssich, I.E. & Hahlbrock, K. (1997) A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. *Biol. Chem.* **378**, 909–913.
59. Brill, E.M., Abrahams, S., Hayes, C.M., Jenkins, C.L. & Watson, J.M. (1999) Molecular characterisation and expression of a wound-inducible cDNA encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne (*Medicago sativa* L.). *Plant Mol. Biol.* **41**, 279–291.

60. Williamson, J.D., Stoop, J.M., Massel, M.O., Conkling, M.A. & Pharr, D.M. (1995) Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein ELI3. *Proc. Natl Acad. Sci. USA* **92**, 7148–7152.
61. Quirino, B.F., Normanly, J. & Amasino, R.M. (1999) Diverse range of gene activity during *Arabidopsis thaliana* leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Mol. Biol.* **40**, 267–278.
62. Prata, R.T.N., Williamson, J.D., Conkling, M.A. & Pharr, D.M. (1997) Sugar repression of mannitol dehydrogenase activity in celery cells. *Plant Physiol.* **114**, 307–314.
63. Stoop, J.M.H., Williamson, J.D., Conkling, M.A., MacKay, J.J. & Pharr, D.M. (1998) Characterization of NAD-dependent mannitol dehydrogenase from celery as affected by ions, chelators, reducing agents and metabolites. *Plant Sci.* **131**, 43–51.
64. Pharr, D.M., Prata, R.T.N., Jennings, D.B., Williamson, J.D., Zamski, E., Yamamoto, Y.T. & Conkling, M.A. (1999) Regulation of mannitol dehydrogenase: relationship to plant growth and stress tolerance. *Hortscience* **34**, 1027–1032.
65. Masuda, N., Yasumo, H., Furusawa, T., Tsukamoto, T., Sadano, H. & Osumi, T. (1998) Nuclear receptor binding factor-1 (NRBF-1), a protein interacting with a wide spectrum of nuclear hormone receptors. *Gene* **221**, 225–233.
66. Yamazoe, M., Shirahige, K., Rashid, M.B., Kaneko, Y., Nakayama, T., Ogasawara, N. & Yoshikawa, H. (1994) A protein which binds preferentially to single-stranded core sequence of autonomously replicating sequence is essential for respiratory function in mitochondria of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **269**, 15244–15252.
67. Owen, G.I. & Zelent, A. (2000) Origins and evolutionary diversification of the nuclear receptor superfamily. *Cell. Mol. Life Sci.* **57**, 809–827.
68. Metz, J.G., Pollard, M.R., Anderson, L., Hayes, T.R. & Lassner, M.W. (2000) Purification of a jojoba embryo fatty acyl-coenzyme A reductase and expression of its cDNA in high erucic acid rapeseed. *Plant Physiol.* **122**, 635–644.
69. Wang, X. & Kolattukudy, P.E. (1995) Solubilization, purification and characterization of fatty acyl-CoA reductase from duck uropygial gland. *Biochem. Biophys. Res. Commun.* **208**, 210–215.
70. O'Brien, P.J. & Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105.
71. Henehan, G.T. & Oppenheimer, N.J. (1993) Horse liver alcohol dehydrogenase-catalyzed oxidation of aldehydes: dismutation precedes net production of reduced nicotinamide adenine dinucleotide. *Biochemistry* **32**, 735–738.
72. Svensson, S., Lundsjo, A., Cronholm, T. & Hoog, J.O. (1996) Aldehyde dismutase activity of human liver alcohol dehydrogenase. *FEBS Lett.* **394**, 217–220.
73. Tsai, C.S. (1982) Multifunctionality of liver alcohol dehydrogenase: kinetic and mechanistic studies of esterase reaction. *Arch. Biochem. Biophys.* **213**, 635–642.
74. Kusano, M., Sakai, Y., Kato, N., Yoshimoto, H., Sone, H. & Tamai, Y. (1998) Hemiacetal dehydrogenation activity of alcohol dehydrogenases in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* **62**, 1956–1961.
75. Todd, A.E., Orengo, C.A. & Thornton, J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**, 1113–1143.
76. Clish, C.B., Levy, B.D., Chiang, N., Tai, H.H. & Serhan, C.N. (2000) Oxidoreductases in lipoxin A4 metabolic inactivation: a novel role for 15-oxoprostaglandin 13-reductase/leukotriene B4 12-hydroxydehydrogenase in inflammation. *J. Biol. Chem.* **275**, 25372–25380.
77. Galperin, M.Y., Walker, D.R. & Koonin, E.V. (1998) Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* **8**, 779–790.
78. Todd, A.E., Orengo, C.A. & Thornton, J.M. (1999) Evolution of protein function, from a structural perspective. *Curr. Opin. Chem. Biol.* **3**, 548–556.
79. Benach, J., Atrian, S., Ladenstein, R. & Gonzalez-Duarte, R. (2001) Genesis of *Drosophila* ADH: the shaping of the enzymatic activity from a SDR ancestor. *Chem. Biol. Interact.* **130–132**, 405–415.
80. Pal, G.P., Jany, K.D. & Saenger, W. (1987) Crystallization of and X-ray investigations on glucose dehydrogenase from *Bacillus megaterium*. *Eur. J. Biochem.* **167**, 123–124.
81. Yamamoto, K., Kusunoki, M., Urabe, I., Tabata, S. & Osaki, S. (2000) Crystallization and preliminary X-ray analysis of glucose dehydrogenase from Bacillus megaterium IWG3. *Acta. Crystallogr. D Biol. Crystallogr.* **56**, 1443–1445.
82. Baldock, C., Rafferty, J.B., Stuitje, A.R., Slabas, A.R. & Rice, D.W. (1998) The X-ray structure of *Escherichia coli* enoyl reductase with bound NAD^+ at 2.1 Å resolution. *J. Mol. Biol.* **284**, 1529–1546.
83. Kater, M.M., Koningstein, G.M., Nijkamp, H.J. & Stuitje, A.R. (1994) The use of a hybrid genetic system to study the functional relationship between prokaryotic and plant multi-enzyme fatty acid synthetase complexes. *Plant Mol. Biol.* **25**, 771–790.
84. Slabas, A.R., Cottingham, I., Austin, A., Fawcett, T. & Sidebottom, C.M. (1991) Amino acid sequence analysis of rape seed (*Brassica napus*) NADH-enoyl ACP reductase. *Plant Mol. Biol.* **17**, 911–914.
85. Jornvall, H., von Bahr-Lindstrom, H., Jany, K.D., Ulmer, W. & Froschle, M. (1984) Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases. *FEBS Lett.* **165**, 190–196.
86. Edgar, A.J. (2002) Molecular cloning and tissue distribution of mammalian L-threonine 3-dehydrogenases. *BMC Biochem.* **3**, 19.
87. True, J.R. & Carroll, S.B. (2002) Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell. Dev. Biol.* **18**, 53–80.
88. Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000) Directed evolution of a (beta alpha) 8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl Acad. Sci. USA* **97**, 9925–9930.
89. Horowitz, N.H. (1945) On the evolution of biochemical syntheses. *Proc. Natl Acad. Sci. USA* **31**, 153–157.
90. Xu, L.L., Singh, B.K. & Conn, E.E. (1988) Purification and characterization of acetone cyanohydrin lyase from Linum usitatissimum. *Arch. Biochem. Biophys.* **263**, 256–263.
91. Trummler, K. & Wajant, H. (1997) Molecular cloning of acetone cyanohydrin lyase from flax (*Linum usitatissimum*). Definition of a novel class of hydroxynitrile lyases. *J. Biol. Chem.* **272**, 4770–4774.
92. Trummler, K., Roos, J., Schwaneberg, U., Effenberger, F., Forster, S., Pfizenmaier, K. & Wajant, H. (1998) Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris* – utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. *Plant Sci.* **139**, 19–27.
93. Breithaupt, H., Pohl, M., Bonigk, W., Heim, P., Schimz, K.L. & Kula, M.R. (1999) Cloning and expression of (*R*)-hydroxynitrile lyase from *Linum usitatissimum* (flax). *J. Mol. Catal. B-Enzym.* **6**, 315–332.
94. Dreveny, I., Gruber, K., Glieder, A., Thompson, A. & Kratky, C. (2001) The hydroxynitrile lyase from almond: a lyase that looks like an oxidoreductase. *Structure* **9**, 803–815.
95. Vetter, J. (2000) Plant cyanogenic glycosides. *Toxicon* **38**, 11–36.

96. Thornton, J.M., Orengo, C.A., Todd, A.E. & Pearl, F.M. (1999) Protein folds, functions and evolution. *J. Mol. Biol.* **293**, 333–342.
97. Nakar, D. & Gutnick, D.L. (2001) Analysis of the wee gene cluster responsible for the biosynthesis of the polymeric bioemulsifier from the oil- degrading strain *Acinetobacter lwoffii* RAG-1. *Microbiology* **147**, 1937–1946.
98. Babiychuk, E., Kushnir, S., Belles-Boix, E., Van Montagu, M. & Inze, D. (1995) *Arabidopsis thaliana* NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. *J. Biol. Chem.* **270**, 26224–26231.
99. Ichinose, Y., Tiemann, K., Schwenger-Erger, C., Toyoda, K., Hein, F., Hanselle, T., Cornels, H. & Barz, W. (2000) Genes expressed in *Ascochyta rabiei*-inoculated chickpea plants and elicited cell cultures as detected by differential cDNA-hybridization. *Z. Naturforsch C.* **55**, 44–54.
100. Smith, S. (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* **8**, 1248–1259.
101. Zhu, G., Marchewka, M.J., Woods, K.M., Upton, S.J. & Keithly, J.S. (2000) Molecular analysis of a Type I fatty acid synthase in *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* **105**, 253–260.
102. Hutchinson, C.R., Kennedy, J., Park, C., Kendrew, S., Auclair, K. & Vederas, J. (2000) Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. *Antonie Van Leeuwenhoek* **78**, 287–295.
103. Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C. & Hutchinson, C.R. (1999) Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368–1372.
104. James, L.C. & Tawfik, D.S. (2001) Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. *Protein Sci.* **10**, 2600–2607.
105. Tao, H. & Cornish, V.W. (2002) Milestones in directed enzyme evolution. *Curr. Opin. Chem. Biol.* **6**, 858–864.
106. Garland, D., Rao, P.V., Del Corso, A., Mura, U. & Zigler, J.S.J. (1991) zeta-Crystallin is a major protein in the lens of *Camelus dromedarius*. *Arch. Biochem. Biophys.* **285**, 134–136.
107. Rao, P.V. & Zigler, J.S.J. (1992) Purification and characterization of zeta-crystallin/quinone reductase from guinea pig liver. *Biochim. Biophys. Acta* **1117**, 315–320.
108. Gonzalez, P., Rao, P.V., Nunez, S.B. & Zigler, J.S.J. (1995) Evidence for independent recruitment of zeta-crystallin/quinone reductase (CRYZ) as a crystallin in camelids and hystricomorph rodents. *Mol. Biol. Evol.* **12**, 773–781.
109. Fujii, Y., Kimoto, H., Ishikawa, K., Watanabe, K., Yokota, Y., Nakai, N. & Taketo, A. (2001) Taxon-specific zeta-crystallin in Japanese tree frog (Hyla japonica) lens. *J. Biol. Chem.* **276**, 28134–28139.
110. Bhaud, Y., Geraud, M.L., Ausseil, J., Soyer-Gobillard, M.O. & Moreau, H. (1999) Cyclic expression of a nuclear protein in a dinoflagellate. *J. Eukaryot. Microbiol.* **46**, 259–267.
111. Guillebault, D., Derelle, E., Bhaud, Y. & Moreau, H. (2001) Role of nuclear WW domains and proline-rich proteins in dinoflagellate transcription. *Protist* **152**, 127–138.
112. Chang, A. & Fink, G.R. (1995) Targeting of the yeast plasma membrane [H^+]ATPase: a novel gene AST1 prevents mislocalization of mutant ATPase to the vacuole. *J. Cell Biol.* **128**, 39–49.
113. Jain, R., Rivera, M.C. & Lake, J.A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806.
114. Lawrence, J.G. & Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417.
115. Field, J., Rosenthal, B. & Samuelson, J. (2000) Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.* **38**, 446–455.
116. Shafqat, J., El-Ahmad, M., Danielsson, O., Martinez, M.C., Persson, B., Pares, X. & Jornvall, H. (1996) Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). *Proc. Natl Acad. Sci. USA* **93**, 5595–5599.
117. Ondarza, R.N., Rendon, J.L. & Ondarza, M. (1983) Glutathione reductase in evolution. *J. Mol. Evol.* **19**, 371–375.
118. Fahey, R.C. & Sundquist, A.R. (1991) Evolution of glutathione metabolism. *Adv. Enzymol. Relat. Areas Mol. Biol.* **64**, 1–53.
119. Selkov, E., Maltsev, N., Olsen, G.J., Overbeek, R. & Whitman, W.B. (1997) A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. Gene 197, GC11-GC26.
120. Lazcano, A. & Miller, S.L. (1999) On the origin of metabolic pathways. *J. Mol. Evol.* **49**, 424–431.
121. Fothergill-Gilmore, L.A. & Michels, P.A. (1993) Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* **59**, 105–235.
122. Larralde, R., Robertson, M.P. & Miller, S.L. (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc. Natl Acad. Sci. USA* **92**, 8158–8160.
123. Conway, T. (1992) The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol. Rev.* **9**, 1–27.
124. Romano, A.H. & Conway, T. (1996) Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* **147**, 448–455.
125. Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.* **343**, 115–124.
126. Gabel, N.W. & Ponnamperuma, C. (1967) Model for origin of monosaccharides. *Nature* **216**, 453–455.
127. Reid, C. & Orgel, L.E. (1967) Synthesis in sugars in potentially prebiotic conditions. *Nature* **216**, 455.
128. Epps, D.E., Nooner, D.W., Eichberg, J., Sherwood, E. & Oro, J. (1979) Cyanamide mediated synthesis under plausible primitive earth conditions. VI. The synthesis of glycerol and glycerophosphates. *J. Mol. Evol.* **14**, 235–241.
129. Weber, A.L. (1987) The triose model: glyceraldehyde as a source of energy and monomers for prebiotic condensation reactions. *Orig. Life Evol. Biosph.* **17**, 107–119.
130. Weber, A.L. & Hsu, V. (1990) Energy-rich glyceric acid oxygen esters: implications for the origin of glycolysis. *Orig. Life Evol. Biosph.* **20**, 145–150.
131. Fukuchi, S. & Otsuka, J. (1992) Evolution of metabolic pathways by chance assembly of enzyme proteins generated from sense and antisense strands of pre-existing genes. *J. Theor. Biol.* **158**, 271–291.
132. Skulachev, V.P. (1994) Bioenergetics: the evolution of molecular mechanisms and the development of bioenergetic concepts. *Antonie Van Leeuwenhoek* **65**, 271–284.
133. Rella, R., Raia, C.A., Pensa, M., Pisani, F.M., Gambacorta, A., De Rosa, M. & Rossi, M. (1987) A novel archaebacterial NAD^+-dependent alcohol dehydrogenase. Purification Properties. *Eur. J. Biochem.* **167**, 475–479.
134. Peretz, M., Bogin, O., Keinan, E. & Burstein, Y. (1993) Stereospecificity of hydrogen transfer by the NADP-linked alcohol dehydrogenase from the thermophilic bacterium *Thermoanaerobium brockii*. *Int. J. Pept. Protein Res.* **42**, 490–495.
135. Demetrius, L. (1998) Role of enzyme-substrate flexibility in catalytic activity: an evolutionary perspective. *J. Theor. Biol.* **194**, 175–194.
136. Giordano, A., Cannio, R., La Cara, F., Bartolucci, S., Rossi, M. & Raia, C.A. (1999) Asn249Tyr substitution at the coenzyme

binding domain activates *Sulfolobus solfataricus* alcohol dehydrogenase and increases its thermal stability. *Biochemistry* **38**, 3043–3054.
137. Raia, C.A., Caruso, C., Marino, M., Vespa, N. & Rossi, M. (1996) Activation of Sulfolobus solfataricus alcohol dehydrogenase by modification of cysteine residue 38 with iodoacetic acid. *Biochemistry* **35**, 638–647.
138. Pietruszko, R. (1979) Nonethanol substrates of alcohol dehydrogenase. In *Biochemistry and Pharmacology of Ethanol* (Majchrowicz, E. & Noble, E.P., eds), pp. 87–106. Plenum Press, New York, USA.
139. Liu, H. & Reynolds, K.A. (2001) Precursor supply for polyketide biosynthesis: the role of crotonyl-CoA reductase. *Metab. Eng.* **3**, 40–48.
140. Darling, P.B., Grunow, J., Rafii, M., Brookes, S., Ball, R.O. & Pencharz, P.B. (2000) Threonine dehydrogenase is a minor degradative pathway of threonine catabolism in adult humans. *Am. J. Physiol. Endocrinol. Metab.* **278**, E877–E884.
141. Wilkins, A.S. (1997) Canalization: a molecular genetic perspective. *Bioessays* **19**, 257–262.
142. Deltour, L., Foglio, M.H. & Duester, G. (1999) Metabolic deficiencies in alcohol dehydrogenase Adh1, Adh3, and Adh4 null mutant mice. Overlapping roles of Adh1 and Adh4 in ethanol clearance and metabolism of retinol to retinoic acid. *J. Biol. Chem.* **274**, 16796–16801.
143. Castresana, J. (2001) Comparative genomics and bioenergetics. *Biochim. Biophys. Acta* **1506**, 147–162.
144. Koonin, E.V., Tatusov, R.L. & Galperin, M.Y. (1998) Beyond complete genomes: from sequence to structure and function. *Curr. Opin. Struct. Biol.* **8**, 355–363.
145. Koonin, E.V., Wolf, Y.I. & Karev, G.P. (2002) The structure of the protein universe and genome evolution. *Nature* **420**, 218–223.
146. Butler, A.B. & Saidel, W.M. (2000) Defining sameness: historical, biological, and generative homology. *Bioessays* **22**, 846–853.
147. Lo, C.L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G. & Chothia, C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res.* **28**, 257–259.
148. Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M. & Servant, F. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40.
149. Ng, K.YeR., Wu, X.C. & Wong, S.L. (1992) Sorbitol dehydrogenase from *Bacillus subtilis*. Purification, characterization, and gene cloning. *J. Biol. Chem.* **267**, 24989–24994.
150. Marini, I., Bucchioni, L., Borella, P., Del Corso, A. & Mura, U. (1997) Sorbitol dehydrogenase from bovine lens: purification and properties. *Arch. Biochem. Biophys.* **340**, 383–391.
151. Lindstad, R.I., Koll, P. & McKinley-McKee, J.S. (1998) Substrate specificity of sheep liver sorbitol dehydrogenase. *Biochem. J.* **330**, 479–487.
152. Oura, Y., Yamada, K., Shiratake, K. & Yamaki, S. (2000) Purification and characterization of a NAD^+-dependent sorbitol dehydrogenase from Japanese pear fruit. *Phytochemistry* **54**, 567–572.
153. Hogenesch, J.B., Ching, K.A., Batalov, S., Su, A.I., Walker, J.R., Zhou, Y., Kay, S.A., Schultz, P.G. & Cooke, M.P. (2001) A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* **106**, 413–415.
154. Shouse, B. (2002) American Association for the Advancement of Science Annual Meeting. Human gene count on the rise. *Science* **295**, 1457.

Supplementary material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/EJB/EJB3704/EJB3704sm.htm

Table S1. Proteins that belong to MDR superfamily.

Table S2. References for Tables 3–8.

Figure 1

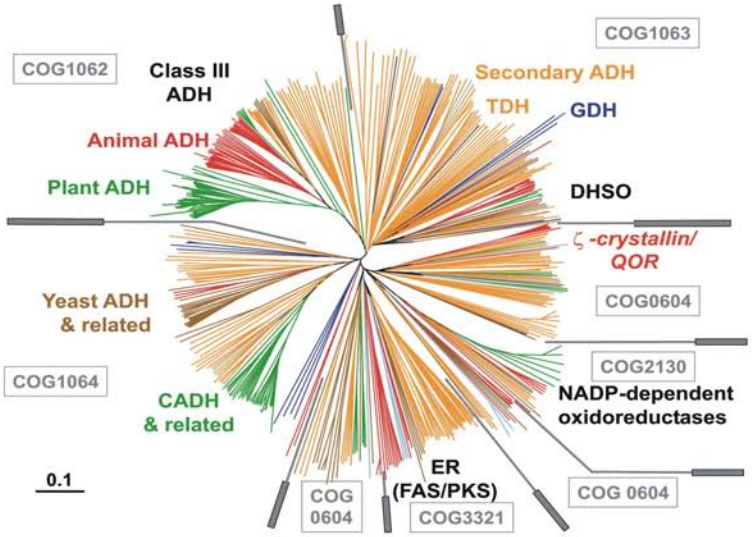


Figure 5

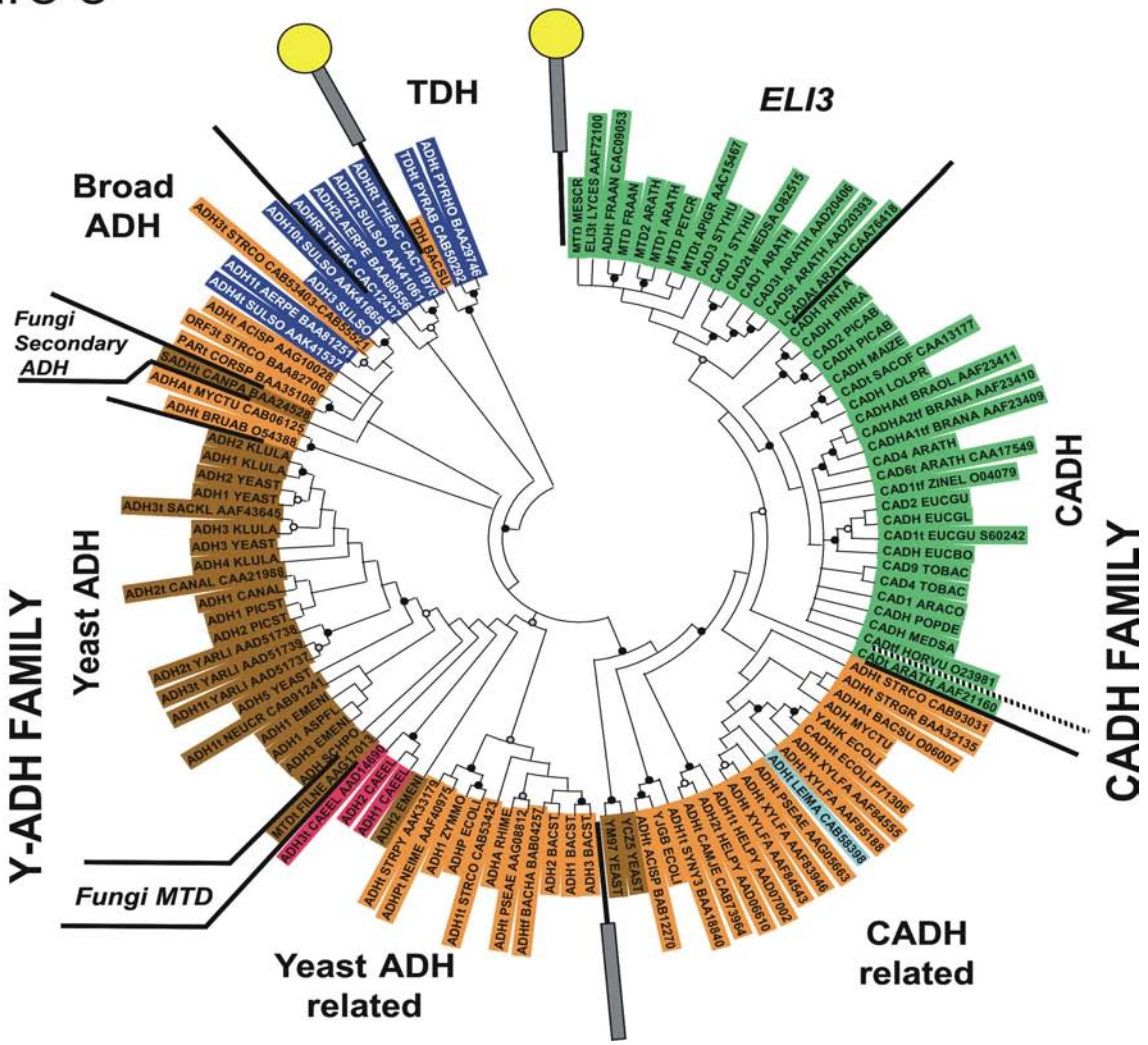


Figure 2

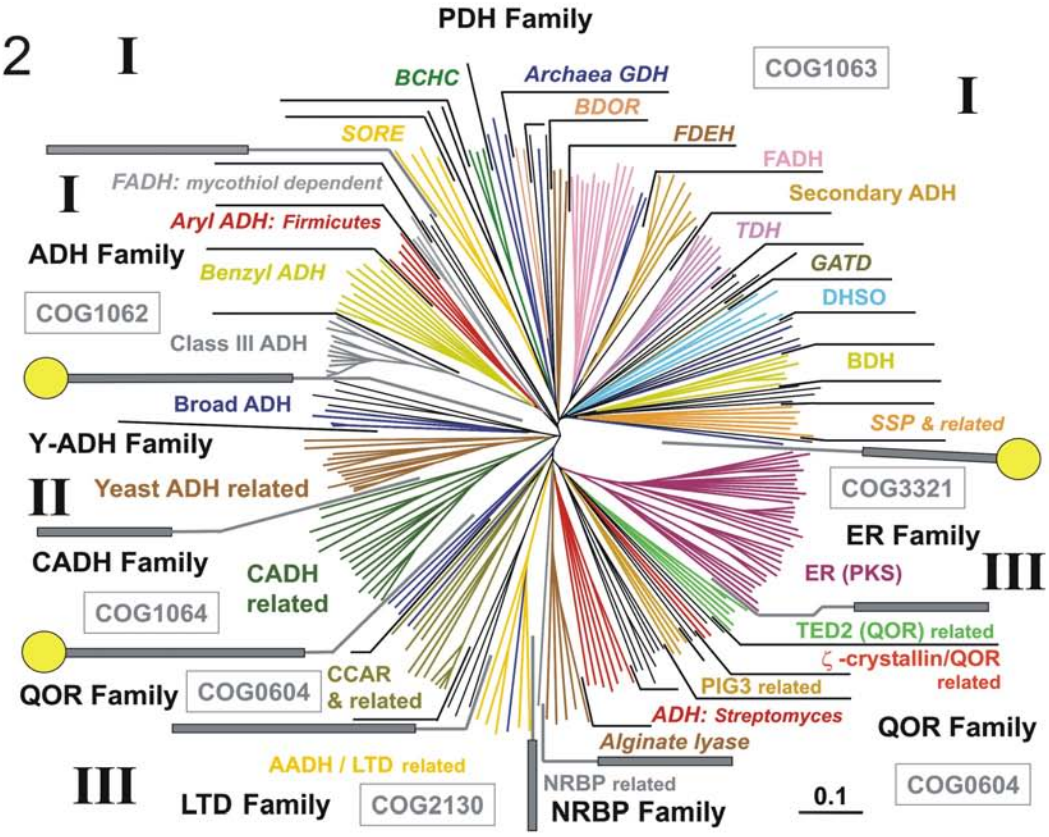
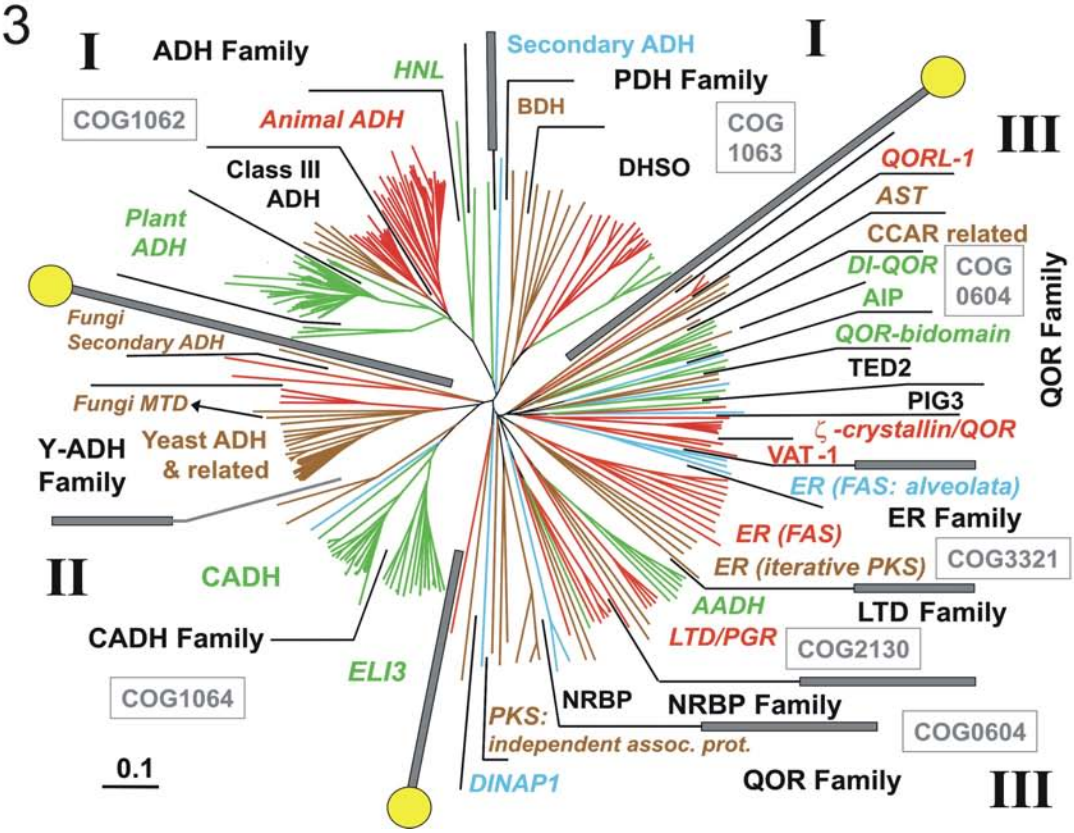


Figure 3



Discusión

1. Problemas con las bases de datos (La Babel bioinformática)

Actualmente se tienen en las bases de datos públicas más de un millón de secuencias de proteínas, las cuales se pueden agrupar en poco más de 1000 superfamilias de proteínas distintas. La asignación de una nueva proteína en alguna de las superfamilias ya descritas suele ser una tarea rutinaria, y existen diversas herramientas que permiten hacerlo con un margen de error pequeño. En contraste, no se tiene consenso sobre los criterios para identificar dentro de una superfamilia, las familias o subfamilias que la conforman. La identificación de la función desempeñada por una nueva secuencia de proteínas (no caracterizada experimentalmente) suele efectuarse a través de comparar dicha secuencia con el resto de las secuencias de proteínas, utilizando herramientas como BLASTP o FASTA. Así, con base en la proteína que muestra una mayor identidad con la proteína a identificar (BLAST's reciprocal best hit), se propone la posible función de la proteína. Sin embargo, debido a que una fracción considerable de las proteínas no ha sido caracterizada, es frecuente que la proteína que corresponde al mejor resultado (hit), resulte ser una proteína no caracterizada experimentalmente y cuya función se propuso con base en su similitud con otra proteína previamente identificada. De esta manera, no es extraño encontrarse con la situación de que la función de una proteína se propone con base en la similitud de ésta con una proteína que es similar a otra que a su vez es la similar de la similar de otra proteína que finalmente sí se determinó experimentalmente su función. Con base en consideraciones como ésta, se ha detectado el problema de que una fracción significativa de las proteínas no caracterizadas experimentalmente, pero cuya secuencia esta registrada en las bases de datos, tienen anotaciones dudosas o francamente incorrectas sobre su función (Pennisi, 1999; Chen y Jeong, 2000). Este problema de la no caracterización de las proteínas registradas en las bases de datos se acentúa por el hecho de que existe también una fracción considerable de enzimas cuya función ya esta caracterizada experimentalmente, pero no registrada en las bases de datos. Esta última situación, como ya se comentó antes en la sección de material y métodos, se da por el hecho de que los autores deben registrar en las bases de datos las nuevas secuencias de proteínas, antes de enviar a publicación sus resultados. Sin embargo, frecuentemente los autores, después de publicados sus resultados, no regresan al registro original de las bases de datos, para incluir la cita de la publicación en donde se caracteriza la proteína.

Con respecto a la superfamilia de proteínas MDR, este problema es particularmente significativo en el caso de las quinona oxidorreductasas, tal y como se discutió en la sección de resultados (página 44 de esta tesis, o página 3311 del artículo original). De esta manera, uno de

los primeros problemas que tuvo que ser resuelto antes de iniciar propiamente el análisis, fue establecer los criterios para la identificación de proteínas ortólogas, para poder así asignar, con confianza razonable, una probable función a las proteínas no caracterizadas.

2. Identificación de proteínas ortólogas

La identificación de proteínas ortólogas y parálogas es un problema que no ha sido fácil de abordar, ya que se requiere no sólo de un análisis de secuencias y de la construcción de árboles filogenéticos, sino también de información acerca de la función que desarrollan las proteínas. Inclusive, el concepto mismo de ortólogo y parálogo parece no ser muy claro para muchos biólogos moleculares, o bien se manejan de forma poco cuidadosa.

El concepto de ortología y paralogía se originó del campo de la sistemática molecular, en donde Walter M. Fitch, en 1970, propuso ambos conceptos para distinguir entre los diferentes tipos de proteínas homólogas (Fitch, 1970). Así, ortólogos y parálogos constituyen los dos tipos principales de homólogos: los primeros se originan a partir de un ancestro común por procesos de especiación, de tal forma que la historia de las proteínas ortólogas refleja la historia de las especies (del griego *orto* = correcto), mientras que los segundos se originan por procesos de duplicación dentro de una misma especie, de tal forma que la historia de las proteínas duplicadas se desarrolla en forma paralela a partir de su origen (del griego *para* = junto a).

En la práctica, existen diversas situaciones que dificultan la identificación de grupos ortólogos: 1) la excesiva duplicación de genes en los genomas eucariontes y la redundancia funcional, 2) la estructura multidominio de muchas proteínas, originada por la fusión de genes, 3) la predominancia de genomas incompletamente secuenciados en eucariontes (Li y cols., 2003), y 4) la transferencia horizontal de genes, predominante en organismos procariontes.

No obstante esto, la importancia de identificar grupos ortólogos dentro de las proteínas ha sido reconocida por diversos grupos de investigación. De esta manera, se han construido diferentes bases de datos que tienen por objetivo la identificación de grupos ortólogos, por ejemplo, la base de datos COG (Clusters of Orthologous Groups) está construida con base en los resultados de comparación por BLAST de todos-contra-todos, en organismos cuyos genomas están totalmente secuenciados y en los cuales se conocen sus proteomas completos. Así, en COG se identifican como grupos ortólogos a los conjuntos de secuencias pertenecientes a distintos proteomas que poseen recíprocamente entre ellas los mejores resultados de BLAST (BLAST's reciprocal best hits), y que pertenecen al menos a tres líneas filogenéticas independientes. Esta forma de análisis permite identificar grupos ortólogos de origen antiguo, pero no distingue

parálogos de origen "reciente"[1] que se originaron en una única línea filogenética. Por esta razón, no es probable encontrar COGs conformados por proteínas que desarrollen una misma función y, en realidad, sólo se puede tener certeza de que las COGs conforman grupos que comparten un mismo ancestro común (monofiléticos). Esta es la razón por la cual las COG corresponden más bien a las familias de proteínas identificadas en este trabajo y no a grupos ortólogos reales.

Las bases de datos *EGO* (Lee y cols., 2002), desarrollada en The Institute for Genomic Research y *OrthoMCL* (Li y cols., 2003), desarrollada en la University of Pennsylvania, son otras bases de datos generadas por procesos automatizados que efectúan también la identificación de grupos ortólogos mediante comparaciones con BLAST de todos-contra-todos, y la identificación de grupos de secuencias de proteínas que poseen recíprocamente entre ellas los mejores resultados de BLAST (BLAST's reciprocal best hits). Sin embargo, *EGO* y *OrthoMCL* tienen la desventaja de sólo emplear un conjunto selecto de especies y no el total de genes disponible en el GenBank. La base de datos *SYSTER*, usando criterios de identificación semejantes a los arriba mencionados, resuelve este último problema, al utilizar el total de secuencias reportadas en el GenBank. Sin embargo, ninguna de estas tres bases de datos se apoya en la construcción de árboles filogenéticos para definir los grupos de proteínas ortólogas.

Por otra parte, cuando se utilizan criterios filogenéticos para la identificación de grupos ortólogos, se tiene la desventaja de usar valores de identidad arbitrarios para definir los límites entre grupos ortólogos, sin considerar que cada grupo de proteínas ortólogas tiene diferencias en cuanto a su capacidad para soportar sustituciones en su secuencia de aminoácidos, esto es, existen proteínas que pueden experimentar modificaciones importantes en su secuencia de aminoácidos sin que se altere su función, mientras que otras son menos tolerantes. Así, la utilización de un valor límite único de identidad puede generar tanto falsos positivos como falsos negativos dentro de los grupos ortólogos así identificados.

En este trabajo, la utilización de criterios no basados en distancias, similares a los utilizados por la base de datos *SYSTERS*, tiene la ventaja de que los límites entre los diferentes grupos ortólogos están determinados por los mismos datos y no por un valor límite de identidad fijado arbitrariamente por el autor/investigador. Así, es posible identificar tanto grupos ortólogos (subfamilias) muy conservados como pobremente conservados. Además, este método es altamente conservativo, por lo que la probabilidad de obtener un falso positivo es extremadamente baja. Sin embargo, es necesario reconocer que la probabilidad de obtener falsos negativos es un poco más alta, aunque a nivel global, al menos dentro de la superfamilia de las MDR, este problema no resultó significativo, ya que fueron muy pocas las secuencias que no pudieron ser

[1] Los parálogos "recientes" se originan por eventos de duplicación posteriores a la especiación, y pueden por tanto estar igualmente relacionados a otros ortólogos en especies diferentes en donde no se dio un proceso de duplicación.

asignadas a una subfamilia y que se mantuvieron por tanto en la zona de indefinición (twilight zone).

Por último, es importante señalar que los límites de los grupos ortólogos identificados mediante "BLAST's reciprocal best hits", no pueden ser definidos con toda precisión si no se consideran criterios adicionales, como la función de las proteínas (determinada experimentalmente) y el patrón filogenético de las secuencias de proteínas que conforman un grupo ortólogo. Así, con la inclusión de estos dos criterios para definir los límites entre los grupos ortólogos (subfamilias de proteínas), se puede tener la certeza de mejorar significativamente la clasificación originalmente propuesta para las MDR en la base de datos *SYSTER*, construida con base en procedimientos automatizados.

3. Taxonomía de proteínas.

La taxonomía de proteínas presenta particularidades y problemas que no se presentan en la taxonomía clásica. Para empezar, la taxonomía de los seres vivos comprende al menos siete diferentes niveles taxonómicos ampliamente aceptados: reino, phylum o división, clase, orden, familia, género y especie. Si se consideran los dominios archaea, bacteria y eukarya como una jerarquía adicional, se tienen entonces 8 niveles taxonómicos diferentes para ordenar y clasificar, con un criterio filogenético, la gran diversidad de los seres vivos actualmente existentes. Idealmente, una taxonomía natural debe reflejar la historia evolutiva de los seres vivos. (Nota: los paleontólogos por su parte, tienen también sus problemas particulares por la necesidad de integrar a las especies vivas con las extintas en un mismo sistema de clasificación (ver por ejemplo: Sereno, 1999). Pero en el caso de las proteínas, el concepto de superfamilia como un grupo monofilético de proteínas que comparten un mismo plegamiento, está ampliamente arraigado en la literatura bioquímica. De hecho, su utilidad está lejos de toda duda y actualmente es posible identificar con certeza la superfamilia de proteínas a la que pertenece una proteína recién secuenciada[2]; e incluso, se puede estimar en alrededor de 1000 los principales plegamientos de proteínas que existen en la naturaleza, conocidos o por conocer, los cuales a su vez conforman probablemente alrededor de 5000 superfamilias/familias[3] de proteínas diferentes (Wolf y cols., 2000).

[2] Esto último suele ser valido para alrededor del 80% de las proteínas identificadas dentro del genoma de un organismo recien secuenciado. Sólo el 20% de las proteínas recien identificadas en general, parecen no mostrar similitud con ninguna de las proteínas previamente identificadas. A estas últimas proteínas se les denomina ORFans.

[3] En la literatura existe ambigüedad con respecto a los términos superfamilia, familia y subfamilia. La cifra indicada no toma en consideración los ORFans, para los cuales existe una gran incertidumbre en cuanto a su número total.

La distribución filogenética de las superfamilias más abundantes de proteínas (y que comprenden alrededor del 80% del universo de las proteínas), sugiere que estas existen desde antes del último ancestro común, lo cual implica que la historia evolutiva de una gran proporción de las superfamilias de proteínas es tan antigua como la misma historia evolutiva de los seres vivos. Esto último en realidad enriquece la utilidad del concepto de superfamilia, pero al mismo tiempo genera un conflicto muy grande desde el punto de vista taxonómico, porque la taxonomía de proteínas, a pesar de tener una historia tan vasta, sólo puede establecerse formalmente empleando jerarquías de un nivel taxonómico inferior al de superfamilia[4].

Es claro que en el universo de las proteínas, cada proteína individual podría ser considerada como equivalente a un individuo (Henikoff y cols., 1997). Siguiendo con la analogía, los conjuntos de proteínas ortólogas[5] harían el equivalente al papel de las especies y conforman por tanto, el único grupo que de manera natural puede definirse sin ambigüedades (al igual que lo que ocurre en el mundo de los seres vivos con el concepto de especie)[6]. Puesto que en este trabajo se han definido a estos conjuntos naturales de proteínas ortólogas como subfamilias para estar en concordancia con la nomenclatura ya establecida en la literatura, es claro que cualquier categoría taxonómica propuesta dentro de una superfamilia, deberá ubicarse entre subfamilia y superfamilia.

Es claro que los conjuntos de subfamilias más relacionados entre sí pueden agruparse en familias, pero la pregunta ahora es si ¿una sola categoría taxonómica intermedia es suficiente para describir la historia evolutiva de las proteínas? A primera vista y en comparación con la taxonomía de los seres vivos, uno pensaría que la respuesta es no; luego entonces, ¿cuál es un número adecuado de categorías taxonómicas necesarias para ilustrar la riqueza evolutiva de las proteínas? Usar el mismo número de categorías taxonómicas que las empleadas para clasificar a los seres vivos es muy probablemente un número excesivo. La taxonomía de los seres vivos es en realidad muy detallada, y no parece práctico tratar de representar con el mismo grado de detalle la evolución de cada superfamilia de proteínas. Por otra parte, la taxonomía de proteínas tiene particularidades que hacen impráctica una clasificación tan detallada: los fenómenos de transferencia horizontal y la fusión de genes por ejemplo, complican cualquier esfuerzo tendente a establecer una clasificación con alto grado de detalle. Además, es claro que cualquier categoría taxonómica propuesta debe tener un significado biológico que le confiera una razón de ser. Las

---

[4] Esto último es deseable, ya que lo más conveniente para construir una taxonomía filogenética de las proteínas es construir un esquema de clasificación acorde a las jerarquías taxonómicas previamente establecidas para los seres vivos.

[5] Recuérdese que son los conjuntos de proteínas generados por procesos de especiación.

[6] De manera estricta, el concepto de especie tiene también sus bemoles y no siempre es tan fácil de aplicar como uno quisiera, pero la discusión acerca de este concepto está más alla del enfoque de este trabajo.

dos categorías propuestas en este trabajo cumplen a nuestra manera de ver con ésta premisa: familia y macrofamilia tienen utilidad por tener un significado biológico preciso: una familia comprende un grupo monofilético[7] de subfamilias que conforman un conjunto de proteínas parálogas[8] relacionadas entre sí; por otra parte, el concepto de macrofamilia tendrá una utilidad real si se cumple nuestra hipótesis de que representa a grupos monofiléticos de familias de proteínas que están presentes desde el último ancestro común. Esto implica también otro supuesto, y es la idea de que todos los principales tipos de plegamientos existentes actualmente, están presentes desde el último ancestro común. Considerando la distribución de las proteínas dentro de las diferentes superfamilias hasta ahora descritas, este esquema parece ser válido para el 80% del universo de las proteínas actuales.

En este mismo esquema se considera la formación de un nuevo tipo de plegamiento dentro de los organismos modernos, como un evento excepcional, pero no imposible. La existencia de un número cada vez mayor de proteínas sin similitud con el resto de las proteínas (ORFans), sugiere que aún operan mecanismos que permiten la formación *de novo* de proteínas, sin requerir por tanto de una proteína homóloga preexistente. De hecho, hasta nuestro conocimiento, no existe ninguna razón para pensar que las fuerzas evolutivas que dirigieron la formación y el origen de las primeras proteínas, han dejado de operar en los organismos actuales[9].

De esta forma, la taxonomía propuesta en este trabajo se aplica directamente, sin ninguna dificultad, en las principales superfamilias de proteínas, las cuales comprenden aproximadamente el 80% de todas las proteínas actuales, descritas o por describir. Todas estas superfamilias poseen un origen muy antiguo y se formaron por duplicación a partir de una proteína ancestral. Así, el último ancestro común a todos los seres vivos, poseía ya, al menos a una proteína perteneciente a cada una de las principales superfamilias de proteínas actuales.

Para las proteínas formadas *de novo* en los organismos modernos (ORFans), y que comprenden aproximadamente el 20% restante de las proteínas, el esquema taxonómico propuesto en este trabajo debe ser modificado, para aceptar la posibilidad de que pueden existir plegamientos únicos a niveles taxonómicos inferiores, como familias y subfamilias. Esto es, se

---

[7] Esto es, que presentan todos un mismo ancestro común. Se contrapone al término polifilético, en donde existen diferentes ancestros.

[8] Recuérdese que son las proteínas generadas por procesos de duplicación de genes.

[9] Recientemente Johannes Söding y Andrei N. Lupas (2003), sugieren que las primeras estructuras proteicas que aparecieron, fueron péptidos que servían como cofactores para las ribozimas, durante el “RNA world”. Estos péptidos posteriormente se asociaron y fusionaron en cadenas polipeptídicas para formar las primeras estructuras independientes (no asociadas a un armazón de RNA), las cuales eventualmente dieron lugar a un nuevo tipo de macromoléculas: las proteínas plegadas.

sugiere restringir el uso de los términos macrofamilia y superfamilia exclusivamente para aquellos grupos de proteínas homólogas que existen desde el último ancestro común, De esta manera, el hecho de que un grupo de proteínas presente un plegamiento único no debe ser razón suficiente para conferirles el rango de macrofamilia o superfamilia, ya que para ello deben cumplir además con la condición de poseer un origen muy antiguo.

3.1 ¿Por qué utilizar el término macrofamilia?

En este momento, es adecuado efectuar un paréntesis y preguntarse ¿por qué a los conjuntos de familias de proteínas dentro de una superfamilia se les denominó macrofamilias? Para contestar esta pregunta, es necesario considerar lo siguiente: las únicas categorías taxonómicas aceptadas de manera general son: reino, phylum o división, clase, orden, familia, genero y especie. Como puede verse, de entrada los términos subfamilia y superfamilia no están considerados dentro de las categorías taxonómicas “oficiales”; sin embargo, es fácil intuir que a la subfamilia le corresponde un nivel taxonómico inferior y a superfamilia uno de mayor jerarquía. Pero, ¿qué otras categorías taxonómicas se han utilizado? Por el momento, sólo los paleontólogos han requerido el empleo de un número alto de categorías taxonómicas; Paul C. Sereno por ejemplo, en uno de sus trabajos clásicos sobre la taxonomía de los dinosaurios del orden Ornithischia (Sereno, 1986), utilizó las siguientes jerarquías:

Megaorden > Superorden > Orden > Parvorden > Nanorden > Hipoorden > Minorden > Suborden > Infraorden > Gigafamilia > Megafamilia > Grandfamilia > Hiperfamilia > Superfamilia > Familia.

Sin embargo, dentro de las quince subcategorías taxonómicas utilizadas por Sereno para los taxa orden y familia, no existe ninguna categoría intermedia entre superfamilia y familia. Por esta razón se tuvo que proponer el concepto de macrofamilia como una categoría taxonómica intermedia. De esta manera, se evitan conflictos con las categorías taxonómicas ya utilizadas por los paleontólogos. Hasta nuestro conocimiento, el término macrofamilia no se ha utilizado hasta ahora en un contexto biológico; únicamente los lingüistas han utilizado el concepto de macrofamilia (e.g., Barbujani y Pilastro, 1993). De esta manera, la propuesta de la jerarquía macrofamilia como un nivel intermedio entre superfamilia y familia, además de ofrecer las ventajas ya señaladas, no entra en conflicto con ninguna categoría taxonómica propuesta hasta ahora en biología.

3.2 Otros esquemas de taxonomía de proteínas.

Es importante señalar que la clasificación filogenética propuesta en este trabajo no entra en conflictos con las clasificaciones fenéticas[10] o estructurales de las proteínas previamente propuestas por otros grupos de investigación [e.g., CATH (Orengo y cols., 1997, 2002); SCOP (Lo

[10] Relativas al fenotipo, esto es, basadas en las características visibles.

Conte y cols., 2000, 2002); Dali/FSSP (Holm y Sander, 1997)]. Estas últimas se enfocan en general a las relaciones que existen entre diferentes superfamilias de proteínas. Así, el grupo de Janet M. Thornton, Christine A. Orengo y cols. propusieron el esquema de clasificación CATH (Orengo y cols., 1997, 2002), en el cual las diferentes superfamilias de proteínas se organizan en grupos homólogos de superfamilias (nivel H), las cuales probablemente comparten un ancestro común; éstos a su vez se agrupan de acuerdo con el plegamiento que presentan, esto es, la forma en que se conectan las diferentes estructuras secundarias, también denominada topología de una proteína (nivel T). Los diferentes plegamientos o topologías se organizan de acuerdo a su arquitectura (nivel A), es decir, la forma en que se orientan en un espacio tridimensional las diferentes estructuras secundarias que componen a una proteína. Finalmente, dependiendo de su contenido de estructuras secundarias α y/o β, se asigna la clase (nivel C) a la que pertenecen. La figura 1 integra bajo un mismo esquema la clasificación filogenética de proteínas propuesta en este trabajo, con la clasificación fenética o estructural CATH propuesta por Orengo y Thornton[11] (Orengo y cols., 1997, 2002).

Por lo que puede apreciarse en el figura 1, es claro que las aproximaciones fenética y filogenética de la taxonomía de proteínas operan a niveles diferentes dentro de la complejidad estructural de las proteínas, por lo cual estas dos aproximaciones, más que contraponerse se complementan. Por otra parte, el hecho de que el origen de las superfamilias de proteínas individuales puede remontarse hasta el último ancestro común, dificulta enormemente la utilización de criterios filogenéticos más alla del nivel de superfamilia, por lo que en los niveles de integración superior a superfamilia sólo los criterios estructurales o fenéticos pueden ser utilizados con certidumbre.

[11] Existen otros esquemas de clasificación que se construyen en forma manual (SCOP) o completamente automatizada (FSSP y DALI) y utilizan todos ellos los niveles de Clase (C), Topología (T) y el de superfamilias homólogas (H). Unicamente el nivel de arquitectura (A) es exclusivo de la base de datos CATH (Orengo y cols., 2002). Si bien existen diferencias en cuanto a las definiciones utilizadas por las diferentes bases de datos para construir sus esquemas de clasificación, la concordancia entre ellas es mayor al 50% (para una revisión ver: Hadley y Jones, 1999; Day y cols., 2003).

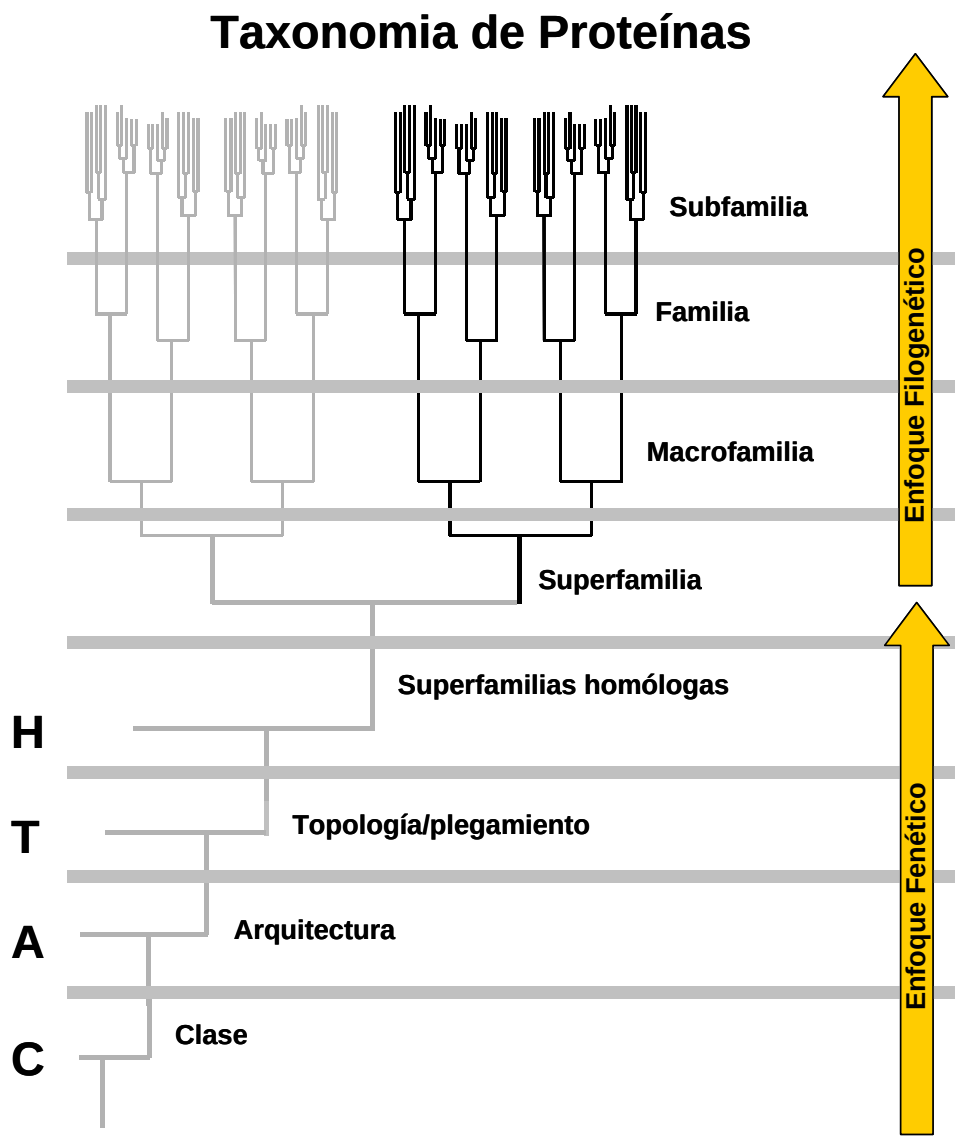


Figura 1. Esquema integral de la taxonomía de proteínas que incluye las categorías taxonómicas propuestas en este trabajo en base a criterios filogenéticos, y el esquema de clasificación de proteínas CATH, en base a criterios estructurales (ver texto para más detalles).

4. ¿Por qué no utilizar los métodos filogenéticos convencionales?

Es importante señalar que actualmente existen dos corrientes distintas para identificar grupos ortólogos. La utilizada en este trabajo se basa en la identificación de los mejores resultados de similitud en BLASTP (BLAST's reciprocal best hits) ya descrita previamente. La segunda

corriente se basa en la construcción de árboles filogenéticos y los resultados del análisis de bootstrap. Estos últimos métodos, además del inconveniente de tener que determinar límites de identidad arbitrarios para delimitar subfamilias y familias de proteínas, tienen también la limitante de no poder manejar el conjunto completo de secuencias de proteínas pertenecientes a la superfamilia de las MDR. Esto es, los métodos filogenéticos convencionales son muy demandantes de recursos de computo, por lo que habitualmente un análisis de este tipo maneja sólo secuencias representativas y no más de un centenar de secuencias por análisis. Así, hubiera resultado prácticamente imposible tratar de incluir la totalidad de las secuencias identificadas en este trabajo como pertenecientes a la superfamilia de las MDRs (583 secuencias): analizar un centenar de secuencias por los métodos tradicionales, implica tiempos de computo del orden de 2 semanas, y el tiempo de computo se incrementa en forma exponencial de acuerdo al número de secuencias empleadas. De esta forma, la identificación simultánea de las 3 macrofamilias, 8 familias y 49 subfamilias que conforman la superfamilia de las MDR, sólo fue posible mediante la comparación de secuencias usando los resultados de BLASTP. Una vez identificadas las macrofamilias y familias que conforman las MDR, fue relativamente sencillo conformar subconjuntos de las MDR que permitieran, mediante la construcción de árboles filogenéticos, validar por análisis de bootstrap, los grupos ya identificados previamente con base en los resultados de BLASTP. Aquí es importante señalar que este es uno de los pocos trabajos en los cuales los grupos ortólogos identificados mediante BLAST, se validan empleando simultáneamente métodos filogenéticos convencionales. Lo habitual es la presentación de resultados por un método u otro, sin ningún tipo de comparación.

Finalmente, es importante señalar que los resultados obtenidos con BLAST no son suficientes por sí solos para delimitar los grupos de proteínas ortólogas. Es necesario analizar en forma paralela, el patrón filogenético de las proteínas que conforman el grupo ortólogo a identificar. La ruptura en la concordancia filogenética dentro de un conjunto de proteínas es un fuerte indicador de la presencia de proteínas parálogas. De esta forma, sólo mediante la integración de los resultados obtenidos con ambos métodos es posible realizar predicciones razonables para delimitar los grupos ortólogos[12].

[12] No debe olvidarse que la comprobación experimental es la única forma para tener certeza sobre la función desarrollada por una proteína. Cualquier inferencia de la función de una proteína realizada por métodos computacionales, es sólo una predicción.

Perspectivas

Rompiendo Paradigmas: el redescubrimiento de las proteínas.

En este punto, es claro que la plasticidad y versatilidad de las proteínas es mucho mayor de lo que nos gustaría reconocer, y estamos actualmente experimentando una revolución sobre los conceptos tradicionales que tenemos de las proteínas. Así, por ejemplo, de los resultados de este trabajo, es claro que el paradigma: “una proteína, una función”, no es válido para una gran diversidad de proteínas: es común que éstas presenten, además de su función principal, dos o más funciones secundarias fisiológicamente relevantes, de tal forma que las proteínas multifuncionales parecen ser mucho más abundantes de lo que se pensó en un principio. Inclusive, se puede observar el caso de proteínas que desarrollen dos actividades catalíticas distintas, ambas con la misma eficiencia, de tal manera que resulte imposible definir cual es la función catalítica primaria y cual la secundaria.

Por otra parte, el desarrollo de técnicas nuevas para estudiar la estructura de proteínas en solución, como es el caso de la resonancia magnética nuclear, han dejado en claro que las proteínas en solución no son entes estáticos, como lo sugiere la cristalografía de rayos X. Así, una proteína en solución puede adoptar múltiples conformaciones (del Río Portilla, 2003).

Si esta última evidencia la conectamos con el hecho de que las proteínas son multifuncionales, no es difícil llegar a la idea de que las diferentes conformaciones de una misma proteína, tendrán diferente afinidad por sus ligandos y presentarán diferentes eficiencias catalíticas. De esta forma, una conformación determinada (probablemente la(s) más estable(s)), podría ser la responsable de la actividad catalítica primaria, mientras que otra conformación diferente (probablemente la menos estable), podría ser la responsable de una de las actividades secundarias de esa misma proteína. Estos conceptos ilustran un posible mecanismo de evolución de las proteínas en donde la selección natural, actuando diferencialmente sobre los diferentes confórmeros de una misma proteína, permite conservar y optimizar alguna(s) de las múltiples actividades desarrolladas por una misma proteína y explican un mecanismo a través del cual, una proteína dada puede adquirir diferentes funciones primarias. De hecho, esta última idea proporciona una base sólida a partir de la cual se pueden discutir los mecanismos de evolución de proteínas indicados en el artículo: Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily.

Toda esta información contribuye a replantear nuestros conceptos básicos sobre la estructura y la función de las proteínas y conforman, en nuestra opinión, un redescubrimiento de las proteínas.

Glosario

Confórmeros: en proteínas, se refiere a las configuraciones tridimensionales alternativas que puede adoptar una misma secuencia de aminoácidos. Los estudios de resonancia magnética nuclear para determinar la estructura de proteínas en solución han demostrado que las proteínas adoptan múltiples conformaciones o confórmeros.

Especiación: proceso a través del cual se generan nuevas especies, siempre a través de aislamiento reproductivo.

Fenético: relativo al fenotipo, esto es, las características visibles.

Homología: se refiere a un carácter o unidad biológica, cuya similitud (en diferentes especies), resulta de la divergencia a partir de un ancestro común. La homología carece de métrica y no debe confundirse con la identidad; la homología existe o no existe, pero sin ningún tipo de métrica.

Monofilético: se refiere a los caracteres o unidades biológicas que poseen un solo ancestro común.

ORFans: proteínas que no muestran similitud con ninguna de las proteínas previamente identificadas.

Ortólogo: uno de los dos principales tipos de homólogos. En biología molecular, se refiere a aquellos genes o proteínas que se originan a partir de un ancestro común por procesos de especiación, de tal forma que la historia de las proteínas ortólogas refleja la historia evolutiva de las especies (del griego *orto* = correcto).

Parálogo: uno de los dos principales tipos de homólogos. En biología molecular, se refiere a aquellos genes o proteínas que se originan por procesos de duplicación dentro de una misma especie, de tal forma que la historia de las proteínas duplicadas se desarrolla en forma paralela a partir de su origen (del griego *para* = junto a).

Polifilético: se refiere a los caracteres o unidades biológicas que poseen dos o más ancestros diferentes.

Taxa: Cada uno de los grupos o subdivisiones de la clasificación de los seres vivientes, que se ordena sistemáticamente según su jerarquía propia

Taxonomía: Ciencia que trata de los principios, métodos y fines de la clasificación. Se aplica en particular, dentro de la biología, para la ordenación jerarquizada y sistemática, con sus nombres, de las diferentes especies de seres vivos.

Referencias

Altamirano MM, Blackburn JM, Aguayo C, Fersht AR. (2000) Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* 403: 617-622.

Altschul SF, Madden TL, Schaffer 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.

Arnon R, Degli Esposti S, Zern MA. (1995). Molecular biological aspects of alcohol-induced liver disease. *Alcohol Clin Exp Res* 19: 247-256.

Atteia A, van Lis R, Mendoza-Hernández G, Henze K, Martin W, **Riveros-Rosas H**, González-Halphen D. (2003). *Bifunctional aldehyde/alcohol (ADHE) in chlorophyte algal mitochondria*. *Plant Mol Biol* 53: 175-188.

Bairoch A, Apweiler R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28: 45-48.

Baraona E, Julkunen R, Tannenbaum L, Lieber CS. (1986). Role of intestinal bacterial overgrowth in ethanol production and metabolism in rats. *Gastroenterol* 90: 103-110.

Barbujani G, Pilastro A. (1993). Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc Natl Acad Sci U S A* 90: 4670-4673.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. (2000). GenBank. *Nucleic Acids Res* 28: 15-18.

Björkhem I. (1972). On the role of alcohol dehydrogenase in ω-oxidation of fatty acids. *Eur J Biochem* 30: 441-451.

Chen R, Jeong SS. (2000). Functional prediction: identification of protein orthologs and paralogs. *Protein Sci* 9: 2344-2353.

Conway T, Ingram LO. (1989). Similarity of Escherichia coli propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from Zymomonas mobilis and Saccharomyces cerevisiae. *J Bacteriol* 171: 3754-3759.

Danielsson O, Atrain S, Luque T, Hjelmqvist L, Gonzàlez-Duarte R, Jörnvall H. (1994). Fundamental molecular differences between alcohol dehydrogenases classes. *Proc Natl Acad Sci USA* 91: 4980-4984.

Day R, Beck DA, Armen RS, Daggett V. (2003). A consensus view of fold space: combining SCOP, CATH, and the Dali Domain Dictionary. *Protein Sci* 12: 2150-2160.

Del Rio Portilla JF. (2003). Determinación de la estructura de proteínas por resonancia magnética nuclear. En: **Riveros-Rosas H**, Flores-Herrera O, Sosa-Peinado A, Vázquez-Contreras E (eds). Mensaje Bioquímico Vol. XXVII. Depto. Bioquímica, Fac. Medicina, UNAM. México. pp 65-83. (ISSN-0188-137X). http://bq.unam.mx/mensajebioquimico

Dixon M, Lutwak-Mann C. (1937). Aldehyde mutase. *Biochem J* 31: 1347.

Estonius M, Hoog JO, Danielsson O, Jornvall H. (1994) Residues specific for class III alcohol dehydrogenase. Site-directed mutagenesis of the human enzyme. *Biochemistry* 33: 15080-15085.

Fitch WM. (1970). Distinguishing homologous from analogous proteins. *Syst Zool* 19: 99-113.

Gerlt JA, Babbit PC. (2000). Can sequence determine function? *Genome Biol*, 1, REVIEWS0005.

Gerlt JA, Babbit PC. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70: 209-246.

Hadley C, Jones DT. (1999). A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. *Structure Fold Des* 7: 1099-1112.

Hedberg JJ, Stromberg P, Hoog JO. (1998). An attempt to transform class characteristics within the alcohol dehydrogenase family. *FEBS Lett* 436: 67-70.

Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, Hood L. (1997). Gene families: the taxonomy of protein paralogs and chimeras. *Science* 278: 609-614.

Hjelmqvist L, Shafqat J, Siddiqi AR, Jornvall H. (1995). Alcohol dehydrogenase of class III: consistent patterns of structural and functional conservation in relation to class I and other proteins. *FEBS Lett* 373: 212-216.

Holm L, Sander C. (1997). Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res* 25: 231-234.

Holmquist B, Vallee BL. (1991). Human liver class III alcohol and glutathione dependent formaldehyde dehydrogenase are the same enzyme. *Biochem Biophys Res Commun* 178: 1371-1377.

James LC, Tawfik DS. (2001). Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. *Protein Sci* 10: 2600-2607.

Jornvall H, von Bahr-Lindstrom H, Jeffery J. (1984). Extensive variations and basic features in the alcohol dehydrogenase-sorbitol dehydrogenase family. *Eur J Biochem* 140:17-23.

Jörnvall H, Höög J-O, von Bahr-Lindström H, Vallee BL. (1987). Mammalian alcohol dehydrogenases of separate classes: Intermediates between different enzymes and intraclass isozymes. *Proc Natl Acad Sci USA* 84: 2580-2584.

Jörnvall H, Danielsson O, Eklund H, Hjelmqvist L, Höög J-O, Parés X, Shafqat J. (1993). Enzyme and isozyme development within the medium-chain alcohol dehydrogenase family. *Adv Exp Med Biol* 328: 533-544.

Jörnvall H, Hoog JO, Persson B. (1999). SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett* 445: 261-264.

Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R. (2000). Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A* 97: 9925-9930.

Krebs HA, Perkins JR. (1970). The physiological role of liver alcohol dehydrogenase. *Biochem J* 118: 635-644.

Krozowski Z. (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *J Steroid Biochem Mol Biol* 51: 125-130.

Kumar S, Tamura K, Jakobsen IB, Nei M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Lara-Ortíz T, **Riveros-Rosas H**, Aguirre J. (2003). Reactive oxygen species generated by microbial NADPH oxidase NoxA regulate sexual development in *Aspergillus nidulans*. *Mol Microbiol* 50: 1241-1255.

Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J y cols. (2002). Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA). *Genome Res* 12: 493-502.

Leloir LF, Muñoz JM. (1938). Ethyl alcohol metabolism in animal tissues. *Biochem J* 32: 299.

Li L, Stoeckert CJ Jr, Roos DS. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178-2189.

Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C. (2000). SCOP: a structural classification of proteins database. *Nucleic Acids Res* 28: 257-259.

Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res* 30: 264-267.

Loeb EM (1943). Primitive intoxicants. *Q J Stud Alcohol* 4: 387.

Lutwak-Mann C. (1938). Alcohol dehydrogenase of animal tissues. *Biochem. J.* 32: 1364.

McManus IR, Contag AO, Olson RE (1960). Characterization of endogenous ethanol in the mammal. *Science* 131: 102.

Nebert DW. (1991). Proposed role of drug-metabolizing enzymes: regulation of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. Mol Endocrinol 5: 1203-1214.

Nordling E, Jornvall H, Persson B. (2002). Medium-chain dehydrogenases/reductases (MDR). *Eur J Biochem* 269: 4267-4276.

O'Brien PJ, Herschlag D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6: R91-R105.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. (1997). CATH - a hierarchic classification of protein domain structures. *Structure* 5: 1093-1108.

Orengo CA, Bray JE, Buchan DW, Harrison A, Lee D, Pearl FM *et al*. (2002). The CATH protein family database: a resource for structural and functional annotation of genomes. *Proteomics* 2: 11-21.

Page RD. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357-358.

Pearson WR, Lipman DJ. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-2448.

Pearson WR. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* 183: 63-98.

Pennisi E. (1999). Keeping genome databases clean and up to date. Science 286: 447-450.

Persson B, Krook M, Jornvall H. (1991). Characteristics of short-chain alcohol dehydrogenases and related enzymes. *Eur J Biochem* 200: 537-543.

Persson B, Zigler JSJ, Jornvall H. (1994). A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur J Biochem* 226: 15-22.

Pietruszko R. (1979). Nonethanol substrates of alcohol dehydrogenase. In Majchrowicz E, Noble EP, Eds. Biochemistry and Pharmacology of Ethanol, Vol. 1. New York: Plenum Press, pp. 87-106.

Popham RE. (1978). The social history of the tavern. In Israel Y, Glaser FB, Kalant H, Popham RE, Schmidt W, Smart RG, Eds. Research Advances in Alcohol and Drug Problems, Vol 4. New York: Plenum Press, 225-302.

Riveros-Rosas H, Julián-Sánchez A, Piña E. (1997). Enzymology of ethanol and acetaldehyde metabolism in mammals. *Arch. Med. Res.* 28 (4): 453-471.

Riveros-Rosas H, Zentella de Piña M, Guinzberg R, Saldaña-Balmori Y, Julián-Sánchez A, Saavedra-Molina A, Piña E. (1999). Antagonism between the metabolic responses induced by epinephrine and piroxicam on isolated rat hepatocytes. *Eur J Pharmacol* 380: 49-59.

Riveros-Rosas H, Julián-Sánchez A, Villalobos-Molina R, Pardo JP, Piña E. (2003). Diversity, taxonomy and evolution of medium-chain dehydrogenase/ reductase superfamily. *Eur J Biochem* 270: 3309-3334.

Scopes RK. (1983). An iron-activated alcohol dehydrogenase. *FEBS Lett* 156: 303-306.

Sereno PC. (1986). Phylogeny of the bird-hipped dinosaurs (order Ornithischia). *National Geographic Research* 2: 234-256.

Sereno PC. (1999). Definitions in phylogenetic taxonomy: Critique and rationale. *Systematic Biology* 48: 329-351.

Shafqat J, El-Ahmad M, Danielsson O, Martinez MC, Persson B, Pares X, Jornvall H. (1996). Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). *Proc Natl Acad Sci USA* 93: 5595-5599.

Soding J, Lupas AN. (2003). More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays* 25: 837-846.

Sun H-W, Plapp BV. (1992). Progressive sequence alignment and molecular evolution of the Zn-containing alcohol dehydrogenase family. *J Mol Evol* 34: 522-535.

Tao H, Cornish VW. (2002). Milestones in directed enzyme evolution . *Curr Opin Chem Biol* 6: 858-864.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.

Uotila L, Koivusalo M. (1989). Glutathione-dependent oxidoreductases: Formaldehyde dehydrogenase. In Dolphin D, Poulson R, Avramovic O, Eds. Coenzymes and Cofactors, glutathione: Chemical, biochemical and medical aspects. Vol III part A. New York: Wiley, 517-551.

Waller G, Theorell H, Sjövall J. (1965). Liver alcohol dehydrogenase as a 3-β-hydroxy-5-β-cholanic acid dehydrogenase. Arch Biochem Biophys 111: 671-684.

Williamson VM, Paquin CE. (1987). Homology of Saccharomyces cerevisiae ADH4 to an iron-activated alcohol dehydrogenase from Zymomonas mobilis. *Mol Gen Genet* 209: 374-381.

Wolf YI, Grishin NV, Koonin EV. (2000). Estimating the number of protein folds and families from complete genome data. *J Mol Biol* 299: 897-905.

Yang ZN, Bosron WF, Hurley TD. (1997). Structure of human chi chi alcohol dehydrogenase: a glutathione-dependent formaldehyde dehydrogenase. *J Mol Biol* 265: 330-343.

Yokoyama S, Harry DE. (1993). Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants. *Mol Biol Evol* 10:1215-1226.

Anexos

Correspondencia con la oficina editorial del European Journal of Biochemistry, previa a la publicación del artículo: Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily.

Table S1. Proteins that belong to MDR superfamily.
(Material suplementario del artículo: Diversity, taxonomy and evolution of ....)
[Incluida sólo en la versión electrónica]

Table S2. References for Tables 3-8.
(Material suplementario del artículo: Diversity, taxonomy and evolution of ....)
[Incluida sólo en la versión electrónica]

December 12, 2002

Professor Richard Perham
Editor-in-Chief
European Journal of Biochemistry

Dear Editor-in-Chief

Attached to this letter please find the article entitled: **DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASES/REDUCTASES SUPERFAMILY**, by **Héctor Riveros-Rosas, Adriana Julián-Sánchez, Rafael Villalobos-Molina, Juan Pablo Pardo, and Enrique Piña**, which I submit for publication to the ***European Journal of Biochemistry***. The manuscript includes two supplementary sections with additional references and a full list of proteins identified as members of medium-chain dehydrogenases/reductases superfamily.

Recently, Nordling *et al.* published in this journal (EJB) one paper about medium-chain dehydrogenases/reductases (MDR) [Eur. J. Biochem. (2002) 269: 4267-4276]. Our paper was submitted for publication almost at the same time that Nordling *et al.* paper, however, we make a mistake and we submit our paper first to the Journal of Molecular Biology on May 14. The editor of JMB, detained six months our paper. In the meantime, the paper of Nordling *et al.* was published. The paper that we submit today, take in consideration the recently published contribution of Naraling *et al.*, and we believe that we make a significative improvement to the work initiated by Nordling *et al.*

Dr. Hans Jornvall and his collaborators (including Dr. Erik Nordling and Dr. Bengt Persson), are the leader group on medium-chain dehydrogenases/reductases research. They could be excellent reviewers for our paper, however, please consider the possibility of a conflict of interest.

As requested in the general submission guidelines I suggest as referees:

Dr. Xavier Parés
Dept of Biochemistry and Molecular Biology
Universitat Autonoma de Barcelona
E-08193 Bellaterra, Spain
Phone: 93 581 3026
Fax: 93 581 1264
e-mail: xavier.pares@uab.es

Dr. Josep A. Biosca
Dept of Biochemistry and Molecular Biology
Universitat Autonoma de Barcelona
E-08193 Bellaterra, Spain
Phone: 93 581 3070
Fax: 93 581 1264
e-mail: josep.biosca@uab.es

Dr. Jaume Farrés
Dept of Biochemistry and Molecular Biology
Universitat Autonoma de Barcelona
E-08193 Bellaterra, Spain
Phone: 93 581 2557
Fax: 93 581 1264
e-mail: jaume.farres@uab.es

Yours sincerely,

Héctor Riveros-Rosas
Depto. Bioquímica, Fac. Medicina
Universidad Nacional Autónoma de México
Apdo. Postal 70159. Ciudad Universitaria
México D.F., 04510 MEXICO.

Tel +52 (55) 5622 - 0829;
FAX +52 (55) 5616 - 2419;
e-mail: hriveros@servidor.unam.mx

Héctor Riveros-Rosas

De: <ejb@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
Enviado: jueves, 12 de diciembre de 2002 18:27
Asunto: EJB Manuscript No. EJB-02-1375

EJB - The FEBS Journal

December 12th, 2002

Manuscript number: EJB-02-1375
Title: DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASES/REDUCTASES SUPERFAMILY
Authors: 1) Hector Riveros-Rosas 2) Adriana Julián-Sánchez 3) Rafael Villalobos-Molina 4) Juan Pablo Pardo 5) Enrique Piña

Dear Dr. Riveros-Rosas,

Thank you for submitting your manuscript to EJB.

As corresponding author, you will receive all future communications about this manuscript. Please remember to quote your manuscript number in all correspondence.

You can keep track of the progress of your manuscript through the Editorial system by periodically logging on to the EJB online submission website at http://ejb.manuscriptcentral.com.

The status of your manuscript will be displayed in the Author Centre. Please address any queries or comments to the Editorial Office.

We will advise you of the Editor's decision in due course.

Yours sincerely
Louise Sanders

Louise Sanders
Editorial Manager
European Journal of Biochemistry
98 Regent Street
Cambridge CB2 1DP, UK
Tel: +44 12223 369020
Fax: +44 1223 369090
Email: ejb@camfebs.co.uk

Héctor Riveros-Rosas

De: <ejb@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
CC: <Anna.tramontano@uniroma1.it>; <ejb@camfebs.co.uk>
Enviado: martes, 18 de febrero de 2003 19:42
Asunto: EJB manuscript EJB-02-1375

EJB - The FEBS Journal

18 Feb 03

Reference no.: EJB-02-1375

Title: DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASES/REDUCTASES SUPERFAMILY

Authors: 1) Hector Riveros-Rosas 2) Adriana Julián-Sánchez 3) Rafael Villalobos-Molina 4) Juan Pablo Pardo 5) Enrique Piña

Editor: Anna Tramontano

Dear Dr Riveros-Rosas

Thank you for submitting your paper for publication in EJB.

I regret to inform you that the Editor responsible for the evaluation of your paper has advised me that it cannot be accepted for publication in its present form. The referees acknowledge the large amount of work that has gone into collecting the data, but there are concerns about the novelty and depth of the analysis and the interest it engenders.

The referees' reports are appended below, and you may wish to consider them when preparing your paper for submission to another journal now or to our journal at a later stage in your research when you feel you can satisfy our referees' concerns.

I am very sorry to have to disappoint you.

Yours sincerely

Richard Perham
Chairman of the Editorial Board

European Journal of Biochemistry
98 Regent Street
Cambridge CB2 1DP, UK
Tel: 44 1223 369020
Fax: 44 1223 369090
Email: ejb@camfebs.co.uk

REFEREES' COMMENTS

Referee 1 Comments:
The authors describe the taxonomy of the medium-chain dehydrogenase/reductase (MDR) superfamily by sequence comparison of several families and subfamilies of proteins showing very different biological roles and biochemical specificities. The topic is certainly interesting in an attempt to rationalize the distribution of the MDR enzymes among distant phyla and the evolutionary correlations that can be drawn from such a study.

As a general comment, this report does not improve the depth of the previous study by Nordling et al. (very often cited by the authors and published in the same Journal) who already identified and clustered the MDR superfamily in eight protein families with clear and convincing informatics approaches. It is interesting that many additional sequences could be found to increase the confidence of the comparative analysis but they do not add that much further information. Moreover the authors claim that public databases that use tools such as BLASTP or FASTA show "several inconsistencies" in following different criteria for the classification and grouping of the proteins. But they seem to use the same tools or, if not so, it is very hard to understand the alternative approaches (such as the cut-off) of their analysis. The Materials and Methods section should be less cryptic on this point
>From the evolutionary point of view, the data presented are insufficient to deduce theories about enzymogenesis routes. It seems that throughout the manuscript the authors base their evolutionary theory from the paradigm "archeal-ancestor" which is a hazardous simplification. In fact, the "common progenote" theory is a delicate matter and still the focus point of hard debates among scientists. This perspective does not consider all the complex mechanisms of horizontal gene transfers between archaea and bacteria which renders very hard, if not impossible, any definitive assignment of ancestor proteins.
As an example, in the S. solfataricus genome, 13 genes encoding putative NAD dependent alcohol dehydrogenases have been identified. 12 of them show significant similarity with bacterial ADHs and hence are xenologs of thermophilic bacterial counterparts.
It is very interesting that some archaeal "benzyl" ADHs can be grouped close to "aryl" ADHs from fungi or plants but the authors do not provide clear differences both in the substrate specificities (what is the difference between "aryl" and "benzyl"? To the best of our knowledge there is no such strict and distinctive specificity) and biological roles.

Referee 2 Comments:
The paper by Riveros-Rosas et al is a huge compilation of all known members of the medium-chain dehydrogenase/reductase superfamily. The discussion focuses on the combination of

structure and function to judge a protein superfamily. Several papers have been published about this protein family with sequence comparisons and evolutionary conclusions. However, within this paper the functional aspect has been stressed.
It would have been nice to arrive at a more distinct final conclusion. The same is true of the summary, which is very weak as compared to the discussion. The only novelty, as judged from the summary, is the large number of proteins studied.
In the paper (and stressed in the conclusion) the authors group the MDRs into three divisions and 49 subfamilies. However, in the summary nothing is mentioned about these divisions. What is the advantage to group the MDRs within three divisions instead of eight families as proposed by Nordling et al. ?
Overall, the paper is well written and sound.

Specific comments
The summary has to be rewritten and be in line with the conclusion of the paper.
Pg 9, ln 17: (g subdivision), refer to tables for clearness.
Pg 9, ln 19: Table IX should read something else.
Pg 9, bottom paragraph: Their members contain Zn2 , probably they do but it has not been shown for all. Use the same nomenclature (Zn2 or zinc) for the ligands throughout the paper, cf pg 10.
Pg 11, paragraph The elicitor-. what does the end of the sentence mean?
Pg 14: The discussion about paralogous and orthologous is interesting but the example with plant ADH/class III ADH is maybe not the best because they can also be classified as orthologous. They can use different substrates but they also share a number of substrates.
Pg 18, last sentence, second paragraph: What enzymes does modern MDRs refer to?
Pg 19, first sentence conclusion. Several of the 49 subfamilies show overlapping activities. In that sense it is hard to agree that all 49 subfamilies have a different function. How is it with animal ADH, plant ADH, class III ADH and fungi ADH? Maybe that they have some different functions but they also possess same activity, e.g. all can metabolise ethanol. Furthermore, here it is also important what does it mean that a protein has a different function.
Figure 1. The color coding is not the best in such a complicated figure. For easier reading it would have been easier to split this figure into one figure and one table. An explanation to bars (and yellow dots in other figures) is requested.
Table 7. The number of protein coding genes given disagree in several cases with numbers found in the literature. Especially is this valid for the eukaryota where the values given are higher except for D. melanogaster. An explanation and/or references are requested.

Minor comments
The abbreviation list covers only a small part of all abbreviations used. At least ten more protein abbreviations are used throughout the manuscript. COG and NCBI could be included as well.

Summary - the level of examples are very different, e.g. ethanol metabolism and lens proteins, which should be examples of functions
Pg 3, ln 7 and pg 19, ln 2: the short-chain ADH should preferably read the short chain dehydrogenases/reductases
Use ADH for alcohol dehydrogenase throughout the paper
Pg 5, ln 18: 21 sequences should preferably read 21 proteins
Use the full name of an abbreviation the first time the name appear in the text, e.g. COG and several proteins
Pg , ln 19: use u.c. s for S-hydroxymethyl-glutathione and S-hydroxymethyl-mycothiol
Pg 24: Use l.c. for names in ref 42 and 48.
Use same type of nomenclature for tables in figure legends as in running text.
Figure 5 looks like a table

Referee 3 Comments:
The manuscript by Riveros-Rosas et al. represents a comprehensive, structural and functional, in silico analysis of the medium-chain dehydrogenase/reductase (MDR) superfamily (583 proteins). The authors use extensive database mining and the BLASTP program in an iterative manner in order to identify all the members of the MDR superfamily. Thus, they classify MDR proteins based on sequence as well as functional similarities, construct unrooted phylogenetic trees, and compare their results with previous partial classification attempts. The authors also infer that eukaryota inherited MDRs from bacteria, since MDR subfamilies in eukaryota are more closely related to their counterparts in bacteria than in archaea. As an ancestral activity, they propose formaldehyde metabolism and broad alcohol fermentation.

This work has been carefully performed and represents the most comprehensive and up-to-date study on this superfamily. In my opinion, it would be very useful to investigators in the field as well as a model to those involved in studies on protein function, taxonomy and evolution.

A recent publication by Nordling et al. (2002), which is extensively cited in the present manuscript, included only 116 sequences from 6 completed genomes (5 eukaryotic plus Escherichia coli). In the present work, the entire set of MDR members has been studied and a different method of grouping protein sequences has been used.

The summary should be rewritten in order to include all the relevant information provided in the manuscript. For instance, it should contain the definitions of protein subfamily and cluster of subfamilies or division, according to authors. A good idea would be to merge the existing information from the summary and conclusions section. Some colloquial expressions, such as "and so on" or "Jornvall's group" should be avoided.

In the Introduction, the starting paragraph on non-MDR alcohol dehydrogenases, from the Fe-dependent and short-chain dehydrogenase/reductase superfamilies, could be omitted.

Nomenclature referring to "cluster of protein subfamilies", "division", and "family" results a bit confusing for non-initiated readers. In page 8, the authors state that at the "highest level of integration" they identify 3 divisions, while at "lower levels of integration" they identify 6-8 families. The authors should be more precise in what they mean by high and low levels of integration. In addition, in some places in the text and the supplementary material, Roman numerals which are used for divisions are also used to designate families (i.e., Family I and Family II, in p. 10; Family III, in p. 18). I suggest to keep Roman numerals for divisions and use specific names for families.

In page 8, the bottom paragraph, "In this way,.." sounds repetitive when compared to previous paragraphs.

Concerning Division I, is ScXDH (xylitol dehydrogenase) considered a member of this division? Why is this enzyme not mentioned anywhere in the text? Within the ADH Family, most sequences bind NAD(H) as stated in the text, but ADH4 (now renamed ADH8) from Rana perezi is an exception since it prefers NADP(H) (Peralba et al., 1999).

The end of the first paragraph in p. 19, "On the other..substituted." is not easy to understand.

There are a few minor points. Some of them deserve further explanation in a manuscript intended for a journal with a broad readership:

- Some of the abbreviations should be defined.
- COG should be defined the first time that it appears in the text.
- The meaning of E-value should be explained.
- Many references include JID numbers that should be deleted.
- Table 7 is not cited in the text.

- In Fig. 1, divisions in roman numerals should be included.
- In Fig. 3, if BDH subfamily refers to fungi BDH (2,3-butanediol dehydrogenase), as stated in Table 1, p. 36), then they should be brown coloured. Add label for CCAR and related proteins.
- In legends of Figs. 1, 2, and 3, the meaning of separating yellow and grey pins and bars should be explained in terms of boundaries for each division or subfamily.

A few misspellings should be corrected to the suggested form in several places in the text:
- In the title, "medium-chain dehydrogenases/reductases" should be singular.
- Classical (p. 3, line 9 from top; p. 10, line 1 from top).

- Leukotriene
- P. 3, line 3 from top: "exhibit". Line 10: "alcohol dehydrogenase families". Line 17: "proteins".
- P. 5, line 4 from top: "dendrograms".
- P. 6, line 4 from bottom: "agreement".
- P. 7, line 3 from top: reference 23 should be cited after authors' names. Line 4 from bottom: "allows".
- P. 8, line 15 from top: "divisions"
- P. 9, line 11 from top: "appear". Line 14: "possesses".
- P. 11, line 10 from top: "shows".
- P. 12, line 13 from bottom: "these".
- P. 14, line 8 from bottom: "hydroxyl".
- P. 15, line 3 from top: "preexisting".
- P. 16, line 13 from bottom: please use capital "S" for sulfur atom of derivative compounds of glutathione and mycothiol. Line 5 from bottom: "another". Line 3: "contain".
- P. 18, line 8 from bottom: "theoretical".
- Figure legends: "bottom", "COG families", "roman numerals". "Identity".

Héctor Riveros-Rosas

De: "Hector Riveros-Rosas" <hriveros@servidor.unam.mx>
Para: <ejb@camfebs.co.uk>; <Anna.tramontano@uniroma1.it>
CC: <epgarza@servidor.unam.mx>; "Juan Pablo Pardo (Labo4)" <pardov@laguna.fmedic.unam.mx>; "rafael villalobos-molina" <rvmolina@hotmail.com>; <adh@laguna.fmedic.unam.mx>
Enviado: martes, 01 de abril de 2003 16:05
Adjuntar: Riveros_Reply Referee 3.pdf; Riveros_Reply Referee 2.pdf; Riveros_Reply Referee 1.pdf; Riveros_suppl references_revised.pdf; Riveros_suppl_MDR proteins_revised.pdf; Riveros-paper_MDR-superfamily_revised.pdf; Riveros-paper_MDR-superfamily_referees only.pdf
Asunto: Re: EJB manuscript EJB-02-1375

April 1, 2003
Reference no.: EJB-02-1375

Dr. Richard Perham
Chairman of the Editorial Board
European Journal of Biochemistry
98 Regent Street
Cambridge CB2 1DP, UK

Dear Dr. Perham
We write you regarding our manuscript (EJB-02-1375) entitled: Diversity, Taxonomy and Evolution of Medium-Chain Dehydrogenase/Reductase Superfamily, by Héctor Riveros-Rosas, Adriana Julián-Sánchez, Rafael Villalobos-Molina, Juan Pablo Pardo, and Enrique Piña.

We sincerely appreciate the referees comments on our manuscript, and we feel pleased by their appreciation as a "well written and sound" paper (referee 2), that "represent the most comprehensive and up-to-date study on this superfamily. In my opinion, it would be very useful to investigators in the field as well as a model to those involved in studies on protein function, taxonomy and evolution" (referee 3), although we do not agree with several of the comments from referee 1.

Attached to this letter you will find a reply to each one of the referees, as well as our modified manuscript. We acknowledge their careful review and we took in consideration all their suggestion to improve our manuscript. The main changes to our manuscript are the following: we wrote a new summary, incorporating the main conclusions of the paper, and deleted the Conclusions section to avoid information duplicity. However, due to the complex and extensive analysis of MDR superfamily in our manuscript, the number of words in the summary exceeds by 35% the allowed word number. We will appreciate if you accept our summary because the main conclusions of our manuscript are listed only in the Summary section.

In reply to referees request, we improved the Materials and Methods section, and included bootstrap analysis results with conventional phylogenetic methods (neighbour-joining, UPGMA and maximum parsimony) to validate the proposed taxonomic categories (subfamilies, families and macrofamilies). Also, we included within the Discussion section, two new topics: 1) mechanisms of evolution within MDR Superfamily, and 2) Taxonomy inside MDR Superfamily. Furthermore, the number of references in our manuscript increases because several of the references included inside the supplementary material where added directly to the main text as referees

requested, although other are new because two new topics were written.

We hope that the new version of our manuscript is now suitable for publication in EJB, and thank the reviewers and editor for their help to have a better written manuscript.

Yours sincerely

Héctor Riveros-Rosas

==========

Héctor Riveros-Rosas
Depto. Bioquimica, Fac. Medicina
Universidad Nacional Autonoma de Mexico
Apdo. Postal 70-159. Ciudad Universitaria
Mexico D.F., 04510 MEXICO

e-mail: hriveros@servidor.unam.mx
FAX (+52 55) 56 16 24 19
TEL (+52 55) 56 22 08 29

==========

Reply to referees.

Referee 1 Comments:

The authors describe the taxonomy of the medium-chain dehydrogenase/reductase (MDR) superfamily by sequence comparison of several families and subfamilies of proteins showing very different biological roles and biochemical specificities. The topic is certainly interesting in an attempt to rationalize the distribution of the MDR enzymes among distant phyla and the evolutionary correlations that can be drawn from such a study.

Authors comments: we agree.

Referee 1 Comments:

As a general comment, this report does not improve the depth of the previous study by Nordling et al. (very often cited by the authors and published in the same Journal) who already identified and clustered the MDR superfamily in eight protein families with clear and convincing informatics approaches. It is interesting that many additional sequences could be found to increase the confidence of the comparative analysis but they do not add that much further information.

Authors answer: We do not agree. To our knowledge, this is the first time that a clear and solid protein taxonomy inside the MDR superfamily is provided, including three different taxonomical categories (subfamily, family and macrofamily[1]). Nordling et al. only identified families, and did not make any major effort to compare the eight families with the taxonomical categories quoted in other protein databases. We performed this comparison, and included the eight families proposed by Nordling et al.

By other side, we do not repeat the procedure used to construct the phylogenetic tree by Nordling et al. Instead, we used an alternative and different non distance-based method that let us to sort the MDR sequences into non-overlapping clusters (subfamilies). More important, the granularity of this clustering is determined by the data, not by a user-supplied data-dependent cut-off value. This last procedure is advantageous over the distance-based clustering methods used by Nordling et al., and it has been tested and validated by Krause and Vingron [Krause et al., (1998) Bioinformatics 14: 430; Krause et al., (2000) Nucleic Acids Res 28: 270; Krause et al., (2002) Nucleic Acids Res 30: 299). Krause and Vingron showed that this method is highly conservative, since the probability to obtain a false positive is extremely low.

---

[1] On the original manuscript we used the term division instead of macrofamily, but in order to give a coherent protein taxonomy, we prefer the term macrofamily because division possess on *sensu stricto*, an higher taxonomic rank than superfamily. However, although macrofamily is not a standardized taxonomic rank, do not exist a taxonomic rank located between family and superfamily, and since the protein superfamily concept is well established, all the protein taxonomic ranks inside a superfamily, must ideally possess a lower taxonomic rank than superfamily. To our knowledge the term macrofamily has been used before only in a linguistic context, but in a biological context, the classic paper of Paul C Sereno (Natl Geograph Res (1986) 2:234-256), used the next ranks between order and family: order > parvorder > nanorder > hyporder > monorder > suborder > infraorder > gigafamily > megafamily > grandfamily > hyperfamily > superfamily > family. Because in biology there is not intermediate rank between superfamily and family, we propose to use macrofamily, as a suitable alternative.

In addition, the use of character-based methods, like parsimony (this last used by Nordling et al.), and maximum likelihood, requires an excessive computer time in comparison to the distance-based methods, precluding its use with the complete set of MDR. Indeed, this was probably the main reason for the use of a limited set of MDR sequences by Nordling et al.

Our analysis included all known members of the MDR superfamily (583 protein sequences vs 116 protein sequences from six complete genomes selected by Nordling et al.). The use of the complete set of MDR proteins, together with the above described procedure, allow us to identify 49 subfamilies inside the MDR superfamily. Nordling et al., analysis cannot undertake this subfamily identification, since the procedure used by Nordling et al., is limited by the computer time required to perform the analysis. Therefore, Nordling et al., only included a limited number of protein sequences, and this small number makes impossible the identification of the subfamilies. Similarly, the use of the classic phylogenetic methods (neighbour joining, UPGMA, minimum evolution, maximum parsimony), did not allow the use of the complete set of MDR protein members, and thus, the utilization of the whole information to discover the fine details in the taxonomy of MDR superfamily.

Finally, we found several mistakes in Nordling et al., paper that introduce confusion in the bioinformatics field. First, the enoyl reductase (ER) family (COG3321) contains four related subfamilies comprising multifunctional polypeptides (fatty acid synthases and polyketide synthases) with a MDR domain containing enoyl reductase activity. Nordling et al., named inappropriately this family as acyl-CoA reductase (ACR), and says: “ACRs form a family that contributes one domain of the fatty acid synthases and erythronolide synthases”. However, in this case, Nordling et al., are referring to the enoyl-ACP reductase domain contained in the multidomain fatty acid synthase from animals, or the enoyl-ACP reductase domain from the iterative polyketide synthase in fungi. The enzyme acyl-CoA reductase is absent in fatty acid synthase, a multidomain enzyme that uses ACP (acyl carrier protein) as carrier for intermediates, not coenzyme A. The acyl-CoA reductases usually are membrane-bound enzymes involved in the biosynthesis of fatty alcohols and waxes, and are clearly different to the enoyl reductases that form a family inside the MDR superfamily. This problem of basic biochemistry was unnoticed by the referees.

Second, on page 4267, at the end of the first paragraph of Nordling et al., paper, they state that “All MDR enzymes utilize NAD(H) or NADP(H) as cofactor”. However, several protein subfamilies of MDR are not oxidoreductases, and do not use NAD(H) or NADP(H) as cofactor to perform their function: for example, hydroxynitrile lyase (HNL, division II) from plants, or alginate lyase (AL, division III) from proteobacteria. We show that through evolution, a great number of new functions were acquired by MDRs.In addition, other slips associated to the incomplete description of protein families on Nordling et al., paper, are pointed out in our manuscript.

Referee 1 Comments:

Moreover the authors claim that public databases that use tools such as BLASTP or FASTA show "several inconsistencies" in following different criteria for the classification and grouping of the proteins. But they seem to use the same tools or, if not so, it is very hard to understand the alternative approaches (such as the cut-off) of their analysis. The Materials and Methods section should be less cryptic on this point

Authors answer: Two different concepts are mixed in the referee’s comment. Certainly that BLASTP and FASTA are used frequently to identify new members of a protein superfamily, but not all protein databases use BLASTP or FASTA to classify proteins inside a protein superfamily.

For example, *Pfam* does not try to identify families or subfamilies inside the MDR superfamily; PROSITE uses motifs to identify two protein families inside MDR superfamily; PIR uses distance-based criteria to identify 119 families inside MDR; CATH uses structural data to identify six superfamilies inside the MDR; COG uses phylogenetic criteria to identify six families; SYSTERS uses a non distance-based method to identify 80 families. All this information is mentioned on the table included inside the Figure 1. We cannot see the "consistence" among the different criteria used by the public databases to classify and grouping proteins.

Today, the classification of new protein sequences into one of the protein superfamilies described, is easy and does not represent any problem; however, the identification of taxonomic categories inside a protein superfamily is the problem.

We recognize that the material and methods section was extremely reduced, and did not mention that the taxonomic categories proposed (subfamilies, families and macrofamilies) were validated by bootstrap analysis with conventional phylogenetic methods, using both distance-based methods (neighbour-joining and UPGMA) and character-based methods (maximum parsimony). To perform the phylogenetic analysis, only subsets of the complete MDR superfamily were used (due to the computer time limitations mentioned above). The subsets employed for phylogenetic analysis included protein sequences that belong to only one kingdom (archaea, bacteria, animals, plants and fungi). Later, subsets of proteins that belong to each one of the proposed macrofamilies or families, were used to validate by bootstrap analyses, the proposed protein subfamilies.

To illustrate this phylogenetic analysis, we included an additional figure (5) that shows the phylogenetic tree constructed with the protein sequences from the macrofamily II inside the MDR superfamily. Additional phylogenetic trees (not showed) were constructed with the protein sequences that belong to macrofamily I and III, and to each of the kingdoms to which the MDR proteins (archaea, bacteria, fungi, animals and plants) belong.

Referee 1 Comments:

From the evolutionary point of view, the data presented are insufficient to deduce theories about enzymogenesis routes.

Authors answer: We do not agree. Currently, two different evolutionary scenarios are envisioned. New catalytic functions of enzymes can evolve by: 1) changing the chemistry of catalysis, while retaining the binding capacity for a common ligand; or 2) retaining the chemistry of catalysis, while changing the substrate specificity. The data presented in our manuscript confirm new perspectives on protein evolution, because in addition to the above-mentioned mechanism of enzyme evolution, we found examples of preexisting enzymes that can be recruited to conform novel pathways where the proteins acquire new activities without conserving the same binding capacity or chemistry of catalysis.

We can mention one interesting example to support this third possibility. A divergent plant ADH with activity as acetone cyanohydrin lyase (P93243) has been described in flax (*Linum usitatissimum*) (Xu et al., (1988) Arch Biochem Biophys 263: 256; Trummler et al., (1997) J Biol Chem 272: 4770; Trummler et al., (1998) Plant Sci 139: 19; Breithaupt et al., (1999) J Mol Catal B-Enzymatic 6: 315).. This protein conforms a novel class of hydroxynitrile lyases (HNLs), and its amino acid sequence shows no overall homology to any cloned HNLs. Indeed, HNLs from plants form a heterogenous group of proteins differing in molecular mass, quaternary structure, presence or absence of flavin adenin dinucleotide, as well as glycosylation. They have convergently evolved from FAD-dependent oxidoreductases, α/β hydrolases, and MDRs (Dreveny et al., (2001) Structure 9:

803). Interestingly, HNL from flax (P93243), is a zinc-containing protein and conserve all amino acid residues important for structural integrity or coordinating zinc; however, flax HNL (P93243) neither displays ADH activity nor is inhibited by reagents interfering with zinc coordination. This information, together with the fact that flax HNL is more related to plant-, animal- and class III ADH, suggest that flax HNL (P93243) evolved late from a plant-/class III ADH, which was recruited for cyanogenesis in plants, a recent secondary pathway used as a defence mechanism against herbivores (Vetter J (2000) Toxicon 38: 11). The existence of multiple phylogenetically independent HNLs in plants supports this proposal. Therefore, this novel activity inside the MDR superfamily was acquired without the conservation of the original binding capacity and the chemistry of catalysis. In conclusion, proteins exhibit an unrecognized huge plasticity.

Another different alternative mechanism for enzyme evolution, observed also inside the members from MDR superfamily correspond to the modular construction or gene fusion, where separate gene products join together and generate new genes containing two or more domains with novel activities (Thornton et al., (1999) J Mol Biol 293: 333; Todd et al., (2001) J Mol Biol 307: 1113). Examples of this modular construction inside the MDR superfamily are: bi-domain oxidoreductase (BDOR) involved in the biosynthesis of exopolysaccharides (Nakar & Gutnick (2001) Microbiology 147: 1937); bifunctional QOR in plants, with an amino-terminal domain related to the short-chain dehydrogenase/reductase superfamily (Babiychuk et al., (1995) J Biol Chem 270: 26224; Ichinose et al., (2000) Z Natursforsch 55: 44); fatty acid synthase (FAS), a multifunctional polypeptide with seven enzymatic domains from animals (Smith S (1994) FASEB J 8: 1248) or alveolata (protozoa) (Zhu et al., (2000) Mol Biochem Parasitol 105: 253); the modular polyketide synthase from bacteria (Smith S (1994) FASEB J 8: 1248), and the iterative polyketide synthase from fungi (Kennedy et al., (1999) Science 284: 1368; Hutchinson et al., (2000) Antonie Van Leeuwenhoek 78: 287) .All of them possess a modular architecture. In this sense, it is important to mention that oligomerization is not conserved among the members of the MDR superfamily. For example, monomers, homodimers, homotrimers, homotetramers and heterodimers, are present in this superfamily, and has been proposed that the degree of oligomerization might be involved with changes on the functional role developed by the proteins (Thornton et al., (1999) J Mol Biol 293: 333; Todd et al., (2001) J Mol Biol 307: 1113).

Taken together, we conclude that the deep-rooted statements “one enzyme, one function” or “one protein family, one function” are not accurate for many enzymes. Several secondary activities might exist in one protein, as the above-mentioned animal ADH or fungi ADH subfamilies (see the first topic from the discussion), and this can be the point of departure to gain new completely different functions. Indeed, we point out the fact that two different and equally efficient catalytic activities can be a feature of one single protein, as described for the LTD/PGR subfamily. This catalytic promiscuity has been recognized as a vital springboard from which new catalytic activities can emerge out of existing folds and active sites (O’Brien & Herschlag (1999) Chem Biol 6: R91; James & Tawfik (2001) Protein Sci 10: 2600).

Data presented in this paper reinforce the idea that a protein can gain or lose a function through a limited number of amino acid changes, and several of such examples recovered from natural protein evolution are provided. MDR belongs to the limited number of protein superfamilies that posses both different mechanisms of reaction and substrate specificity (Gerlt & Babbit (2001) Annu Rev Biochem 70: 209; Todd et al., (2001) J Mol Biol 307: 1113)[49;77]. Indeed, several laboratories (e.g., Altamirano et al., (2000) Nature 403: 617; Jurgens et al., (2000) PNAS 97: 9925; Tao & Cornish (2002) Curr Opin Chem Biol 6: 858), have mimicked evolution of paralog proteins *in vitro*, showing generation of new catalytic or binding properties by modifications of a preexisting

protein scaffold, and forget that Nature, through the evolution, made a lot of such successful experiments.

It is interesting to mention that several subfamilies inside the MDR superfamily evolved as nonenzyme homologs, i.e., they conform novel proteins that lost their catalytic activity. The ζ-crystallin/QOR is probably the most detailed example. This protein is expressed in a taxon specific way in the lens of phylogenetically distant guinea pig, camel, and Japanese tree frog (*Hyla japonica*) (Garland et al., (1991) Arch Biochem Biophys 285: 134; Rao & Zigler (1992) Biochem Biophys Acta 1117: 315; Gonzalez et al., (1995) Mol Biol Evol 12: 773; J Biol Chem (2001) J Biol Chem 276: 28134), and constitutes about 10% of total water-soluble proteins of the lens. The recruitment of ζ-crystallin as a lens protein in the three former groups has been shown to have occurred independently through the modification or acquisition of unique alternative lens-specific promoters, suggesting that ζ-crystallin has a function in the lens that confers selective advantage. In this sense, it has been proposed that oxidoreductases may have been recruited as enzyme-crystallins (ε–, ρ–, λ–, η–, π–, ζ–) because they markedly increase the levels of pyridine nucleotides in the lens, which might be beneficial since high levels of reduced nucleotides serve as near UV filters and make lens less susceptible to photo damage (Zigler & Rao (1991) FASEB J 5: 223; Jong et al., (1994) Prog Retinal Eye Res 13: 391). Other examples of nonenzymes into the MDR superfamily are: a) the nuclear receptor binding proteins (NRBP1) that function as transcription factor in yeast (Yamazoe et al., (1994) J Biol Chem 269: 15244; Owen & Zelent (2000) Cell Mol Life Sci 57: 809), or nuclear receptor co-operator in animals (Masuda et al., (1998) Gene 221: 225); b) the dinoflagellate nuclear associated protein (DINAP) that corresponds to the quantitatively major nuclear protein in the dinoflagellate *Crypthecodinium cohnii*, and although DINAP did not bind directly to DNA, it activated basal transcription activity (Bhaud et al., (1999) J Eukaryot Microbiol 46: 259; Guillebauldt et al., (2001) Protist 152: 127); and c) the membrane traffic protein (AST) in fungi (Chang & Fink (1995) J Cell Biol 128: 39).

On the other hand, the subcellular location is not conserved across the members of the MDR superfamily. Although the great majority are soluble cytoplasmic proteins, some of them are located in mitochondria (Fungi ADH), and nucleus (DINAP; NRBP1; class III ADH in animals), and other inclusive possess a membrane location (VAT-1, and probably BDOR), or has been recruited as structural proteins (ζ-crystallin/QOR).

All these examples serve as a cogent reminder that Nature is not restricted to chemically- or substrate- conserved strategies for divergent evolution; instead, divergent evolution is opportunistic and one active site architecture, can be used to develop mechanistically distinct catalytic or non-catalytic functions. In other words, inside one protein superfamily (e.g., MDR), the functional diversity is more complex than sequence diversity.

[Note: We included this last discussion in our paper].

Referee 1 Comments:

It seems that throughout the manuscript the authors base their evolutionary theory from the paradigm "archeal-ancestor" which is a hazardous simplification.

Authors answer: To our knowledge, the 'universal tree of life' generally accepted is constructed with small subunit ribosomal RNA, and locates archaea and eucarya as sister groups (e.g. Doolittle (99) Science 284: 2124). Also, it is generally accepted that eukaryotic cells are the symbiotic result of bacteria (the symbiont) and archaea (the host) . Inside this context, it is not clear to us, what is the

meaning of the paradigm “archeal-ancestor”. We do not know any generally accepted paper that suggests that the last universal common ancestor was an archaea.

Referee 1 Comments:
In fact, the "common progenote" theory is a delicate matter and still the focus point of hard debates among scientists.

Authors answer: Please observe that we avoided the use of the term “Progenote”. The current literature uses the term progenote in two different ways: It either signifies an organizational level that preceded cells with prokaryotic organization, or it is used to denote the last common ancestor of all extant life. Unfortunately, both definitions are potentially contradictory. Therefore, to avoid confusion, it has been suggested (Gogarten & Olendzenski. The Progenote. Encyclopedia of Molecular Biology, T. Creighton (ed). Wiley) that the last common ancestor of all extant life should be denoted as the universal ancestor or cenancestor, and the term progenote should be reserved to denote a hypothetical pre-prokaryotic stage in cellular evolution. Therefore, we use in the manuscript the term “last common ancestor” to avoid confusion.

To our knowledge, there are no papers questioning the existence of the last common ancestor. We agree that the metabolic capabilities of the last common ancestor are the focus of intense debates, but we believe that this does not preclude us to make a contribution to this debate. Indeed, our analysis shows for first time that fatty acid synthesis, a primary metabolic pathway, evolved late in comparison to other primary metabolic pathways. We believe that this sole finding, justifies the discussion above the ancestral activities inside MDR

Referee 1 Comments:
This perspective does not consider all the complex mechanisms of horizontal gene transfers between archaea and bacteria which renders very hard, if not impossible, any definitive assignment of ancestor proteins. As an example, in the S. solfataricus genome, 13 genes encoding putative NAD dependent alcohol dehydrogenases have been identified. 12 of them show significant similarity with bacterial ADHs and hence are xenologs of thermophilic bacterial counterparts.

Authors answer: We do not discuss all possible mechanisms of horizontal gene transfer (in our opinion, this is far away from the main objective from our paper). We only mentioned two clear-cut examples, where a phylogenetically discordant pattern was obtained (one for *Entamoeba histolytica* in eukarya, and one for *Halobacterium* sp. in archaea). However, even today, the most convincing methods to reveal lateral gene transfer are based on phylogenetic inference (Ragan MA (2001) FEMS Microbiol Lett 201: 187).

With respect to the ancestral functions inside MDR, it has been proposed that the protein families that are present in all domains of life can be used as an estimate for the genome content of the last universal common ancestor (e.g., Woese & Fox (1977) J Mol Evol 10: 1; Mushegian & Koonin (1996) PNAS USA 93: 10268; Kyrpides et al., (1999) J Mol Evol 49: 413). This idea was taken by us as a way to detect the possible ancestral subfamilies (and functions) inside the MDR. We accept that there is no guarantee that a contemporary gene found in bacteria, archaea and eukarya was present in their last common ancestor —it could have arisen more recently in one domain and spread to the others, but such scenario seems more appealing for dispensable genes (antibiotic resistance, biosynthetic capabilities not useful in many environments) than for genes encoding indispensable cellular functions (Doolittle WF (1999) Trends Cell Biol 9: M5). Since we looked for primary metabolic activities inside the MDR superfamily as probable ancestral activities, we are less prone to errors.

On other hand, we do not agree with referee 1 proposal that 12 of the "13 genes encoding putative NAD dependent alcohol dehydrogenase (in *Sulfolobus solfataricus* genome) ... are xenologs of thermophilic bacterial counterparts". Several lines of evidences are against this interpretation. First, the new figure 5 in our paper include 4 protein sequences from *S. solfataricus* showing a higher identity-similarity with other archeal protein sequences than to bacterial ones. The *S. solfataricus* protein sequences included inside figure 5 belong to both, the broad ADH subfamily and other non-characterized archeal cluster. If we consider all the 13 *S. solfataricus* proteins identified recently as members of the MDR, four belong to the broad ADH subfamily (ADH4, ADH5, ADH7, ADH10), other three belong to, the above mentioned, non-characterized archeal cluster related to broad ADH subfamily (ADH2, ADH8, and ADH13)[2]. Interestingly, this archeal non-characterized cluster shows an uncertain position between the macrofamily II and III, and is located at the base of phylogenetic trees inside macrofamily II (see new Figure 5 in our manuscript). Furthermore, these proteins shows a wide distribution inside archaea, and include both crenarchaeota (*S. solfataricus, S. tokadaii, Aeropyrum pernix, Pyrobaculum aerophilum*) and euryarchaeota (*Ferroplasma acidarmanus, Thermoplasma volcanium, T. acidophilum*). In contrast, the number of related proteins that belong to these clusters in bacteria is fewer, in spite of the fact that the number of bacterial proteins in the public databases surpasses the number of archeal proteins. These data suggest an ancient origin for these proteins, and it is against to the possibility of a horizontal gene transfer origin.

Second, the closer bacterial relatives to the above-mentioned *S. solfataricus* proteins belong to both, thermophilic and mesophilic bacteria. The massive gene transfer between archaea and thermophilic bacteria is well documented (e.g., Aravind et al., (1998) Trends Genet 14: 442; Nelson et al., (1999) Nature 399: 323), and it seems likely that bacterial hyperthermophily has evolved secondarily within moderately thermophilic bacteria by continuous acquisition of thermotolerance genes from preadapted hyperthermophiles, namely the archaea (Aravind et al., (1998) idem). However, the horizontal gene transfer between mesophilic bacteria and archaea is not well documented, probably because these organisms live on different environments, which preclude the possibilities of a free gene interchange. A noteworthy example is *Dehalococcoides ethenogenes*, a mesophilic anaerobic bacterium that possesses two archeal-type genes (Nesbo et al., (2001) Mol Biol Evol 18: 362), in this case their co-occurrence in the same environment (mesophilic [30^{o}C], anaerobic estuarine sediments), along with several methanogenic archaeal strains, opens the possibility to horizontal gene transfer.

We sincerely would like to know the evidence that referee 1 has to suggest that 12 of the 13 MDR proteins present in *S. solfataricus* are xenologs of thermophilic bacteria, i.e. 12 of the MDR proteins in *S. solfataricus* are the product of an horizontal transfer event whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog) —Koonin et al., (2001) Annu Rev Microbiol 55: 709—.

[Note: since the addition of new MDR proteins of *S. solfataricus* did not modify the conclusions of our manuscript, we included these proteins only in the supplementary table S1].

Referee 1 Comments:

It is very interesting that some archaeal "benzyl" ADHs can be grouped close to "aryl" ADHs from fungi or plants but the authors do not provide clear differences both in the substrate specificities

[2] ADH6 and ADH12 belong to an archaeal cluster related to CCAR subfamily (QOR family). ADH1, ADH3, ADH9 and ADH11 belong to several small non-characterized clusters of proteins within PDH family.

(what is the difference between "aryl" and "benzyl"? To the best of our knowledge there is no such strict and distinctive specificity) and biological roles.

Authors answer: In first place, to our knowledge there is no “benzyl” ADH in archaea. We did not mention the existence of these enzymes in any part of our paper. Second, it is not clear to us how the referee handled to group the (in our opinion non-existent) “archaeal benzyl ADHs” with the “aryl ADHs from fungi or plants”; archeal and bacterial proteins are included on Figure 2, and eukaryotic proteins are included on a different Figure —3—. Third, to our knowledge there is no aryl or benzyl ADH on plants. The only reported plant enzymes with the capacity to act as benzyl ADH, belong to the elicitor-inducible defence-related proteins (ELI3) subfamily. Their members are present only in eudicot plants, and show different, but related defence activities: cinnamyl alcohol dehydrogenase, benzyl ADH, or mannitol dehydrogenase. This subfamily, absent in fungi, evolved from CADH subfamily.

If we left besides the above imprecisions, on Figure 2 are indicated inside the macrofamily II, an aryl/alkyl ADH subfamily present in firmicutes (Gram-Positive bacteria), and a benzyl ADH (present in proteobacteria (α and γ subdivisions) and firmicutes). As indicated on Table 4 (formerly table 2 from our initial manuscript), the protein members from Aryl/Alkyl ADH: Firmicutes subfamily, belong to a highly conserved gene cluster encoding haloalkane metabolism on the plasmid PrtI1 (Poelarends et al., (2000) J Bacteriol 182: 2191 and 2725). The protein members from Benzyl ADH shows affinity for a wide range of (substituted) aromatic alcohols, but are not capable of oxidizing aliphatic alcohols (e.g., MacKintosh & Fewson (1988) Biochem J 255: 653 and 743; Shaw & Harayama (1990) Eur J Biochem 191: 705 —for additional references, see supplementary references in our manuscript—). From the above description, is clear that both protein subfamilies belong to different metabolic pathways. They share the capacity to oxidize aromatic alcohols, although do not share the capacity to oxidize alkyl alcohols. Even if the above-mentioned two subfamilies possess the same catalytic activity, they can be paralogous among them. See below:

Paralogous proteins not necessarily posses different functions, since by definition, two proteins are said to be paralogous if they are derived from a duplication event, but orthologous if they are derived from a speciation event (e.g., Li W-H (1997) Molecular evolution. Sinauer; Patthy L (1999) Protein evolution. Blackwell Science; Page RDM & Holmes EC (1999) Molecular evolution: A phylogenetic approach. Blackwell Science; Graur D & Li W-H (1999) Fundamentals of Molecular Evolution. Second edition. Sinauer; Nei M & Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press; Sonnhammer & Koonin (2002) Trends Genet 18: 619). Therefore, after a duplication event, initially, the two resulting proteins posses identical properties, and only after evolution they could acquire different functions. A change in the functional role of a protein, not necessarily involve a change in substrates or mechanism of reaction. The recruitment of a duplicated protein into a different metabolic pathway, a different physiological role, or even a change in the spatiotemporal pattern of expression, expressing a protein in novel tissues and/or developmental stages (True & Carroll (2002) Annu Rev Cell Dev Biol 18: 53), could be a enough evolutive reason to conserve the duplicated protein, conforming in consequence, a novel paralogous protein with a different functional role. This last explanation is obligated since some outstanding papers provide inexact definitions (e.g., Altamirano et al., (2000) Nature 403: 617; Gerlt & Babbit (2000) Genome Biol 1: REVIEWS00005; Gerlt & Babbit (2001) Annu Rev Biochem 70: 209).

In order to clarify the above-mentioned concepts in our manuscript, we included this last discussion in our paper.

Referee 2 Comments:

The paper by Riveros-Rosas et al is a huge compilation of all known members of the medium-chain dehydrogenase/reductase superfamily. The discussion focuses on the combination of structure and function to judge a protein superfamily. Several papers have been published about this protein family with sequence comparisons and evolutionary conclusions. However, within this paper the functional aspect has been stressed. It would have been nice to arrive at a more distinct final conclusion. The same is true of the summary, which is very weak as compared to the discussion. The only novelty, as judged from the summary, is the large number of proteins studied.

Authors answer: In first place, thank you for the careful revision of our manuscript. We take in consideration the above comments and rewrite completely the summary. In addition, we included two new sections to the discussion in order to arrive to more distinct conclusions. One concern to the mechanisms of enzymogenesis inside the MDR superfamily, and the other concern to the protein taxonomy inside MDR superfamily. We hope, with this new section, to satisfy the inquiry of referee 2.

Referee 2 Comments:

In the paper (and stressed in the conclusion) the authors group the MDRs into three divisions and 49 subfamilies. However, in the summary nothing is mentioned about these divisions. What is the advantage to group the MDRs within three divisions instead of eight families as proposed by Nordling et al. ?

Authors answer: We modified the summary and included a new section into the discussion in order to propose a logic sketch to classify the protein members of the MDR superfamily into three different taxonomic categories.

On the original manuscript we used the term division instead of macrofamily, but in order to give a coherent protein taxonomy, we prefer the term macrofamily because division possess on *sensu stricto*, a higher taxonomic rank than superfamily. However, although macrofamily is not a standardized taxonomic rank, do not exist a taxonomic rank located between family and superfamily, and since the protein superfamily concept is well established, all the protein taxonomic ranks inside a superfamily, must ideally possess a lower taxonomic rank than superfamily. To our knowledge the term macrofamily has been used before only in a linguistic context, but in a biological context, the classic paper of Paul C Sereno (Natl Geograph Res (1986) 2:234-256), used the next ranks between order and family: order > parvorder > nanorder > hyporder > monorder > suborder > infraorder > gigafamily > megafamily > grandfamily > hyperfamily > superfamily > family. Because in biology there is not intermediate rank between superfamily and family, we propose to use macrofamily, as a suitable alternative.

The use of the complete set of MDR proteins, together with the criteria and procedures described under Results, allow us to identify inside the MDR superfamily, 49 subfamilies, and two additional taxonomic levels containing eight families and three macrofamilies (formerly divisions). From these three taxonomic levels (subfamily, family and macrofamily), only the subfamily level, as defined by us, comprises a natural unit that can be used to sort the protein members of a protein superfamily, with clear-cut rules. Thus, each subfamily encloses a set of ideally orthologous proteins that perform the same function, and delineate a closed group (see Results).

A protein family, as discussed previously, must comprise a cluster of monophyletic subfamilies, i.e., highly-related paralogous proteins, that derive all from a common ancestor. They

possess a significative sequence identity and/or similarity, and might share or not, common substrates or mechanisms of reaction.

In contrast, a protein macrofamily inside the MDR comprises a cluster of related protein families with a wide phylogenetic distribution, i.e., with protein members from the three domains of life, that originate from a common ancestor (monophyletics). Furthermore, inside each macrofamily, at least one subfamily possesses a physiologic role related to the primary metabolic pathways (with a probably ancient origin). Thus, the advantage to cluster protein families into macrofamilies lies on the fact that not all families are equally related, and this is probably due to the fact that some protein families are more ancient than others. Indeed, inside each macrofamily of the MDR, there is a probable ancestral group (see the previous section), that might be tracked to the last universal common ancestor. If this last is true, then the number of macrofamilies inside MDR superfamily, reflects the original number of MDR proteins that existed in the last universal common ancestor. It is important to mention that Castresana (Biochim Biophys Acta (2001) 1506: 147), after analysing the phylogenetic distribution and evolution of the bioenergetic pathways, concluded that the last universal common ancestor contained several members of each gene family. This agrees with the idea that the last universal common ancestor was a metabolically sophisticated organism.

Finally, it is interesting to point out that in comparison to the other taxonomic categories, the superfamily concept is not the focus of an extensive discussion, and there is an almost consensus agreement that in addition to sequence similarities, and a common evolutionary origin, 3D structure data should be taken in consideration. Thus, a superfamily can be considered as groups of homologous protein families (and/or macrofamilies) with a monophyletic origin, that share, at least, a barely detectable sequence similarity, but showing similar 3D structure (Koonin et al., (1998) Curr Opin Struct Biol 8: 355; Koonin et al., (2002) Nature 420: 218).

The inclusion of phylogenetic criteria to define subfamilies, families, macrofamilies and superfamilies can be subscribed toward the actual tendency to construct a natural taxonomy of proteins and protein families. The COG database —that comprises “ancient families conserved across a wide phylogenetic range” [140]—is one of the most successful examples.

Referee 2 Comments:
Overall, the paper is well written and sound.

Authors answer: Thank you for this appreciation, and for all specific comments that help us to improve the manuscript.

Specific comments
The summary has to be rewritten and be in line with the conclusion of the paper.
Pg 9, ln 17: (g subdivision), refer to tables for clearness.

Authors answer: done. We rewrite the summary incorporating the conclusions of the paper. In consequence we deleted the conclusions section to avoid duplicated information.

Pg 9, ln 19: Table IX should read something else.

Authors answer: Table 9 shows the number of identified genes that belong to MDR in completely sequenced species. There is great variability with respect to the total number of genes identified in each organism, even inside the same taxonomic category, as well as variability with respect to the number of genes identified in MDR superfamily. We believe that the data are illustrative by themselves, but if the referee considerate appropriate we can move it to the supplementary material.

Pg 9, bottom paragraph: Their members contain Zn2 , probably they do but it has not been shown for all. Use the same nomenclature (Zn2 or zinc) for the ligands throughout the paper, cf pg 10.

Authors answer: done. We made the requested changes.

Pg 11, paragraph The elicitor-. what does the end of the sentence mean?

Authors answer: We modify the paragraph: The elicitor-inducible defense-related proteins (ELI3) are present only in eudicot plants, and show different, but related defence activities: cinnamyl alcohol dehydrogenase, benzyl alcohol dehydrogenase, or mannitol dehydrogenase. ELI3 expression is elicited by fungal pathogens [60], wound [61], salicylic acid [62], leaf senescence [63]. In celery, there is a down-regulation by sugars or salt stress [64-66].

Pg 14: The discussion about paralogous and orthologous is interesting but the example with plant ADH/class III ADH is maybe not the best because they can also be classified as orthologous. They can use different substrates but they also share a number of substrates.

Authors answer: There is a widespread confusion about the exact meaning of the terms "orthology" and "paralogy" (see Sonnhammer & Koonin (2002) Trends Genet 18: 619). Many researchers seem to believe that orthologous are simply genes or proteins with the same function in different organisms, whereas paralogous are simply homologs within one organism. This does not agree with the original definitions of orthology and paralogy given by Fitch (Syst Zool (1970) 19: 99): two proteins are said to be paralogous if they are derived from a duplication event, but orthologous if they are derived from a speciation event (e.g., Li WH (1997) Molecular Evolution. Sinauer; Page & Holmes (1999) Molecular Evolution: A Phylogenetic Approach. Blackwell Science; Graur & Li (1999) Fundamentals of Molecular Evolution. Sinauer; Nei & Kumar (2000) Molecular Evolution and Phylogenetics. Oxford University Press). Therefore, initially a duplication event will produce two proteins possessing identical properties, and only after evolution they might acquire different functions. In contrast, true orthologous proteins must perform the same function. This explanation is obligatory since some outstanding papers provide inexact definitions, e.g., Altamirano et al., (2000) Nature 403: 617; Gerlt & Babbit (2000) Genome Biol 1: REVIEWS0005; Gerlt & Babbit (2001) Annu Rev Biochem 70: 209.
[Note: We included this material in our manuscript].

Pg 18, last sentence, second paragraph: What enzymes does modern MDRs refer to?

Authors answer: We modify and complete the sentence: Indeed, some MDRs fulfil all these requirements (e.g., Broad ADH subfamily [129;132;133], or animal ADH subfamily [15;134].

Pg 19, first sentence conclusion. Several of the 49 subfamilies show overlapping activities. In that sense it is hard to agree that all 49 subfamilies have a different function. How is it with animal ADH, plant ADH, class III ADH and fungi ADH? Maybe that they have some different functions but they also possess same activity, e.g. all can metabolise ethanol. Furthermore, here it is also important what does it mean that a protein has a different function.

Authors answer: You are right: a change in the functional role of a protein is not necessarily related with a change in substrates or mechanism of reaction. The recruitment of a duplicated protein into a different metabolic pathway, a different physiological role, or even a change in the spatiotemporal

pattern of expression, expressing a protein in novel tissues and/or developmental stages (True & Carroll (2002) Annu Rev Cell Dev Biol 18: 53), could be a good evolutive reason to conserve the duplicated protein, and resulting in a novel paralogous protein with a different functional role. [Note: we included this material in our manuscript].

Figure 1. The color coding is not the best in such a complicated figure. For easier reading it would have been easier to split this figure into one figure and one table. An explanation to bars (and yellow dots in other figures) is requested.

Authors answer: done.

Table 7. The number of protein coding genes given disagree in several cases with numbers found in the literature. Especially is this valid for the eukaryota where the values given are higher except for D. melanogaster. An explanation and/or references are requested.

Authors answer: done. Although please take in consideration that the extent of the proteomes for completely sequenced eukaryotic organisms (i.e. the total number of protein-coding sequences) is still not defined, even for the yeast proteome, completely sequenced more than six years ago (Harrison et al., (2002) Nucleic Acids Res 30: 1083).

Minor comments

The abbreviation list covers only a small part of all abbreviations used. At least ten more protein abbreviations are used throughout the manuscript. COG and NCBI could be included as well.

Authors answer: done.

Summary - the level of examples are very different, e.g. ethanol metabolism and lens proteins, which should be examples of functions

Authors answer: We rewrite the summary and suppress these examples.

Pg 3, ln 7 and pg 19, ln 2: the short-chain ADH should preferably read the short chain dehydrogenases/reductases. Use ADH for alcohol dehydrogenase throughout the paper

Pg 5, ln 18: 21 sequences should preferably read 21 proteins

Use the full name of an abbreviation the first time the name appear in the text, e.g. COG and several proteins

Pg , ln 19: use u.c. s for S-hydroxymethyl-glutathione and S-hydroxymethyl-mycothiol

Pg 24: Use l.c. for names in ref 42 and 48.

Use same type of nomenclature for tables in figure legends as in running text.

Authors answer: done. Thank you for the careful review and improvements to our manuscript.

Figure 5 looks like a table

Authors answer: We change this figure for a table.

Referee 3 Comments:
The manuscript by Riveros-Rosas et al. represents a comprehensive, structural and functional, in silico analysis of the medium-chain dehydrogenase/reductase (MDR) superfamily (583proteins). The authors use extensive database mining and the BLASTP program in an iterative manner in order to identify all the members of the MDR superfamily. Thus, they classify MDR proteins based on sequence as well as functional similarities, construct unrooted phylogenetic trees, and compare their results with previous partial classification attempts. The authors also infer that eukaryota inherited MDRs from bacteria, since MDR subfamilies in eukaryota are more closely related to their counterparts in bacteria than in archaea. As an ancestral activity, they propose formaldehyde metabolism and broad alcohol fermentation.
This work has been carefully performed and represents the most comprehensive and up-to-date study on this superfamily. In my opinion, it would be very useful to investigators in the field as well as a model to those involved in studies on protein function, taxonomy and evolution.

Authors answer: We sincerely appreciate your comments.

Referee 3 Comments:
A recent publication by Nordling et al. (2002), which is extensively cited in the present manuscript, included only 116 sequences from 6 completed genomes (5 eukaryotic plus Escherichia coli). In the present work, the entire set of MDR members has been studied and a different method of grouping protein sequences has been used.
The summary should be rewritten in order to include all the relevant information provided in the manuscript. For instance, it should contain the definitions of protein subfamily and cluster of subfamilies or division, according to authors. A good idea would be to merge the existing information from the summary and conclusions section. Some colloquial expressions, such as "and so on" or "Jornvall's group" should be avoided.

Authors answer: done. We rewrite the summary incorporating the conclusions of the paper. In consequence we deleted the conclusions section to avoid duplicated information.

Referee 3 Comments:
In the Introduction, the starting paragraph on non-MDR alcohol dehydrogenases, from the Fe-dependent and short-chain dehydrogenase/reductase superfamilies, could be omitted.

Authors answer: We prefer do not omit this paragraph, because in our experience, researchers outside the MDR or SDR area, seems to believe that all alcohol dehydrogenases belong to the same protein superfamily, but this is only our personal appreciation. If the referee considerate that the information of this paragraph is well known for the readers of the European Journal of Biochemistry, then we accept to delete this paragraph.

Referee 3 Comments:
Nomenclature referring to "cluster of protein subfamilies", "division", and "family" results a bit confusing for non-initiated readers. In page 8, the authors state that at the "highest level of integration" they identify 3 divisions, while at "lower levels of integration" they identify 6-8 families. The authors should be more precise in what they mean by high and low levels of integration. In addition, in some places in the text and the supplementary material, Roman numerals which are used

for divisions are also used to designate families (i.e., Family I and Family II, in p. 10; Family III, in p. 18). I suggest to keep Roman numerals for divisions and use specific names for families.

Authors answer: Thank you for your corrections and suggestions. We made several changes in respect of MDR taxonomy. First, we change the term division by macrofamily. On the original manuscript we used the term division, but in order to give a coherent protein taxonomy, we prefer the term macrofamily because division possesses on *sensu stricto*, a higher taxonomic rank than superfamily. However, although macrofamily is not a standardized taxonomic rank, do not exist a taxonomic rank located between family and superfamily, and since the protein superfamily concept is well established, all the protein taxonomic ranks inside a superfamily, must ideally possess a lower taxonomic rank than superfamily. To our knowledge the term macrofamily has been used before only in a linguistic context, but in a biological context, the classic paper of Paul C Sereno (Natl Geograph Res (1986) 2:234-256), used the next ranks between order and family: order > parvorder > nanorder > hyporder > monorder > suborder > infraorder > gigafamily > megafamily > grandfamily > hyperfamily > superfamily > family. Because in biology there is not intermediate rank between superfamily and family, we propose to use macrofamily, as a suitable alternative.

Second, we write a new topic under Discussion about the “taxonomy inside the MDR superfamily”. Thus, we classify MDR proteins based on phylogenetic, sequence and functional similarities, and propose a logic sketch to classify the members of the MDR superfamily into three different taxonomic categories: subfamily, family and macrofamily. Formal definitions are provided in order to allow their application to other protein superfamilies.

Referee 3 Comments:
In page 8, the bottom paragraph, "In this way,.." sounds repetitive when compared to previous paragraphs.

Authors answer: Thanks, you are right, we deleted the paragraph.

Referee 3 Comments:
Concerning Division I, is ScXDH (xylitol dehydrogenase) considered a member of this division? Why is this enzyme not mentioned anywhere in the text?

Authors answer: Yes, xylitol dehydrogenase is a member of the sorbitol dehydrogenase (DHSO) subfamily. We added the next explanatory note in Table 3: The members of this subfamily receive the official name of L-iditol 2-dehydrogenase, and possess alternative names as glucitol dehydrogenase, xylitol dehydrogenase or polyol dehydrogenase, in addition to sorbitol dehydrogenase. This subfamily catalyses the reversible oxidation of D-sorbitol and other polyalcohols, like xylitol and L-iditol, to the corresponding keto-sugars (Ng et al. (1992) J Biol Chem 267: 24989; Marini et al. (1997) Arch Biochem Biophys 340: 383; Lindstad et al. (1998) Biochem J 330: 479; Oura et al. (2000) Phytochemistry 54: 567).

Referee 3 Comments:
Within the ADH Family, most sequences bind NAD(H) as stated in the text, but ADH4 (now renamed ADH8) from Rana perezi is an exception since it prefers NADP(H) (Peralba et al., 1999).

Authors answer: Thank you for the observation. The original manuscript included this exception on the Table II, but now we added to the main text this information.

Referee 3 Comments:
The end of the first paragraph in p. 19, "On the other..substituted." is not easy to understand.

> **Authors answer:** You are right. We modify the paragraph: "On the other hand, the supply of precursors for fatty acid synthesis in bacteria and eukaryota is provided by acetyl-CoA carboxylase, an ancient enzyme also presents in archaea, which contains, instead of fatty acids, isoprenoid-based membranes. This suggests that the origin of acetyl-CoA carboxylase predates that of fatty acid synthesis, because fatty acids are absent in archaea. Apparently, the role of acetyl-CoA carboxylase in the supply of precursors for fatty acid synthesis is a later recruitment in the evolution of this enzyme. Thus, TDH and CCAR probably belong to ancient metabolic pathways that were subsequently substituted by other metabolic pathways".

Referee 3 Comments:
There are a few minor points. Some of them deserve further explanation in a manuscript intended for a journal with a broad readership:
- Some of the abbreviations should be defined.
- COG should be defined the first time that it appears in the text.
- The meaning of E-value should be explained.
- Many references include JID numbers that should be deleted.
- Table 7 is not cited in the text.
- In Fig. 1, divisions in roman numerals should be included.
- In Fig. 3, if BDH subfamily refers to fungi BDH (2,3-butanediol dehydrogenase), as stated in Table 1, p. 36), then they should be brown coloured. Add label for CCAR and related proteins.
- In legends of Figs. 1, 2, and 3, the meaning of separating yellow and grey pins and bars should be explained in terms of boundaries for each division or subfamily.

> **Authors answer:** Done.

Referee 3 Comments:
A few misspellings should be corrected to the suggested form in several places in the text:
- In the title, "medium-chain dehydrogenases/reductases" should be singular.
- Classical (p. 3, line 9 from top; p. 10, line 1 from top).
- Leukotriene
- P. 3, line 3 from top: "exhibit". Line 10: "alcohol dehydrogenase families". Line 17: "proteins".- P. 5, line 4 from top: "dendrograms".
- P. 6, line 4 from bottom: "agreement".
- P. 7, line 3 from top: reference 23 should be cited after authors' names. Line 4 from bottom: "allows".
- P. 8, line 15 from top: "divisions"
- P. 9, line 11 from top: "appear". Line 14: "possesses".
- P. 11, line 10 from top: "shows".
- P. 12, line 13 from bottom: "these".
- P. 14, line 8 from bottom: "hydroxyl".
- P. 15, line 3 from top: "preexisting".
- P. 16, line 13 from bottom: please use capital "S" for sulfur atom of derivative compounds of glutathione and mycothiol. Line 5 from bottom: "another". Line 3: "contain".
- P. 18, line 8 from bottom: "theoretical".

- Figure legends: "bottom", "COG families", "roman numerals". "Identity".

Authors answer: Done. Thank you very much for the careful review and improvements to our manuscript.

Héctor Riveros-Rosas

De: "Laraine Kerr" <kerr@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>
Enviado: miércoles, 02 de abril de 2003 10:29
Asunto: EJB-02-1375

2 April 2003

Dear Dr Riveros-Rosas

Thank you for your email dated 2 April 2003.

You will need to submit the new version via our online submission system EJB Manuscript Central, as you submitted the original version. To do this you should submit the paper as though it were an original manusript, ie go through the submission process again - it is not a revision, and the system won't let you submit it as a revision.

You will see that there is a radio button where you can indicate that your manusript is a new version of a previous manuscript, and a box where you can enter the number of the original version. This will ensure that the manuscript is assigned to Professor Tramontano as the Editor and she will ask us to send it back to the original referees.

In addition, please upload your detailed response to Professors Perham and Tramontano and the referees as a separate file for review. Please do not combine this with the file of the manuscript but upload them separately.

You may include a short covering letter to Professor Perham in the Authors' comments section if you wish.

Your User Account, User ID and password are the same as they were when you submitted the original version.

I hope the above is of assistance and that you can now submit the new version of your manuscript. Please do not hesitate to contact me if you encounter any problems.

Kind regards

Laraine Kerr

Editorial Secretary

Héctor Riveros-Rosas

De: <ejb@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
Enviado: miércoles, 02 de abril de 2003 13:44
Asunto: EJB Manuscript No. EJB-03-0296

EJB - The FEBS Journal

April 2nd, 2003

Manuscript number: EJB-03-0296
Title: DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASE/REDUCTASE SUPERFAMILY
Authors: 1) Héctor Riveros-Rosas 2) Adriana Julián-Sánchez 3) Rafael Villalobos-Molina 4) Juan Pablo Pardo 5) Enrique Piña

Dear Dr. Riveros-Rosas,

Thank you for submitting your manuscript to EJB.

As corresponding author, you will receive all future communications about this manuscript. Please remember to quote your manuscript number in all correspondence.

You can keep track of the progress of your manuscript through the Editorial system by periodically logging on to the EJB online submission website at http://ejb.manuscriptcentral.com.

The status of your manuscript will be displayed in the Author Centre. Please address any queries or comments to the Editorial Office.

We will advise you of the Editor's decision in due course.

Yours sincerely
Louise Sanders

Louise Sanders
Editorial Manager
European Journal of Biochemistry
98 Regent Street
Cambridge CB2 1DP, UK
Tel: +44 12223 369020
Fax: +44 1223 369090
Email: ejb@camfebs.co.uk

Héctor Riveros-Rosas

De: <ejb@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
CC: <Anna.tramontano@uniroma1.it>; <ejb@camfebs.co.uk>
Enviado: martes, 06 de mayo de 2003 8:16
Asunto: EJB manuscript EJB-03-0296

EJB - The FEBS Journal

6 May 03

Reference no.: EJB-03-0296

Title: DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASE/REDUCTASE SUPERFAMILY

Authors: 1) Héctor Riveros-Rosas 2) Adriana Julián-Sánchez 3) Rafael Villalobos-Molina 4) Juan Pablo Pardo 5) Enrique Piña

Editor: Anna Tramontano

Dear Dr. Riveros-Rosas

Thank you for submitting your paper for publication in EJB.

I am pleased to inform you that your paper has received favourable comments from the referees and is likely to be acceptable for publication in the Journal after incorporation of their comments, copies of which are appended below.

Please note that no guarantee of acceptance of the revised version can be given in advance. If accepted, the dates of receipt of both the original and revised papers will be printed.

Please submit the revised version online at your earliest convenience, and within four weeks at the latest, otherwise the manuscript will be treated as a new submission.

To submit the revised version:

1. Respond to the referees' comments on line (see details below).
2. Upload a revised version of the manuscript in Word format (see details below)
3. Upload the figures as a single .pdf file or separate files in the format used for the original version, for review purposes. This is particularly important if you have made any changes to the figures in the revised version.

In addition, please send to the Editorial Office:
- A completed Copyright Assignment Form

- A CD, floppy disk or zip disk containing .tif or .eps files of the figures and a set of good quality hardcopy figures (it is essential to follow the Guidelines for Publication of Quality Artwork sent to you as a separate email)

To respond to the referees' comments and submit your revised manuscript:

1. Log on to http://ejb.manuscriptcentral.com and enter your Author Centre where you will find your manuscript reference number followed by .R1 in a box marked 'Manuscripts for Revision'.
2. Click on the 'View Comments and Respond to Referees' button, and enter your responses.
3. Once you have entered your Responses to Referees, click on the 'Save Response' button. This will return you to the Submission screen.
4. Click on the title of the revised manuscript in this box and change any details as required.
5. Upload the revised manuscript including the figure files - you MUST upload a Word file of the revised text. Alterations to the text or new information added to the text should be highlighted IN BOLD TYPE.
6. SUBMIT the revised manuscript. Your revised manuscript will not register as having been submitted until you click on the 'Submit' button. You will receive an onscreen notification of submission and your manuscript number followed by .R1. If you do not, please contact the Editorial Office.

Please check that your revised manuscript includes a short running title (50 characters maximum including spaces), a list of 5 keywords and references in the correct format for EJB.

We look forward to receiving the revised version of your manuscript.

Yours sincerely
Vanessa Wilkinson

Dr Vanessa Wilkinson
European Journal of Biochemistry
98 Regent St
Cambridge CB2 1DP, UK
Tel: 44 1223 369020
Fax: 44 1223 369090
Email: ejb@camfebs.co.uk

REFEREES' COMMENTS

Referee 1 Comments:
The manuscript "Diversity, taxonomy and evolution of medium-chain dehydrogenase/ reductase superfamily" has been

improved clearly after the revision. The introduction of macrofamilies, families and subfamilies makes the grouping of MDRs fairly easy to follow. However, it would have been nice to have a clearer definition of these levels. They are explained in the running text but for a general understanding of the nomenclature, also in a general perspective, a table/figure is requested.

Many of the aspects of this manuscript concerns nomenclature.
Suggestions - change head in table 3 - 8 (cf table 9) to..subfamilies that conform PDH family of MDR (COG 1063 family) ..
Further, is it correct to call COG 1063 a family within this paper? It is a bit confusing with COG families and PDH family/ADH family etc. The running text is more stringent. Cf. Table 5 where COG 1064 seems to be a higher level of family as compared to the CAD and Y-ADH families.
Try to use the same nomenclature/level of family in all figures. Fig 1 - ER is given the same level as DHSO. Fig 5 - Y-ADH is given the same level as CADH. Y-ADH related should probably read Fungi ADH and related within this figure.
Page 13, bottom paragraph. Macrofamily II/CADH family comprises three subfamilies. Table 2 presents only two subfamilies under CAD (probably correct).
Check, throughout the manuscript (including tables and figures), how the nomenclature is used. If the nomenclature is not used stringent within this paper other authors won't use it.
(A minor point according to the nomenclature is the use of headings in the running text. Use macrofamily I as a main heading/paragraph for the families PDH and ADH etc).

Finally, it would have been nice to mention the Drosophila ADH (SDR) in the discussion under functional similarities or Enzymogenesis as an example of one activity but different protein structure solutions.

Referee 2 Comments:
The revised manuscript by Riveros-Rosas et al. has been substantially improved with respect to the previous version. In my opinion, the authors have answered most of the points raised by the reviewers plus they have extensively rewritten some parts (Summary and Discussion) and added new data (e.g., Fig. 5) and topics for discussion.

Some of the new added paragraphs need to be adequately integrated in the manuscript:

- In page 9, lines 2 and 13 from bottom, the statement on the requirement of cutoff values is unnecessarily repeated.
- In page 17, the newly added paragraph at the bottom has many language problems. Firstly, the sentence "metric measurement for protein homology" has no clear meaning. Homology is not a countable word or a property that can be quantified. A gene or a protein is homologous (paralogous or orthologous) or is not.

Instead "identity" or "similarity" are recommended. Secondly, "metric measurement" sounds redundant and inappropriate. Instead, "measure of sequence identity" could be used. Also, the sentence "they (sequence similarities) should be focused over functional similarities" is unclear. Alternatively, the entire sentence from "Since." to "functional similarities" could be deleted, because is overly verbose and does not add new information.
- In the Discussion, the new paragraph on "Proteinogenesis versus Enzymogenesis" can be omitted or reduced since is mostly review-like.
- Page 25, lines 12-14: The same idea (that fatty acids are absent in archaea) is repeated twice.

There are some minor points that should be taken care of:

- "CADH family", as in page 10, should be used throughout the text, figures, figure legends and tables instead of "CAD family" (CAD is not defined under abbreviations).
- Page 12, line 9 from top, please note that the sentence ".with the exception of bi-domain oxidoreductase subfamily (BDOR), that appears to be (represented by) a transmembrane protein" has no sense since a subfamily is not a protein. Words within parenthesis could be added to make it clearer. A similar inappropriate use of language is seen further below, in line 18: "Of these five subfamilies, four are tetrameric proteins and three are present in archaea". Here the word "include" could substitute "are".
- Page 19, line 4 from top: Delete "(P93243)", since it has been previously introduced.
Line 10 form top: Use "herbivorous".
- Page 23: There may be some problems with the computer use of Greek characters: "a6beta6" and "a2".
- Page 25, line 1 from bottom: Use "amphibia". Add "at least" or "minimally" before "two enzymes".
- Page 26, line 15 from bottom: "physiological".
- Page 27, line 10 from top: "allows".
Line 14 from top: "classical".
- References:
- Ref [43]: Delete either "edition" or "edn".
- Ref [46]: Use initial and final page numers: "reviews0005.1-000.10".
- Ref [48]: "formaldehyde".
- Ref [49].: "Farres, J.".
- Ref [50]: Do not use upper case letters: "Pares, X.".
- Ref [66]: "mitochondria".
- Ref [76]: "15-oxoprostaglandin".
- Ref [81]: The volume number is not clearly seen.
- Ref [119]: Delete "Pt 1".

- Figures:

- Figure 3: Several color labels need to be corrected: "Secondary ADH" should be light blue since it corresponds to protista (Entamobidae). "Class III ADH" should be written in black rather than gray in order to be consistent. The angled line pointing at

Plant ADH subfamily should be in black rather than green.

- Figure 4: What is the meaning of the dotted line?

- Figure 5: The same colour code as that shown in Fig. 1 should be used: orange for bacteria and dark blue for archaea.

- Tables:

- Table 1, page 49, line 9 from top: A question mark ("?") should be changed to a Greek letter (zeta).

- Table 2, page 50, line 2 from top: Delete "paper".

- Table 3, page 53, line 8 from bottom: Use "have" instead of "has".
Line 1 from bottom: "..of supplementary.".
- Table 4, page 54, line 7 from top: "belongs".
Line 14 from top: "This shows."
Line 23 from bottom: ".on glutathione".
Line 14 from bottom: Some animal ADH are also heterodimers (e.g., some isozymes from human class I ADH).
Line 10 from bottom: "comprises".
Line 7 from bottom: "uses".
- Table 8, page 60, line 10 from bottom: "mycotoxin biosynthesis".
Line 4 from bottom: "belongs".
- Table 9, page 61, line 1 from top: "members in organisms with complete genome sequences".

Héctor Riveros-Rosas

De: <wilkinson@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
Enviado: miércoles, 07 de mayo de 2003 8:02
Asunto: EJB-03-0296 - Essential Information

EJB - The FEBS Journal

GUIDELINES FOR PUBLICATION QUALITY ARTWORK FOR EJB

In order for manuscripts to be considered for publication in EJB, please send to the Editorial Office:

1. A completed Copyright Assignment Form (see Instructions and Forms on the EJB web site: http://ejb.manuscriptcentral.com)
2. A good quality set of hardcopy figures, including two copies of any colour figures
3. A CD, floppy disk or zip disk containing electronic copies of the figures

Hardcopies of the figures

It is essential that you supply good quality hardcopies of your figures. Publication of the manuscript will be delayed if these are not provided.

For line drawings, a good quality laser-printed copy should be supplied. For gels, electron micrographs, colour figures, etc. glossy prints are required.

For all figures, please check that text and lines are clear, and photographs, e.g. gel pictures, are not blurred or speckled.

The axes of each graph should be lettered with the numerical scale and the measured quantity with units according to the style of the Journal.

All hardcopies of the figures should be labelled, preferably on the back, with the figure number and the authors' names. Please indicate the top of the figure.

For more information on preparing print quality graphics, please see http://www.blackwell-science.com/elecmed/digilla.htm

Electronic copies of the figures

Digital images should be provided as .eps or .tiff files only.

Where materials have been scanned to produce these images, the scanning resolution should be as follows to ensure adequate reproduction: lineart, >800 d.p.i.; half-tones (including gel photographs), >300 d.p.i.

Black and white images should be supplied as 'grayscale' and colour images should be supplied as CMYK.

For more information please consult Electronic Artwork Guidelines (see Instructions and Forms on the EJB web site: http://ejb.manuscriptcentral.com)

Suggested cover figures

Authors with a colour figure that would make a good image for the cover are invited to submit a hardcopy of the figure 12 cm by 12 cm without any labels or scale bars. An electronic copy of the figure should also be provided at 600 d.p.i, together with a short legend for the figure.

Yours sincerely
Louise Sanders

Louise Sanders
Editorial Manager
European Journal of Biochemistry
98 Regent Street
Cambridge CB2 1DP, UK
Tel: +44 1223 369020
Fax: +44 1223 369090
e-mail: ejb@camfebs.co.uk

Reply to editor and referees.

Response to editor:

We included a new figure (6) and a brief paragraph about the analogous enzymes shared between the MDR and the short-chain dehydrogenase/reductase superfamilies. We sincerely appreciate the comments, observations and corrections of the three referees, to substantially improve our manuscript. We are embarrassed because one of the referees found a lot of minor mistakes in our manuscript, and we are surprised because the referee corrected inclusive titles of articles in the reference section (that correspond to errors found in the original medline records!). We believe that this carefully review and discussion must be extensively acknowledge, and therefore we included an explicit acknowledge to the referees in our final manuscript.

Response to referees:

Referee 1 Comments:
The manuscript "Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily" has been improved clearly after the revision. The introduction of macrofamilies, families and subfamilies makes the grouping of MDRs fairly easy to follow. However, it would have been nice to have a clearer definition of these levels. They are explained in the running text but for a general understanding of the nomenclature, also in a general perspective, a table/figure is requested.

> **Authors answer:** In first place, thank you for the observations, corrections and suggestions. We are pleased to include a figure showing a general perspective about the proposed protein taxonomy.

Referee 1 Comments:
Many of the aspects of this manuscript concerns nomenclature. Suggestions - change head in table 3 - 8 (cf table 9) to..subfamilies that conform PDH family of MDR (COG 1063 family) ..
Further, is it correct to call COG 1063 a family within this paper? It is a bit confusing with COG families and PDH family/ADH family etc. The running text is more stringent.
Cf. Table 5 where COG 1064 seems to be a higher level of family as compared to the CAD and Y-ADH families.

> **Authors answer:** We change head in tables 3-8, and heading/paragraph for the protein families in the running text. Also we avoid to call COG as families. In sensu stricto, they are similar, but not equal to the protein families described here. Thank you for the correction. On other hand, it is interesting to observe that four of the COGs identified with the automated procedure described by Koonin and Tatusov, are identical to the protein families proposed by Nordling et al., and this paper.

Referee 1 Comments:
Try to use the same nomenclature/level of family in all figures. Fig 1 - ER is given the same level as DHSO. Fig 5 -Y-ADH is given the same level as CADH. Y-ADH related should probably read Fungi ADH and related within this figure.

Authors answer: Done. We corrected all the figures with the aim to use exactly the same names for protein families/subfamilies indicated on tables and running text. The figure 1 in particular, shows the protein subfamilies indicated in the SWISSPROT database; they do not correspond to the protein families and subfamilies proposed in this work. To avoid confusion, we mention this last in a more explicit way.

Referee 1 Comments:
Page 13, bottom paragraph. Macrofamily II/CADH family comprises three subfamilies. Table 2 presents only two subfamilies under CAD (probably correct). Check, throughout the manuscript (including tables and figures), how the nomenclature is used. If the nomenclature is not used stringent within this paper other authors won't use it.

Authors answer: Done, and thank you for the correction. There are two subfamilies on the CADH family, and we check that the same names are used throughout the figures, tables and running text.

Referee 1 Comments:
(A minor point according to the nomenclature is the use of headings in the running text. Use macrofamily I as a main heading/paragraph for the families PDH and ADH etc).

Authors answer: Done. We change the heading/paragraphs as you suggested.

Referee 1 Comments:
Finally, it would have been nice to mention the Drosophila ADH (SDR) in the discussion under functional similarities or Enzymogenesis as an example of one activity but different protein structure solutions.

Authors answer: It is a nice suggestion. As you requested, we included the next paragraph with the aim to point out examples of analogous enzymes shared between MDR and SDR: "It is worth to mention the existent parallelism between the MDR and the short-chain dehydrogenase/reductase (SDR) superfamilies, since both share several analogous enzymes. Thus, the SDR superfamily contains an analogous alcohol dehydrogenase found in Drosophila [79], a glucose dehydrogenase from Bacillus [80;81], an enoyl reductase from bacteria and plants [82-84], a sorbitol dehydrogenase from Klebsiella [85], and a threonine dehydrogenase in animals [86]. These enzymes represent different protein structure solutions to the same activities observed in MDR".

Referee 2 Comments:
The revised manuscript by Riveros-Rosas et al. has been substantially improved with respect to the previous version. In my opinion, the authors have answered most of the points raised by the reviewers plus they have extensively rewritten some parts (Summary and Discussion) and added new data (e.g., Fig. 5) and topics for discussion.

Authors answer: In first place, we acknowledge the corrections, observations and suggestions to our manuscript. We are convinced that they substantially contributed to improve our manuscript.

Referee 2 Comments:

Some of the new added paragraphs need to be adequately integrated in the manuscript: In page 9, lines 2 and 13 from bottom, the statement on the requirement of cutoff values is unnecessarily repeated. In page 17, the newly added paragraph at the bottom has many language problems. Firstly, the sentence "metric measurement for protein homology" has no clear meaning. Homology is not a countable word or a property that can be quantified. A gene or a protein is homologous (paralogous or orthologous) or is not. Instead "identity" or "similarity" are recommended. Secondly, "metric measurement" sounds redundant and inappropriate. Instead, "measure of sequence identity" could be used. Also, the sentence "they (sequence similarities) should be focused over functional similarities" is unclear. Alternatively, the entire sentence from "Since." to "functional similarities" could be deleted, because is overly verbose and does not add new information.

Authors answer: We offer our apologies for the involuntary mistakes introduced in the new sections of our manuscript. We deleted the indicated sentences as you suggested.

Referee 2 Comments:

- In the Discussion, the new paragraph on "Proteinogenesis versus Enzymogenesis" can be omitted or reduced since is mostly review-like.

- Page 25, lines 12-14: The same idea (that fatty acids are absent in archaea) is repeated twice.

Authors answer: Thank you for the observations. We reduced the first paragraph and deleted the repeated information.

Referee 2 Comments:

There are some minor points that should be taken care of:

-- "CADH family", as in page 10, should be used throughout the text, figures, figure legends and tables instead of "CAD family" (CAD is not defined under abbreviations).

-- Page 12, line 9 from top, please note that the sentence ".with the exception of bi-domain oxidoreductase subfamily (BDOR), that appears to be (represented by) a transmembrane protein" has no sense since a subfamily is not a protein. Words within parenthesis could be added to make it clearer. A similar inappropriate use of language is seen further below, in line 18: "Of these five subfamilies, four are tetrameric proteins and three are present in archaea".

Here the word "include" could substitute "are".

-- Page 19, line 4 from top: Delete "(P93243)", since it has been previously introduced.

-- Line 10 form top: Use "herbivorous".

-- Page 23: There may be some problems with the computer use of Greek characters: "a6beta6" and "a2".

-- Page 25, line 1 from bottom: Use "amphibia". Add "at least" or "minimally" before "two enzymes".

-- Page 26, line 15 from bottom: "physiological".

-- Page 27, line 10 from top: "allows".

-- Line 14 from top: "classical".

-- References:

-- Ref [43]: Delete either "edition" or "edn".

-- Ref [46]: Use initial and final page numers: "reviews0005.1-000.10".

-- Ref [48]: "formaldehyde".

-- Ref [49].: "Farres, J.".
-- Ref [50]: Do not use upper case letters: "Pares, X.".
-- Ref [66]: "mitochondria".
-- Ref [76]: "15-oxoprostaglandin".
-- Ref [81]: The volume number is not clearly seen.
-- Ref [119]: Delete "Pt 1".
-

Authors answer: Thank you for all these corrections. We are embarrassed because you find a lot of minor mistakes. We corrected all of them, and modify the figures with the aim to use exactly the same names for protein families/subfamilies in tables and running text.

Referee 2 Comments:
- Figures:
-- Figure 3: Several color labels need to be corrected: "Secondary ADH" should be light blue since it corresponds to protista (Entamobidae). "Class III ADH" should be written in black rather than gray in order to be consistent. The angled line pointing at Plant ADH subfamily should be in black rather than green.
-- Figure 4: What is the meaning of the dotted line?
-- Figure 5: The same colour code as that shown in Fig. 1 should be used: orange for bacteria and dark blue for archaea.

Authors answer: Thank you for the corrections. We made the requested changes and deleted the angled line pointing at plant ADH subfamily.

Referee 2 Comments:
-- Tables:
-- Table 1, page 49, line 9 from top: A question mark ("?") should be changed to a Greek letter (zeta).
-- Table 2, page 50, line 2 from top: Delete "paper".
-- Table 3, page 53, line 8 from bottom: Use "have" instead of "has". Line 1 from bottom: "..of supplementary.".
-- Table 4, page 54, line 7 from top: "belongs".
-- Line 14 from top: "This shows."
-- Line 23 from bottom: ".on glutathione".
-- Line 14 from bottom: Some animal ADH are also heterodimers (e.g., some isozymes from human class I ADH).
-- Line 10 from bottom: "comprises".
-- Line 7 from bottom: "uses".
-- Table 8, page 60, line 10 from bottom: "mycotoxin biosynthesis".
- Line 4 from bottom: "belongs".
-- Table 9, page 61, line 1 from top: "members in organisms with complete genome sequences".

Authors answer: Thank again for the corrections. We attended all of them and sincerely appreciate all the improvements to our manuscript.

Héctor Riveros-Rosas

De: <wilkinson@camfebs.co.uk>
Para: <hriveros@servidor.unam.mx>; <adh@laguna.fmedic.unam.mx>
Enviado: lunes, 02 de junio de 2003 7:10
Asunto: RE: EJB-03-0296.R1

RE: EJB-03-0296.R1 - DIVERSITY, TAXONOMY AND EVOLUTION OF MEDIUM-CHAIN DEHYDROGENASE/REDUCTASE SUPERFAMILY

Dear Dr. Riveros-Rosas

Thank you for submitting your paper for publication in the
European Journal of Biochemistry. I am pleased to inform you
that it has been accepted for publication in the journal.

Before the manuscript can be sent to the publisher, please send an electronic copy of the supplementary material in Word format (.doc) to my email address. In addition, please send to the Editorial Office:

1. A signed copy of the Copyright Assignment Form (see Instructions and Forms in your Author Centre).

2. Electronic copies of the figures on a CD, floppy or zip disk as TIFF or EPS files. (No other file format is acceptable so, if these are not available, do not send other file formats.)

3. Good quality hardcopies of the figures

In order to avoid delay to publication, please follow the Guidelines for Publication Quality Artwork that will be sent to you as a separate email.

Yours sincerely
Vanessa Wilkinson

Dr Vanessa Wilkinson
Assistant Editorial Manager
European Journal of Biochemistry
98 Regent St
Cambridge CB2 1DP
UK

Tel: +44 1223 369020
Fax: +44 1223 369090
Email: wilkinson@camfebs.co.uk

Identified proteins as members of MDR superfamily

In this work, we implemented a procedure which assigns identifiers to protein sequences from multiple sources; only sequence names reported by the non-redundant Swiss-Prot database were used directly; for protein sequences obtained from other databases (TrEMBLE, PIR, GenBank, etc.), provisional protein sequences identifiers were assigned. They included a SwissProt - like identifiers (GENE_ORGANISM) marked with a lower-case letter “t,” followed by the accession number assigned by the database (GenBank, PIR, TrEMBLE, etc.). In this way, the names used can be easily referred to the original database source.

Below protein name, family and subfamily to which each protein belong is indicated between parenthesis. Subfamily is indicated only if the protein can be unambiguously assigned to one subfamily.

Table S1. Proteins that belong to MDR superfamily

Archaea	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Archaeoglobus fulgidus Euryarchaeota; Archaeoglobales;	--	ADHt_ARCFU_O28179 (PDH Fam)	--
Methanobacterium thermoautotrophicum Euryarchaeota; Methanobacteriales;	--	--	--
Methanococcus jannaschii Euryarchaeota; Methanococcales;	--	--	--
Pyrococcus abyssi Euryarchaeota; Thermococcales;	--	TDHt_PYRAB_CAB50292 (PDH Fam: TDH)	--
Pyrococcus furiosus Euryarchaeota; Thermococcales;	--	TDHt_PYRFU_AAL81115 (PDH Fam: TDH, not included)	--
Pyrococcus horikoshii Euryarchaeota; Thermococcales;	--	ADHt_PYRHO_BAA29746 (PDH Fam: TDH)	--
Haloferax volcanii Euryarchaeota; Halobacteriales;	--	ADHt_HALVO_AAB71801 (PDH Fam: FADH)	--
Halobacterium sp. NRC-1 Euryarchaeota; Halobacteriales; Halobacteriaceae; Halobacterium.	--	ADH2t_HALSP_AAG20655 (PDH Fam: FADH) VNG1023Ct_HALSP_AAG19435 (PDH Fam) GDHt_HALSP_AAG18991 (PDH Fam: GDH archaea)	ADH4t_HALSP_AAG20032 (QOR Fam: CCAR related) QORt_HALSP_AAG19273 (LTD Fam: AADH/LHD related, probable horizontal gene transfer)

Thermoplasma acidophilum Euryarchaeota; Thermoplasmales;	ADHRt_THEAC_CAC11970 (Y-ADH Fam: Broad ADH) ADHRt_THEAC_CAC12437 (Y-ADH Fam: Broad ADH)	DHSOt_THEAC_CAC11883 (PDH Fam) DHG_THEAC (PDH Fam: GDH archaea)	--
Aeropyrum pernix Crenarchaeota; Aeropyrum.	ADH1t_AERPE_BAA81251 (Y-ADH Fam: Broad ADH) ADH2t_AERPE_BAA80556 (Y-ADH Fam: Broad ADH)	ADH3t_AERPE_BAA80235 (PDH Fam)	--
Sulfolobus solfataricus Crenarchaeota; Sulfolobales;	ADH10t_SULSO_AAK41665 (Y-ADH Fam: Broad ADH) ADH3_SULSO (Y-ADH Fam: Broad ADH) ADH4t_SULSO_AAK41537 (Y-ADH Fam: Broad ADH) ADH7t_SULSO_AAK42584 (Y-ADH Fam: Broad ADH) ADH5t_SULSO_AAK41859 (Y-ADH Fam: Broad ADH) ADH2t_SULSO_AAK41061 (Y-ADH Fam: Broad ADH related) ADH8t_SULSO_AAK42630 (Y-ADH Fam: Broad ADH related) ADH13t_SULSO_AAK42983 (Y-ADH Fam: Broad ADH related)	GDHt_SULSO_CAA09918 (PDH Fam: GDH archaea) ADH1t_SULSO_AAK40795 (PDH Fam) ADH3t_SULSO_AAK41467 (PDH Fam) ADH9t_SULSO_AAK42635 (PDH Fam) ADH11t_SULSO_AAK42827 (PDH Fam)	ADH6t_SULSO_AAK42487 (QOR Fam: CCAR related) ADH12t_SULSO_AAK42913 (QOR Fam: CCAR related)

Bacteria	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Thermotoga maritima Thermotogales; Thermotoga.	--	ADH1t_THEMA_AAD35521 (PDH Fam: FDEH) ADH2t_THEMA_AAD35497 (PDH Fam) ADH3t_THEMA_AAD35386 (PDH Fam: SSP related)	--
Borrelia burgdorferi Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.	--	--	--
Treponema pallidum Spirochaetales; Spirochaetaceae; Treponema	--	--	--
Treponema denticola Spirochaetales; Spirochaetaceae; Treponema		tdent_7303 (6821-7822) (Frame: +2) tdent_7325 (17544-18662) (Frame: +3)	
Deinococcus radiodurans Thermus/Deinococcus group; Deinococcus.	--	TDHt_DEIRA_AAF12215 (PDH Fam: TDH) ADHt_DEIRA_AAF11826 (PDH Fam: FADH) ADHt_DEIRA_AAF12293 (PDH Fam: FADH)	QORt_DEIRA_AAF12387 (QOR Fam) QORt_DEIRA_AAF10634 (QOR Fam)
Chlamydia muridarum Chlamydiales; Chlamydiaceae; Chlamydia	--	--	--
Chlamydia trachomatis Chlamydiales; Chlamydiaceae; Chlamydia	--	--	--
Chlamydia pneumoniae Chlamydiales; Chlamydiaceae; Chlamydia	--	--	--
Chlamydia pneumoniae J138 Chlamydiales; Chlamydiaceae; Chlamydia	--	--	--
Chlamydia pneumoniae AR39 Chlamydiales; Chlamydiaceae; Chlamydia	--	--	--
Buchnera sp. APS Proteobacteria; gamma subdivision; Buchnera.	--	--	--

Vibrio cholerae Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio		TDHt_VIBCH_AAF96783 (PDH Fam: TDH)	
Pasteurella piscicida. (Photobacterium Damsela (Subsp. Piscicida)). Og Plasmid. Proteobacteria; Gamma Subdivision; Vibrionaceae; Photobacterium.		ADH3_PASPI (ADH Fam: Class III ADH)	
Escherichia coli Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	ADHP_ECOLI (Y-ADH Fam: Fungi ADH related) YAHK_ECOLI (CAD Fam: CADH related) YJGB_ECOLI (CAD Fam: CADH related) CADHt_ECOLI_P71306 (CAD Fam: CADH related)	ADH3_ECOLI (ADH Fam: Class III ADH) FADHt_ECOLI_Q59399 (ADH Fam: Class III ADH) GATD_ECOLI (PDH Fam: GATD) YPHC_ECOLI (PDH Fam: FDEH) IDND_ECOLI (PDH Fam) YDJJ_ECOLI (PDH Fam: DHSO) RSPB_ECOLI (PDH Fam: SSP) TDH_ECOLI (PDH Fam: TDH) YBDR_ECOLI (PDH Fam: FADH) YCJQ_ECOLI (PDH Fam) YDJL_ECOLI (PDH Fam) YJJN_ECOLI (PDH Fam: SSP) YGGP_ECOLI (PDH Fam: SORE) B2931t_ECOLI_B65078 (not included: is a fragment from YGGP_ECOLI) B2932t_ECOLI_C65078 (not included: is a fragment from YGGP_ECOLI)	QOR_ECOLI (QOR Fam: TED 2 related) YNCB_ECOLI (LTD Fam: AADH/LHD related) YHDH_ECOLI (QOR Fam)

Klebsiella pneumoniae Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.		SORE_KLEPN (PDH Fam: SORE)	
Morganella morganii Proteobacteria; gamma subdivision; Enterobacteriaceae; Morganella.		XDHt_MORMO_AAA25324 (PDH Fam: DHSO)	
Salmonella typhimurium Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.			QOR_SALTY (QOR Fam: TED 2 related)
Haemophilus influenzae Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.	--	ADH3_HAEIN (ADH Fam: Class III ADH) Y053_HAEIN (PDH Fam: SSP)	--
Pseudomonas aeruginosa PA01 Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.	ADHt_PSEAE_AAG08812 (Y-ADH Fam: Fungi ADH related) ADHt_PSEAE_AAG05663 (CAD Fam: CADH related)	ADHt_PSEAE_AAG07484 (PDH Fam: BDH) ADH3t_PSEAE_AAG07017 (ADH Fam: Class III ADH) ADHt_PSEAE_AAG05576 (PDH Fam: FADH) ADHt_PSEAE_AAG05546 (PDH Fam: FADH) ADHt_PSEAE_AAG05507 (PDH Fam: secondary ADH) BDDHt_PSEAE_AAG07540 (PDH Fam: BDH)	QOR_PSEAE (QOR Fam: TED 2 related) PORt_PSEAE_AAG05037 (LTD Fam: AADH/LHD related) ORFt_PSEAE_AAG05585 (LTD Fam: AADH/LHD related) PORt_PSEAE_AAG04526 (QOR Fam: PIG3 related) PORt_PSEAE_AAG08619 (QOR Fam) PORt_PSEAE_AAG06955 (QOR Fam: Alginate lyase)

Pseudomonas putida. Proteobacteria; gamma subdivision; Pseudomonas group; Pseudomonas.		FADH_PSEPU (PDH Fam: FADH) FDEH_PSEPU (PDH Fam: FDEH) FDMt_PSEPU_AAA25818 (PDH Fam: FADH) XYLB_PSEPU (ADH Fam: Benzyl ADH) XYLWt_PSEPU_O30864 (PDH Fam: TDH related: benzyl ADH in plasmid pDK1) TMBWt_PSEPU_Q51992 (PDH Fam: TDH related: benzyl ADH) XYLWt_PSEPU_Q51972 (PDH Fam: TDH related: benzyl ADH in plasmid TOL pWWo) BDHt_PSEPU_AAB58982 (PDH Fam: BDH)	
Pseudomonas syringae pv. glycinea Proteobacteria; gamma subdivision; Pseudomonas group; Pseudomonas.			ADHt_PSESY_AAC38656 (QOR Fam: CCAR related) PKS1t_er-f_PSESY_AAD03047 (ER Fam: ER-modular PKS)
Pseudomonas sp. Proteobacteria; gamma subdivision; Pseudomonas group; Pseudomonas.		ADHt_PSESP_AAD37848 (ADH Fam: Benzyl ADH, different to PSEPU) TERD_PSESP (ADH Fam: Benzyl ADH, different to PSEPU NTNWt_PSESP_AAC38358 (PDH Fam: TDH related: benzyl ADH) ADH3t_PSESP_CAB64351 (ADH Fam: Class III ADH)	AGLt_PSESP_AAD15785 (QOR Fam: Alginate lyase)
Acinetobacter baumannii Proteobacteria; gamma subdivision; Pseudomonas group; Moraxellaceae;		ADH3t_ACIBA_AAD46162 (ADH Fam: Class III ADH)	
Acinetobacter calcoaceticus Proteobacteria; gamma subdivision; Pseudomonas group; Moraxellaceae;		XYLBt_ACICA_AAC32671 (ADH Fam: Benzyl ADH)	

Acinetobacter lwoffii Proteobacteria; gamma subdivision; Pseudomonas group; Moraxellaceae;		Hpt_ACILW_CAB57202 (PDH Fam: BDOR related)	
Acinetobacter sp. M-1 Proteobacteria; gamma subdivision; Moraxellaceae; Acinetobacter.	ADHt_ACISP_BAB12270 (CAD Fam: CADH related)		
Acinetobacter sp. SE19 Proteobacteria; gamma subdivision; Moraxellaceae;	ADHt_ACISP_AAG10028 (Y-ADH Fam: Broad ADH)		
Acinetobacter sp. ADP1 Proteobacteria; gamma subdivision; Pseudomonas group; Moraxellaceae;		BADHt_ACISP_AAD34026 (ADH Fam: Benzyl ADH)	
Methylobacter marinus. Proteobacteria; gamma subdivision; Methylococcaceae; Methylobacter.		FADH_METMR (PDH Fam: FADH)	
Stenotrophomonas maltophilia. Proteobacteria; gamma subdivision; Xanthomonas group; Stenotrophomonas.	ALKBt_STEMA_AAA97866 (False positive; Y-ADH Fam, Fungi-like)		
Xylella fastidiosa Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.	ADHt_XYLFA_AAF83946 (CAD Fam: CADH related) ADHt_XYLFA_AAF84543 (CAD Fam: CADH related) ADHt_XYLFA_AAF85188 (CAD Fam: CADH related) ADHt_XYLFA_AAF84555 (CAD Fam: CADH related)	ADHt_XYLFA_AAF84536 (PDH Fam: secondary ADH)	CHPt_XYLFA_AAF83261 (QOR Fam)
Xanthomonas campestris pv. campestris. Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.		TDH_XANCP (PDH Fam: TDH)	
Rickettsia prowazekii Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia	--	--	--
Paracoccus denitrificans. Proteobacteria; Alpha Subdivision; Rhodobacter Group; Paracoccus.		FADH_PARDE (ADH Fam: Class III ADH)	

Rhizobium leguminosarum bv. trifolii Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae;		DHSOt_RHILE_O87588 (PDH Fam: DHSO)	
Rhizobium tropici. Plasmid pRtrCFN299a. Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae;		XYLB1tf_RHITR_O52367 (ADH Fam: Benzyl ADH) XYLB2tf_RHITR_O52372 (ADH Fam: Benzyl ADH)	
Sinorhizobium meliloti. **(Rhizobium meliloti).** Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium.	ADHA_RHIME (Y-ADH Fam: Fungi ADH related)	TDH_RHIME (PDH Fam: TDH)	FIX23-4t_RHIME_S18956 (ER Fam: ER modular PKS)
Agrobacterium tumefaciens Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae;			QORt_AGRTU_BAA87717 (QOR Fam: TED2 related)
Brucella abortus Proteobacteria; alpha subdivision; Rhizobiaceae group; Brucella.	ADHt_BRUAB_O54388 (Y-ADH Fam: Fungi ADH related)		
Caulobacter crescentus. Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.			QORt_CAUCR_AAK23548 (QOR Fam)
Rhodopseudomonas palustris Proteobacteria; alpha subdivision; Bradyrhizobium group;			QORt_RHOPA_AAC23924 (QOR Fam: CCAR related)
Rhodobacter capsulatus Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.		BCHC_RHOCA (PDH Fam: BCHC)	QORt_RHOCA_AAC32303 (QOR Fam: TED 2 related)
Rhodobacter sphaeroides **(Rhodopseudomonas Sphaeroides).** Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.		ADHI_RHOSH (ADH Fam: Class III ADH) BCHC_RHOSH (PDH Fam: BCHC)	
Methylobacterium extorquens Proteobacteria; alpha subdivision; Methylobacterium.			ADHAt_METEX_Q49134 (QOR Fam: CCAR related)
Sphingomonas aromaticivorans Proteobacteria; alpha subdivision; Zymomonas group; Sphingomonas.		XYLBt_SPHAR_O85841 (ADH Fam: Benzyl ADH)	

Zymomonas mobilis. Bacteria; Proteobacteria; Alpha subdivision; Zymomonas group;	ADH1_ZYMMO (Y-ADH Fam: Fungi ADH related)		ADHt_ZYMMO_AAD19419 (QOR Fam: PIG 3)
Neisseria meningitidis MC58 Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.	ADHPt_NEIME_AAF40975 (Y-ADH Fam: Fungi ADH related)	ADHt_NEIME_AAF41759 (PDH Fam: FADH) ADHt_NEIME_AAF41031 (PDH Fam: BDH) Clas III present (not reported yet) 1321431-1325664 (frame=+3)	
Neisseria meningitidis Z2491 Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.	ADHt_NEIME_CAB84010 (Y-ADH Fam: Fungi ADH related)	ADHt_NEIME_CAB84840 (PDH Fam: FADH) ADHt_NEIME_CAB84091 (PDH Fam: BDH) ADH3t_NEIME_CAB84746 (ADH Fam: Class III ADH)	--
Alcaligenes eutrophus. **(Ralstonia eutropha).** Proteobacteria; beta subdivision; Burkholderia group; Ralstonia.		ADH_ALCEU (PDH Fam: secondary ADH)	
Rubrivivax gelatinosus Proteobacteria; beta subdivision; Comamonadaceae; Rubrivivax.		BCHCt_RUBGE_BAA94034 (PDH Fam: BCHC)	QORt_RUBGE_BAA94068 (QOR Fam)
Thauera aromatica Proteobacteria; beta subdivision; Rhodocyclus group; Thauera.	ADHt_THAAR_O87871 (Y-ADH Fam)		
Thiobacillus sp. (Plasmid pT3.2I.) Proteobacteria; beta subdivision; Thiobacillus.		ADHt_THISP_Q9ZEX7 (PDH Fam)	
Sorangium cellulosum Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Sorangiaceae; Sorangium.			EPOAt_SORCE_AAF62880 (ER Fam: ER-PKS) EPODt_er-fa_SORCE_AAF62883 (ER Fam: ER-PKS) EPODt_er-fb_SORCE_AAF62883 (ER Fam: ER-PKS) EPOFt_SORCE_AAF62885 (ER Fam: ER-PKS)

Campylobacter jejuni Proteobacteria; epsilon subdivision; Campylobacter group; Campylobacter.	ADHt_CAMJE_CAB73964 (CAD Fam: CADH related)	--	--
Helicobacter pylori J99 Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.	ADH1t_HELPY_AAD07002 (CAD Fam: CADH related) ADH2t_HELPY_AAD06610 (CAD Fam: CADH related)	--	--
Helicobacter pylori (Campylobacter pylori). Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.	CADHt_HELPY_O25732 (CAD Fam: CADH related)	--	--
Bacillus subtilis. Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.	ADHAt_BACSU_O06007 (CAD Fam: CADH related)	ADHBt_BACSU_O06012 (PDH Fam: FADH) DHSO_BACSU (PDH Fam: DHSO) FADHt_BACSU_Q45604 (PDH Fam: FADH) TDH_BACSU (PDH Fam: TDH) YJMDt_BACSU_CAB13090 (PDH Fam: SSP) IDHt_BACSU_CAB12443 (PDH Fam: BDH)	YFMJt_BACSU_BAA22324 (LTD Fam: AADH/LHD related) YOGAt_BACSU_CAB1376 (QOR Fam: CCAR related) ADHt_BACSU_O07615 (QOR Fam)
Bacillus halodurans Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus	ADHtf_BACHA_BAB04257 (Y-ADH Fam: Fungi ADH related)	DHSOt_BACHA_BAB03906 (PDH Fam) DHSO2t_BACHA_BAA75341 (PDH Fam) IDHt_BACHA_BAB07668 (PDH Fam: BDH) UCPt_BACHA_BAB05670 (PDH Fam)	QORt_BACHA_BAB04082 (QOR Fam: QOR/ζ-crystallin related) QORt_BACHA_BAB04654 (QOR Fam: QOR/ζ-crystallin related) AGLt_BACHA_BAB04457 (QOR Fam: Alginate lyase)
Bacillus stearothermophilus. Firmicutes; bacillus/clostridium group; bacillaceae; bacillus.	ADH1_BACST (Y-ADH Fam: Fungi ADH related) ADH2_BACST (Y-ADH Fam: Fungi ADH related) ADH3_BACST (Y-ADH Fam: Fungi ADH related)		

Thermoanaerobacter ethanolicus Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter group; Thermoanaerobacter.		SADHt_THEET_AAB06720 (PDH Fam: secondary ADH)	
Thermoanaerobacter brockii **(Thermoanaerobium brockii)**. Firmicutes; Bacillus/clostridium group; Thermoanaerobacter group; Thermoanaerobacter.		ADH_THEBR (PDH Fam: secondary ADH)	
Mycoplasma genitalium. Firmicutes; Bacillus/Clostridium group; Mollicutes; Mycoplasmataceae; Mycoplasma.	--	--	--
Mycoplasma pneumoniae. Firmicutes; Bacillus/Clostridium group; Mollicutes; Mycoplasmataceae; Mycoplasma.	--	ADH_MYCPN (PDH Fam: secondary ADH)	--
Ureaplasma urealyticum. Firmicutes; Bacillus/Clostridium group; Mollicutes; Mycoplasmataceae; Ureaplasma	--	--	--
Clostridium beijerinckii Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.		ADH_CLOBE (PDH Fam: secondary ADH)	
Lactococcus lactis. Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus.			ADHt_LACLA_AAC56039 (QOR Fam: BRP) QORt_LACLA_AAK04815 (QOR Fam: alginate lyase)
Lactobacillus casei Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Lactobacillus.		SOREt_LACCA_AAF24128 (PDH Fam: SORE)	
Pediococcus pentosaceus. (Plasmid) Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Pediococcus.			QORt_PEDPE_AAD39631 (QOR Fam: BRP)
Amycolatopsis methanolica Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales;		FADH_AMYME (ADH Fam: FADH: Mycothiol-dependent)	
Corynebacterium glutamicum. Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.			RXA00288t_CORGL_CAC26028 (QOR Fam: alginate lyase) RXN01619t_CORGL_CAC26270 (NRBP Fam: NRBP related)

Corynebacterium sp. ST-10 Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae;	PARt_CORSP_BAA35108 (Y-ADH Fam: Broad ADH)		
Mycobacterium leprae Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae;			MASAt_MYCLE_AAA17069 (ER Fam: ER-PKS) PKSDt_MYCLE_AAA17357 (ER Fam: ER-PKS)
Mycobacterium tuberculosis Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	ADH_MYCTU (CAD Fam: CADH related) ADHAt_MYCTU_CAB06125 (Y-ADH Fam: Fungi ADH related)	ADHBt_MYCTU_CAB02404 (ADH Fam: Aryl/Alkyl ADH: Firmicutes) ADHDt_MYCTU_CAA16144 (ADH Fam: Aryl/Alkyl ADH: Firmicutes) ADHEt_MYCTU_CAB09757 (ADH Fam) ADHE2t_MYCTU_CAA17296 (ADH Fam: FADH: Mycothiol-dependent) ADH3t_MYCTU_O53904 (PDH Fam) ADH4t_MYCTU_O69693 (PDH Fam) RV1895t_MYCTU_CAB10037 (PDH Fam: secondary ADH)	FADB4t_MYCTU_CAB06274 (QOR Fam) FADB5t_MYCTU_CAB10029 (QOR Fam: PER actinomycetes) QORt_MYCTU_CAA15984 (QOR Fam: TED2 related) RV0149t_MYCTU_CAB07055 (QOR Fam) RV3777t_MYCTU_CAB02456 (QOR Fam: PIG 3 related) MASt_er-f_MYCTU_CAB06108 (QOR Fam: ER-PKS) PKS1t_er-f_MYCTU_CAB06103 (QOR Fam: ER-PKS) PKS2t_er-f_MYCTU_CAB10012 (QOR Fam: ER-PKS) PKS4t_er-f_MYCTU_CAA15858 (QOR Fam: ER-PKS) PKS5t_er-f_MYCTU_CAA17592 (QOR Fam: ER-PKS) PKS7t_er-f_MYCTU_CAB06632 (QOR Fam: ER-PKS) PKS8t_er-f_MYCTU_CAB09098 (QOR Fam: ER-PKS) PKS12t_er-fa_MYCTU_CAA17262 (QOR Fam: ER-PKS) PKS12t_er-fb_MYCTU_CAA17262 (QOR Fam: ER-PKS) PPSCt_er-f_MYCTU_CAB06099 (QOR Fam: ER-PKS)
Rhodococcus rhodochrous Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae;		ADHt_RHORH_O69045 (ADH Fam: Aryl/Alkyl ADH: Firmicutes)	

Saccharopolyspora erythraea (Streptomyces erythraeus). Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.			ERY2t_er-f_SACER_AAA26494 (ER Fam: ER-PKS)
Streptomyces caelestis Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			PKS4-5t_er-f_STRCA_AAC46026 (ER Fam: ER-PKS)
Streptomyces cinnamonensis Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			CCARt_STRCI_AAD53915 (QOR Fam: CCAR)
Streptomyces coelicolor A3(2) Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;	ADH1t_STRCO_CAB53423 (Y-ADH Fam: Fungi ADH related) ADH3t_STRCO_CAB53403-CAB55521 (Y-ADH Fam: Fungi ADH related) ORF3t_STRCO_BAA82700 (Y-ADH Fam: Broad ADH) ADHt_STRCO_CAB93031 (CAD Fam: CADH related)	ORFt_STRCO_CAA20560 (PDH Fam) ADHt_STRCO_CAB53146 (PDH Fam: FADH) ADH5t_STRCO_CAB39715 (ADH Fam: Benzyl ADH) ADH6t_STRCO_CAA22751 (ADH Fam: Benzyl ADH) ADH7t_STRCO_BAA34382 (ADH Fam: Mycothiol-dependent FADH) BDORt_STRCO_CAB58328 (PDH Fam: BDOR) DHSO1t_STRCO_CAB46780 (PDH Fam) DHSO2t_STRCO_CAB46402 (PDH Fam: SSP related) FADHt_STRCO_CAB46803 (PDH Fam) FADH2t_STRCO_CAB45580 (PDH Fam: FADH) ADHt_STRCO_CAB61162 (PDH Fam: SORE related) TDHt_STRCO_CAB71246 (PDH Fam: TDH) PDHt_STRCO_CAB61801 (PDH Fam) PDHt_STRCO_CAB62734 (PDH Fam)	ADHt_STRCO_CAB59716 (QOR Fam: pER actinomycetes) ADHt_STRCO_CAB70647 (QOR Fam: pER actinomycetes) ADHt_STRCO_CAB45348 (QOR Fam: pER actinomycetes) ADHt_STRCO_CAB52974 (QOR Fam: pER actinomycetes) CCARt_STRCO_CAA22721 (QOR Fam: CCAR) ORDt_STRCO_CAB62729 (LTD Fam: AADH/LHD related) ORDt_STRCO_CAB65629 (QOR Fam: pER actinomycetes) ORF2t_STRCO_CAA44234 (QOR Fam: pER actinomycetes) ORF4t_STRCO_CAA44236 (QOR Fam: pER actinomycetes) PORt_STRCO_CAB61611 (QOR Fam) QORt_STRCO_CAB46946 (QOR Fam: PIG 3 related) PKSt_er-f_STRCO_CAB71915 (ER Fam: ER-PKS)

Streptomyces collinus Firmicutes; Actinomycetes; Streptomycetes; Streptomycetaceae;			CCARt_STRCL_AAA92890 (QOR Fam: CCAR)
Streptomyces fradiae Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			CCARt_STRFR_CAA57474 (QOR Fam: CCAR) TS4-5t_STRFR_AAB66506 (ER Fam: ER-PKS)
Streptomyces griseus Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;	ADHt_STRGR_BAA32135 (CAD Fam: CADH related)		
Streptomyces lavendulae Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			ORFt_STRLA_AAD28446 (QOR Fam: CCAR related)
Streptomyces hygroscopicus Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			PKSt_er-fa_STRHY_CAA60460 (ER Fam: ER-PKS) PKSt_er-fb_STRHY_CAA60460 (ER Fam: ER-PKS) PKSt_er-fc_STRHY_CAA60460 (ER Fam: ER-PKS) PKSt_er-f_STRHY_CAA60459 (ER Fam: ER-PKS) PKSt_er-f_STRHY_CAA60462 (ER Fam: ER-PKS) PKS3t_er-f_STRHY_AAC38063 (ER Fam: ER-PKS)
Streptomyces lividans Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			PORt_STRLI_AAC25771 (QOR Fam: pER actinomycetes)
Streptomyces purpurascens Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			FASt_STRPU_CAA43934 (QOR Fam)
Streptococcus pyogenes. Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.	ADHt_STRPY_AAK33179 (Y-ADH Fam: Fungi ADH related)		

Streptomyces sp. MA6548. Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			PKSt_er-fa_STRSP_CAA71463 (ER Fam: ER-PKS) PKSt_er-fb_STRSP_CAA71463 (ER Fam: ER-PKS) PKSt_er-f_STRSP_AAC68815 (ER Fam: ER-PKS)
Streptomyces venezuelae Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae;			PKS1t_er-f_STRVE_AAC69330 (ER Fam: ER-PKS)
Synechocystis sp. (strain PCC 6803). Cyanobacteria; Chroococcales; Synechocystis.	ADH1t_SYNY3_BAA18840 (CAD Fam: CADH related)	ADH3_SYNY3 (ADH Fam: Class III ADH)	--
Anabaena azollae Cyanobacteria; Nostocales; Nostocaceae;		FADHt_ANAAZ_O24687 (ADH Fam: Class III ADH)	
Aquifex aeolicus Aquificales; Aquificaceae; Aquifex	--	--	ADH1t_AQUAE_O67374 (QOR Fam: CCAR related)

Fungi	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Aspergillus flavus. Ascomycota; Euascomycetes; Plectomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. **Aspergillus flavus.** Ascomycota; Eurotiales; Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.	ADH1_ASPFL (Y-ADH Fam: Fungi ADH)		
Aspergillus terreus. Ascomycota; Euascomycetes; Plectomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.			ENRt_ASPTE_Q9Y7D0 (QOR Fam: PKS-IAP) PKSt_er-f_ASPTE_AAD34559 (ER Fam: ER-iterative PKS)
EMERICELLA NIDULANS (ASPERGILLUS NIDULANS). ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES; EUROTIALES; TRICHOCOMACEAE; EMERICELLA **Emericella nidulans** (Aspergillus nidulans). Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.	ADH1_EMENI (Y-ADH Fam: Fungi ADH) ADH2_EMENI (Y-ADH Fam: Fungi ADH related) ADH3_EMENI (Y-ADH Fam: Fungi ADH)		
Trichoderma harzianum Ascomycota; Euascomycetes; Pyrenomycetes; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Trichoderma.			IC11_TRIHA (QOR Fam)
Fusarium sporotrichioides Eukaryotae; mitochondrial eukaryotes; Ascomycota; Euascomycetes; Pyrenomycetes; Hypocreales; Nectriaceae; Fusarium.	PN0170t_FUSSP_PN0170 (Y-ADH Fam: Not included: 12 aa)		
Botryotinia fuckeliana Ascomycota; Leotiales; Sclerotiniaceae; Botryotinia. /note="organism name synonym: **Botrytis cinerea**"			BCLHHt_BOTFU_AAC24957 (LTD Fam: AADH/LHD related) PKSt_BOTFU_AAC24951 (ER Fam: ER-iterative PKS)
Cochliobolus heterostrophus. Ascomycota; Pleosporales; Pleosporaceae; Cochliobolus.			PKSt_COCHE_AAB08104 (ER Fam: ER-iterative PKS)

Cochliobolus carbonum Ascomycota; Euascomycetes; Loculoascomycetes; Dothideales; Pleosporaceae; Cochliobolus.			CPHCTt_COCCA_CAA63129 (QOR Fam: PKS-IAP)
Candida parapsilosis Ascomycota; Hemiascomycetes; Saccharomycetales; Candidaceae; Candida.	SADHt_CANPA_BAA24528 (Y-ADH Fam: Fungi secondary ADH)		
Candida albicans Ascomycota; Hemiascomycetes; Saccharomycetales; Candidaceae; Candida.	ADH1_CANAL (Y-ADH Fam: Fungi ADH) ADH2t_CANAL_CAA21988 (Y-ADH Fam: Fungi ADH)		
Galactocandida mastotermitis. Ascomycota; Hemiascomycetes; Saccharomycetales; Candidaceae; Galactocandida. **Candida sp. HA167.** Ascomycota; Saccharomycetes; Saccharomycetales; anamorphic Saccharomycetales; Candida.		XDHt_GALMA_AAC24597 (PDH Fam: DHSO)	
Candida maltosa (Yeast). Ascomycota; Saccharomycetes; Saccharomycetales; anamorphic Saccharomycetales; Candida.		FADH_CANMA (ADH Fam: Class III ADH)	
KLUYVEROMYCES LACTIS (YEAST). ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES; SACCHAROMYCETACEAE; KLUYVEROMYCES. **Kluyveromyces lactis** Ascomycota; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.	ADH1_KLULA (Y-ADH Fam: Fungi ADH) ADH2_KLULA (Y-ADH Fam: Fungi ADH) ADH3_KLULA (Y-ADH Fam: Fungi ADH) ADH4_KLULA (Y-ADH Fam: Fungi ADH)		
OS **KLUYVEROMYCES MARXIANUS** (YEAST) **(KLUYVEROMYCES FRAGILIS).** ASCOMYCOTINA; HEMIASCOMYCETES.	ADH1_KLUMA (Y-ADH Fam: Fungi ADH: not included: considered redundant) ADH3t_KLUMA_S17252 (Y-ADH Fam: Fungi ADH: not included: considered redundant)		

Pichia pastoris (Yeast). Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.		FADHt_PICPA_O74685 (ADH Fam: Class III ADH)	
Pichia stipitis (Yeast). Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.	ADH1_PICST (Y-ADH Fam: Fungi ADH) ADH2_PICST (Y-ADH Fam: Fungi ADH)	XYL2_PICST (PDH Fam: DHSO)	
Saccharomyces bayanus Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.	ADH2tf_SACBE_AAF43642 (Y-ADH Fam: Fungi ADH: not included: 58 aa)		
SACCHAROMYCES CEREVISIAE (BAKER'S YEAST). ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES; SACCHAROMYCETACEAE; SACCHAROMYCES. **Saccharomyces cerevisiae** Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.	ADH1_YEAST (Y-ADH Fam: Fungi ADH) ADH2_YEAST (Y-ADH Fam: Fungi ADH) ADH3_YEAST (Y-ADH Fam: Fungi ADH) ADH5_YEAST (Y-ADH Fam: Fungi ADH) YCZ5_YEAST (CAD Fam: CADH related) YM97_YEAST (CAD Fam: CADH related)	YAG0_YEAST (PDH Fam: BDH) YAG1_YEAST (PDH Fam: BDH) DHSO_YEAST (PDH Fam: DHSO) PSRt_YEAST_S64902 (PDH Fam: DHSO) FADH_YEAST (ADH Fam: Class III ADH) YDL246Ct_YEAST_CAA98826 (PDH Fam: DHSO)	QOR_YEAST (QOR Fam: TED2 related) AST1_YEAST (QOR Fam: AST) AST2_YEAST (QOR Fam: AST) YCZ2_YEAST (QOR Fam: PKS-IAP) YL60_YEAST (QOR Fam: PKS-IAP) MRF1_YEAST (NRBP Fam: NRBP related) YM27_YEAST (QOR Fam: AST) YMN1_YEAST (LTD Fam: AADH/LHD related) YNN4_YEAST (QOR Fam: PKS-IAP)
Saccharomyces kluyveri. Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.	ADH3t_SACKL_AAF43645 (Y-ADH Fam: Fungi ADH)		
Yarrowia lipolytica Ascomycota; Hemiascomycetes; Saccharomycetales; Dipodascaceae; Yarrowia.	ADH1t_YARLI_AAD51737 (Y-ADH Fam: Fungi ADH) ADH2t_YARLI_AAD51738 (Y-ADH Fam: Fungi ADH) ADH3t_YARLI_AAD51739 (Y-ADH Fam: Fungi ADH)		

Gibberella fujikuroi. Ascomycota; Pyrenomycetes; Hypocreales; Hypocreaceae; Gibberella.			PKSt_GIBFU_AAD43562 (ER Fam: ER-iterative PKS)
Neurospora crassa Ascomycota; Sordariales; Sordariaceae; Neurospora.	ADH1t_NEUCR_CAB91241 (Y-ADH Fam: Fungi ADH)		ORFt_NEUCR_CAB91421 (QOR Fam: 77 aa)
SCHIZOSACCHAROMYCES POMBE (FISSION YEAST). ASCOMYCOTA; ARCHIASCOMYCETES; SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE; SCHIZOSACCHAROMYCES.	ADH_SCHPO (Y-ADH Fam: Fungi ADH related)	DHSO_SCHPO (PDH Fam: DHSO) FADH_SCHPO (ADH Fam: Class III ADH) FADH2t_SCHPO_O74540 (ADH Fam: Class III ADH)	ADH2t_SCHPO_CAA21911 (QOR Fam: CCAR related) QORt_SCHPO_CAA21450 (QOR Fam: TED 2 related) ADHt_SCHPO_CAA21281 (QOR Fam) MRF1_SCHPO (NRBP Fam: NRBP related) QORt_SCHPO_CAA16853 (QOR Fam: DI-QOR/ARP related)
Filobasidiella neoformans var. neoformans Basidiomycota; Hymenomycetes; Tremellales; Tremellaceae; Filobasidiella.	MTDt_FILNE_AAG17013 (Y-ADH Fam: Fungi MTD)		

Metazoa	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Alligator mississippiensis (American Alligator). Chordata; Vertebrata; Tetrapoda; Reptilia; Archosauria.		ADH1_ALLMI (ADH Fam: Animal ADH)	
Apteryx australis (Brown Kiwi). Chordata; Vertebrata; Tetrapoda; Aves; Palaeognathae; Apterygiformes.		ADH1_APTAU (ADH Fam: Animal ADH)	
Struthio camelus (Ostrich). Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae; Struthio.		ADH1_STRCA (ADH Fam: Animal ADH) ADH2_STRCA (ADH Fam: Animal ADH)	
Gallus gallus (Chicken). Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		ADH1_CHICK (ADH Fam: Animal ADH) ADHFt_CHICK_O42483 (ADH Fam: Animal ADH)	FAS_er-f_CHICK (ER Fam: ER-FAS)
Coturnix coturnix japonica (Japanese Quail). Chordata; Vertebrata; Tetrapoda; Aves; Neognathae; Galliformes.		ADH3_COTJA (ADH Fam: Animal ADH)	
Anas platyrhynchos (Domestic Duck). Chordata; Vertebrata; Tetrapoda; Aves; Neognathae; Anseriformes.		ADH_ANAPL (ADH Fam: Animal ADH)	
Naja naja (Indian Cobra). Chordata; Vertebrata; Tetrapoda; Reptilia; Lepidosauria; Serpentes.		ADH1_NAJNA (ADH Fam: Animal ADH)	
Uromastyx hardwickii (Indian spiny-tailed lizard). Chordata; Craniata; Vertebrata; Euteleostomi; Lepidosauria; Squamata; Iguania; Acrodonta; Agamidae; Leiolepinae; Uromastyx.		ADHA_UROHA (ADH Fam: Animal ADH) ADHB_UROHA (ADH Fam: Animal ADH) ADHX_UROHA (ADH Fam: Class III ADH)	

Oryctolagus cuniculus. Chordata; Vertebrata; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		ADH1_RABIT (ADH Fam: Animal ADH) ADHX_RABIT (ADH Fam: Class III ADH) ADHP_RABIT (ADH Fam: Animal ADH) ADHQ_RABIT (ADH Fam: Animal ADH)	LB4D_RABIT (LTD Fam: LHD)
Peromyscus maniculatus (Deer mouse). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae; Peromyscus.		ADHA_PERMA (ADH Fam: Animal ADH) ADH2_PERMA (ADH Fam: Animal ADH)	
Mus musculus (Mouse). Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		ADHA_MOUSE (ADH Fam: Animal ADH) ADHX_MOUSE (ADH Fam: Class III ADH) ADH7_MOUSE (ADH Fam: Animal ADH) DHSO_MOUSE (PDH Fam: DHSO)	QOR_MOUSE (QOR Fam: ζ-crystallin/QOR) VAT1_MOUSE (QOR Fam: VAT1) NRBPt_MOUSE_AAH03864 (NRBP Fam: NRBP)
Rattus norvegicus (Rat). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		ADHA_RAT (ADH Fam: Animal ADH) ADHX_RAT (ADH Fam: Class III ADH) ADH6t_RAT_Q64160 (ADH Fam: Animal ADH, not included: fragment 37 aa) ADH7_RAT (ADH Fam: Animal ADH) ADH2_RAT_CAA62241 (ADH Fam: Animal ADH) DHSO_RAT (PDH Fam: DHSO)	DTElt_RAT_AAB88912 (LTD Fam: LHD) NRBPt_RAT_BAA34804 (NRBP Fam: NRBP) FAS_er-f_RAT (ER Fam: ER-FAS)
Geomys bursarius (Plains pocket gopher). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Geomyidae; Geomys.		ADH1t_GEOBU_2147094 (ADH Fam: Animal ADH)	
Geomys knoxjonesi. Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Geomyidae; Geomys.		ADH1t_GEOKN_2147096 (ADH Fam: Animal ADH)	

Geomys attwateri. Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Geomyidae; Geomys.		ADH1t_GEOAT_Q9Z2M2 (ADH Fam: Animal ADH, not included: considered redundant)	
Geomys texensis. Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Geomyidae; Geomys.		ADH1t_GEOTE_Q9Z2M1 (ADH Fam: Animal ADH, not included: considered redundant)	
Cavia porcellus (domestic guinea pig). Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.			QOR_CAVPO (QOR Fam: ζ-crystallin/QOR)
Callithrix sp. (Marmoset). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.		DHSOt_CALSP_AAB69288 (PDH Fam: DHSO)	
Homo sapiens (Human). Chordata; Vertebrata; Tetrapoda; Mammalia; Oc Eutheria; Primates.		ADHA_HUMAN (ADH Fam: Animal ADH) ADHB_HUMAN (ADH Fam: Animal ADH) ADHG_HUMAN (ADH Fam: Animal ADH) ADHP_HUMAN (ADH Fam: Animal ADH) ADH6_HUMAN (ADH Fam: Animal ADH) ADH7_HUMAN (ADH Fam: Animal ADH) ADHX_HUMAN (ADH Fam: Class III ADH) DHSO_HUMAN (PDH Fam: DHSO)	QOR_HUMAN (QOR Fam: ζ-crystallin/QOR) QORL_HUMAN (QOR Fam: QORL-1) LHDt_HUMAN_BAA08382 (LTD Fam: LHD) PIG3t_HUMAN_AAC39528 (QOR Fam: PIG3) UPPt_HUMAN_BAA91605 (QOR Fam: QORL-1) VAT1_HUMAN (QOR Fam: VAT1) VAT1t_HUMAN_AAA93230 (QOR Fam: VAT1) CGI63t_HUMAN_AAD34058 (NRBP Fam: NRBP) FAS_er-f_HUMAN (ER Fam: ER-FAS) SCGI-63t_HUMAN_AAH01419 (NRBP Fam: NRBP)
Macaca mulatta (Rhesus macaque) Chordata; Vertebrata; Tetrapoda; Mammalia; Eutheria; Primates.		ADH_MACMU (ADH Fam: Animal ADH)	

Papio hamadryas (Hamadryas baboon). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Papio.		ADH_PAPHA (ADH Fam: Animal ADH) ADHGt_PAPHA_O97959 (ADH Fam: Animal ADH)	
Equus caballus (Horse). Chordata; Vertebrata; Tetrapoda; Mammalia; Eutheria; Perissodactyla.		ADHE_HORSE (ADH Fam: Animal ADH) ADHS_HORSE (ADH Fam: Animal ADH) ADHX_HORSE (ADH Fam: Class III ADH)	
Ovis aries (sheep) Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Caprinae; Ovis.		DHSO_SHEEP (PDH Fam: DHSO)	
Bos taurus. Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.			CRYZt_BOVIN_AAD10290 (QOR Fam: ζ-crystallin/QOR) QORt_BOVIN_631585 (QOR Fam: ζ-crystallin/QOR, not included, possible allelic form)
Sus scrofa (Pig). Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		DHSOtf_PIG_Q29318 (PDH Fam: DHSO)	LHDt_PIG_BAA08381 (LTD Fam: LHD)
Lama guanicoe (guanaco). Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Lama.			QOR_LAMGU (QOR Fam: ζ-crystallin/QOR)
Rana perezi (Perez's frog) (Western Mediterranian green frog). Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.		ADH1_RANPE (ADH Fam: Animal ADH) ADH4_RANPE (ADH Fam: Animal ADH)	
Xenopus laevis (African clawed frog). Chordata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.		ADH1tf_XENLE_3511118 (ADH Fam: Animal ADH) ADH4tf_XENLE_3511120 (ADH Fam: Animal ADH)	
Gadus callarias (Baltic Cod). Chordata; Vertebrata; Pisces; Gnathostomata; Osteichthyes; Actinopterygii; Gadiformes.		ADH_GADCA (ADH Fam: Animal ADH)	

Gadus morhua (Atlantic cod). Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae; Gadus.		ADHH_GADMO (ADH Fam: Class III ADH) ADHL_GADMO (ADH Fam: Class III ADH)	
Sparus aurata (gilthead sea bream) Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha; Perciformes; Percoidei; Sparidae; Sparus.		ADHX_SPAAU (ADH Fam: Class III ADH)	
Torpedo californica (Pacific electric ray) Metazoa; Chordata; Vertebrata; Chondrichthyes; Elasmobranchii; Rajiformes; Torpedinoidei; Torpedinidae; Torpedo.			VAT1_TORCA (QOR Fam: VAT1)
Myxine glutinosa (Atlantic hagfish). Chordata; Vertebrata; Pisces; Agnatha (Cyclostomata).		ADHX_MYXGL (ADH Fam: Class III ADH)	
Branchiostoma lanceolatum (amphioxus). Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.		ADHXt_BRALA_AAF73255 (ADH Fam: Class III ADH)	
Branchiostoma floridae (Florida lancelet). Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.		ADHXt_BRAFL_AAF73254 (ADH Fam: Class III ADH)	
Drosophila melanogaster (Fruit Fly). Arthropoda; Insecta; Diptera.	ALT1t_DROME_AAF59234 (Macrofam II?)	ADHX_DROME (ADH Fam: Class III ADH) DHSO1t_DROME_AAD00903 (PDH Fam: DHSO) DHSO2t_DROME_AAD00902 (PDH Fam: DHSO) CG4836t_DROME_AAF55731 (PDH Fam)	GH07626t_er-f_DROME_AAF51148 (ER Fam: ER-FAS) CG3524t_er-f_DROME_AAF51149 (ER Fam: ER-FAS) CG17374t_er-f_DROME_AAF45403 (ER Fam: ER-FAS) CG16935t_DROME_AAF58322 (NRBP Fam: NRBP)
silkworm. **Bombyx mori** Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Bombycoidea; Bombycidae; Bombyx.		DHSO_BOMMO (PDH Fam: DHSO)	P270t_BOMMO_AAB53258 (ER Fam: ER-FAS) P260t_BOMMO_AAB53257 (ER Fam: ER-FAS)

Bemisia argentifolii **silverleaf whitefly.** Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aleyrodiformes; Aleyrodoidea; Aleyrodidae; Aleyrodinae; Bemisia.		KERt_BEMAR_AAD02817 (PDH Fam: DHSO)	
OCTOPUS VULGARIS (OCTOPUS). MOLLUSCA; CEPHALOPODA; COLEOIDEA; OCTOPODA.		ADHX_OCTVU (ADH Fam: Class III ADH)	
Caenorhabditis elegans. Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.	ADH1_CAEEL (Y-ADH Fam: Fungi ADH related) ADH2_CAEEL (Y-ADH Fam: Fungi ADH related) ADH3t_CAEEL_AAD14690 (Y-ADH Fam: Fungi ADH related) ZK829t_CAEEL_Q23624 (Macrofam II?)	ADHX_CAEEL (ADH Fam: Class III ADH) DHSO1t_CAEEL_CAA94842 (PDH Fam: DHSO) DHSO2t_CAEEL_CAA94841 (PDH Fam: DHSO)	ADHt_CAEEL_CAB07384 (QOR Fam: ζ-crystallin/QOR) FASt_er-f_CAEEL_CAB04244 (ER Fam: ER-FAS) NRBPt_CAEEL_CAB04958 (NRBP Fam: NRBP) M106t_CAEEL_Q09593 (LTD Fam) ORFt_CAEEL_CAA19533 (NRBP Fam: NRBP related)

Metaphita	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Arabidopsis thaliana (thale cress) Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	CAD1_ARATH (CAD Fam: ELI3) MTD1_ARATH (CAD Fam: ELI3) MTD2_ARATH (CAD Fam: ELI3) CAD4_ARATH (CAD Fam: CADH) CAD5t_ARATH_AAD20393 (CAD Fam: ELI3) CAD6t_ARATH_CAA17549 (CAD Fam: CADH) LCADAt_ARATH_CAA76418 (CAD Fam: ELI3) CAD3t_ARATH_AAD20406 (CAD Fam: ELI3) CADt_ARATH_AAF21160 (CAD Fam: CADH related)	ADHt_ARATH_BAB10455 (ADH Fam: Plant ADH related)) ADH1_ARATH (ADH Fam: Plant ADH) ADHX_ARATH (ADH Fam: Class III ADH) ADH4t_ARATH_O65459 (ADH Fam: Plant ADH) DHSOt_ARATH_BAB11045 (PDH Fam: DHSO)	AADHt_ARATH_CAB75803 (LTD Fam: AADH) ADHt_ARATH_CAB45500 (QOR Fam: DI-QOR) ADHt_ARATH_AAF26116 (LTD Fam: AADH) ARPt_ARATH_CAA89858 (QOR Fam: QOR-bidomain) AIPt_ARATH_AAC98029 (QOR Fam: ARP) NRBPt_ARATH_CAB75790 (NRBP Fam: NRBP) QOH3t_ARATH_CAA89262 (LTD Fam: AADH) QOH4t_ARATH_CAA89838 (LTD Fam: AADH) CRYZt_ARATH_AAB60917 (LTD Fam: AADH) QORt_ARATH_O65423 (QOR Fam: PIG3 related)
Leavenworthia uniflora. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Leavenworthia.		ADH1tf_LEAUN_O49109 (ADH Fam: Plant ADH, not included: considered redundant) ADH2tf_LEAUN_O50049 (ADH Fam: Plant ADH, not included: considered redundant) ADH3tf_LEAUN_O49113 (ADH Fam: Plant ADH, not included: considered redundant)	
Leavenworthia stylosa. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Leavenworthia.		ADH1tf_LEAST_O49115 (ADH Fam: Plant ADH) ADH2t_LEAST_O49110 (ADH Fam: Plant ADH) ADH3t_LEAST_O49112 (ADH Fam: Plant ADH)	

Leavenworthia crassa. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Leavenworthia.		ADH1tf_LEACR_O49114 (ADH Fam: Plant ADH, not included: considered redundant) ADH2tf_LEACR_O50060 (ADH Fam: Plant ADH, not included: considered redundant) ADH3tf_LEACR_O49111 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabis gemmifera. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabis.		ADHt_ARAGE_O23817 (ADH Fam: Plant ADH)	
Arabis flagellosa. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabis.		ADHtf_ARAFL_Q9ZWL2 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabis hirsuta. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabis.		ADHtf_ARAHR_Q9ZWL0 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabis glabra (tower rockcress). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabis.		ADHtf_ARAGL_Q9ZWL3 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabis stelleri. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabis.		ADHtf_ARAST_Q9ZWL4 (ADH Fam: Plant ADH, not included: considered redundant)	

Arabidopsis lyrata subsp. kawasakiana. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARALY_Q9ZWL1 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabidopsis suecica. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARASU_Q9ZWK5 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabidopsis wallichii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARAWA_Q9ZWK6 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabidopsis korshinskyi. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARAKO_Q9ZWK7 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabidopsis griffithiana. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARAGR_Q9ZWK8 (ADH Fam: Plant ADH, not included: considered redundant)	
Arabidopsis himalaica. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		ADHtf_ARAHI_Q9ZWK9 (ADH Fam: Plant ADH, not included: considered redundant)	
Brassica napus (rape). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.	CADHA1tf_BRANA_AAF23409 (CAD Fam: CADH) CADHA2tf_BRANA_AAF23410 (CAD Fam: CADH)		

Brassica oleracea. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.	CADHAtf_BRAOL_AAF23411 (CAD Fam: CADH)	ADHtf_BRAOL_Q9ZWK4 (ADH Fam: Plant ADH)	
Brassica rapa Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.	CADHAtf_BRARA_AAF23412 (CAD Fam: CADH: not included, considered redundant)		
Eucalyptus gunnii (cider tree) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Myrtales; Myrtaceae; Eucalyptus.	CAD1t_EUCGU_S60242 (CAD Fam: CADH) CAD2_EUCGU (CAD Fam: CADH)		
Eucalyptus botryoides. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Myrtales; Myrtaceae; Eucalyptus.	CADH_EUCBO (CAD Fam: CADH)		
Eucalyptus globulus (Blue gum). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Myrtales; Myrtaceae; Eucalyptus.	CADH_EUCGL (CAD Fam: CADH)		
Gossypium robinsonii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHCtf_GOSRO_O49051 (ADH Fam: Plant ADH, not included: considered redundant) ADHDtf_GOSRB_O65899 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium raimondii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHCtf_GOSRA_O49052 (ADH Fam: Plant ADH, not included: considered redundant)	

Gossypium hirsutum (Upland cotton). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHAtf_GOSHI_O82791 (ADH Fam: Plant ADH) ADHCtf_GOSHI_O49058 (ADH Fam: Plant ADH) ADH2At_GOSHI_Q39782 (ADH Fam: Plant ADH) ADH2Bt_GOSHI_Q39783 (ADH Fam: Plant ADH) ADH2Dt_GOSHI_Q42763 (ADH Fam: Plant ADH) ADH1tf_GOSHI_Q42764 (ADH Fam: Plant ADH) ADH2Ctf_GOSHI_Q42765 (ADH Fam: Plant ADH)	
Gossypium barbadense (Sea-island cotton) (Egyptian cotton). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHAtf_GOSBA_O82792 (ADH Fam: Plant ADH) ADHCtf_GOSBA_O49061 (ADH Fam: Plant ADH)	
Gossypium tomentosum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHCtf_GOSTO_O49059 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium mustelinum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHCtf_GOSMU_O49060 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium darwinii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHCtf_GOSDA_O49062 (ADH Fam: Plant ADH, not included: considered redundant)	

Gossypium bickii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSBI_O65284 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium australe. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSAU_O65288 (ADH Fam: Plant ADH)	
Gossypium populifolium. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSPO_O65298 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium costulatum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSCO_O65290 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium nobile. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSNO_O65294 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium pilosum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSPI_O65909 (ADH Fam: Plant ADH, not included: considered redundant)	

Gossypium pulchellum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSPU_O65302 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium rotundifolium. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSRO_O65303 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium herbaceum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSHE_O65304 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium marchantii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSMA_O65293 (ADH Fam: Plant ADH)	
Gossypium sturtianum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSST_O65903 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium londonderriense. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSLO_O65911 (ADH Fam: Plant ADH, not included: considered redundant)	

Gossypium exiguum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSEX_O65912 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium enthyle. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSEN_O65913 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium cunninghamii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSCU_O65926 (ADH Fam: Plant ADH, not included: considered redundant)	
Gossypium nelsonii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		ADHDtf_GOSNE_O65928 (ADH Fam: Plant ADH, not included: considered redundant)	
Pisum sativum (Garden pea). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Pisum.		ADH1_PEA (ADH Fam: Plant ADH) ADHX_PEA (ADH Fam: Class III ADH)	
Trifolium repens (Creeping white clover). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolium.		ADH1_TRIRP (ADH Fam: Plant ADH)	
Stylosanthes humilis (Townsville stylo). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Stylosanthes.	CAD1_STYHU (CAD Fam: ELI3) CAD3_STYHU (CAD Fam: ELI3)		

Medicago sativa Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.	CADH_MEDSA (CAD Fam: CADH) CAD2t_MEDSA_O82515 (CAD Fam: ELI3)		
Vigna unguiculata (cowpea) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Fabales; Fabaceae; Papilionoideae; Vigna.			TED2t_VIGUN_CAA69914 (QOR Fam: TED2 related)
Vigna radiata (mung bean) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Vigna.			AIPt_VIGRA_AAA87182 (QOR Fam: ARP)
Phaseolus acutifolius (tepary bean). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolus.		ADHt_PHAAC_Q43016 (ADH Fam: Plant ADH)	
Glycine max (Soybean). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.		ADH1tf_GLYMA_O82478 (ADH Fam: Plant ADH) ADH2t_GLYMA_Q9ZT38 (ADH Fam: Plant ADH)	
Cicer arietinum (Chickpea) (Garbanzo). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Cicer.			ORDt_CICAR_BAA78050 (QOR Fam: QOR-bidomain)
Fragaria ananassa (Strawberry). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Rosales; Rosaceae; Fragaria.	MTD_FRAAN (CAD Fam: ELI3) ADHt_FRAAN_CAC09053 (CAD Fam: ELI3)	ADH_FRAAN (ADH Fam: Plant ADH)	

Malus domestica (apple tree) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Rosales; Rosaceae; Malus.		ADH_MALDO (ADH Fam: Plant ADH) DHSOt_MALDO_Q9ZR22 (PDH Fam: DHSO)	
Pyrus communis (Pear). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Rosales; Rosaceae; Pyrus.		ADH3tf_PYRCO_O22649 (ADH Fam: Plant ADH) ADH4tf_PYRCO_O22650 (ADH Fam: Plant ADH)	
Prunus persica (peach) Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Rosales; Rosaceae; Prunus.		DHSOt_PRUPE_BAA94084 (PDH Fam: DHSO)	
Fragaria vesca Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Rosales; Rosaceae; Fragaria.			RIPt_FRAVE_CAA04767 (QOR Fam: ARP)
Eriobotrya japonica Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Rosales; Rosaceae; Eriobotrya.		DHSOt_ERIJA_BAA95897 (PDH Fam: DHSO)	
Populus deltoides (Poplar). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.	CADH_POPDE (CAD Fam: CADH)		
Linum usitatissimum (Flax) (Linseed). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Linaceae; Linum.		MNLt_LINUS_P93243 (ADH Fam: HNL)	
Petunia hybrida (Petunia). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.		ADH1_PETHY (ADH Fam: Plant ADH) ADH2tf_PETHY_Q07321 (ADH Fam: Plant ADH)	

Solanum tuberosum (Potato). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		ADH1_SOLTU (ADH Fam: Plant ADH) ADH2_SOLTU (ADH Fam: Plant ADH) ADH3_SOLTU (ADH Fam: Plant ADH) ADH4t_SOLTU_Q43169 (ADH Fam)	
Lycopersicon esculentum (tomato) Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.	ELI3t_LYCES_AAF72100 (CAD Fam: ELI3)	ADH2_LYCES (ADH Fam: Plant ADH) ADH3Btf_LYCES_Q41242 (ADH Fam: Plant ADH) ADH3Atf_LYCES_Q41241 (ADH Fam: Plant ADH)	
Nicotiana tabacum (common tobacco) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.	CAD4_TOBAC (CAD Fam: CADH) CAD9_TOBAC (CAD Fam: CADH)	ADH1t_TOBAC_Q42953 (ADH Fam: Plant ADH)	AADHt_TOBAC_BAA89423 (LTD Fam: AADH)
Lithospermum erythrorhizon Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; unclassified euasterids I; Boraginaceae; Lithospermum.			LEDI4t_LITER_BAA83082 (QOR Fam: DI-QOR)
Triphysaria versicolor. Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Lamiales; Orobanchaceae; Triphysaria.			QOR1t_TRIVE_AAG53944 (QOR Fam: DI-QOR)
Aralia cordata. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Apiales; Araliaceae; Aralia.	CAD1_ARACO (CAD Fam: CADH)		

Petroselinum crispum (Parsley) (Petroselinum hortense). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.	MTD_PETCR (CAD Fam: ELI3)		
Apium graveolens. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Apium.	MTDt_APIGR_AAC15467 (CAD Fam: ELI3)		
Zinnia elegans. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Zinnia.	CAD1tf_ZINEL_O04079 (CAD Fam: CADH)		TED2t_ZINEL_BAA06460 (QOR Fam: TED2)
Lactuca sativa (Garden lettuce). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.		GRPt_LACSA_Q40249 (ADH Fam: Plant ADH)	
Paeonia lutea. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADHIAtf_PAELU_O22352 (ADH Fam: Plant ADH, not included: considered redundant) ADH2tf_PAELU_O22358 (ADH Fam: Plant ADH, not included: considered redundant) ADH1Btf_PAELU_O24269 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia rockii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH1Atf_PAERO_O22353 (ADH Fam: Plant ADH) ADH1Btf_PAERO_O24283 (ADH Fam: Plant ADH) ADH2tf_PAERO_O22362 (ADH Fam: Plant ADH)	

Paeonia anomala. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH1Atf_PAEAN_O22354 (ADH Fam: Plant ADH, not included: considered redundant) ADH2tf_PAEAN_O22364 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia delavayi. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH1Btf_PAEDE_O22355 (ADH Fam: Plant ADH, not included: considered redundant) ADH2tf_PAEDE_O22359 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia brownii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAEBR_O22356 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia californica. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAECA_O22357 (ADH Fam: Plant ADH, not included: considered redundant) ADH1Atf_PAECA_O24250 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia suffruticosa subsp. spontanea. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH1Atf_PAESU_O24289 (ADH Fam: Plant ADH) ADH1Btf_PAESU_O24291 (ADH Fam: Plant ADH) ADH2tf_PAESU_O22360 (ADH Fam: Plant ADH)	
Paeonia szechuanica. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAESZ_O22361 (ADH Fam: Plant ADH, not included: considered redundant) ADH1Atf_PAESZ_O24290 (ADH Fam: Plant ADH, not included: considered redundant)	

Paeonia tenuifolia. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAETE_O22365 (ADH Fam: Plant ADH, not included: considered redundant) ADH1Atf_PAETE_O24311 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia veitchii. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAEVE_O22366 (ADH Fam: Plant ADH, not included: considered redundant) ADH1Atf_PAEVE_O24317 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia lactiflora (Chinese peony). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH1Atf_PAELA_O24268 (ADH Fam: Plant ADH, not included: considered redundant)	
Paeonia humilis. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Saxifragales; Paeoniaceae; Paeonia.		ADH2tf_PAEHU_AAF37598 (ADH Fam: Plant ADH)	
Vitis vinifera (Grape). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.		ADH1t_VITVI_AAB65840 (ADH Fam: Plant ADH) ADH2t_VITVI_AAG01382 (ADH Fam: Plant ADH) ADH3t_VITVI_AAG01383 (ADH Fam: Plant ADH) ADH7t_VITVI_AAF44336 (ADH Fam: Plant ADH)	
Mesembryanthemum crystallinum (Common ice plant). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.	MTD_MESCR (CAD Fam: ELI3)		

Hordeum vulgare (Barley). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.	CADtf_HORVU_O23981 (CAD Fam: CADH)	ADH1_HORVU (ADH Fam: Plant ADH) ADH2_HORVU (ADH Fam: Plant ADH) ADH3_HORVU (ADH Fam: Plant ADH)	
Hordeum spontaneum (Barley). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.		ADH1tf_HORSP_O65114 (ADH Fam: Plant ADH, not included: considered redundant)	
Zea mays (Maize). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.	CADH_MAIZE (CAD Fam: CADH)	ADH1_MAIZE (ADH Fam: Plant ADH) ADH2_MAIZE (ADH Fam: Plant ADH) ADHX_MAIZE (ADH Fam: Class III ADH)	
Oryza sativa (Rice). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.		ADH1_ORYSA (ADH Fam: Plant ADH) ADH2_ORYSA (ADH Fam: Plant ADH) ADHX_ORYSA (ADH Fam: Class III ADH)	
Pennisetum americanum (Pearl millet). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pennisetum.		ADH1_PENAM (ADH Fam: Plant ADH)	
Zea luxurians (Teosinte). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.		ADH1_ZEALU (ADH Fam: Plant ADH)	
Lolium perenne Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Lolium.	CADH_LOLPR (CAD Fam: CADH)		
Tripsacum dactyloides (Gama grass). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Tripsacum.		ADHtf_TRIDA_O49214 (ADH Fam: Plant ADH)	

Arundo donax (giant reed). Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Arundo.		ADH1tf_ARUDO_O24501 (ADH Fam: Plant ADH)	
Muhlenbergia setarioides. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Muhlenbergia.		ADH1tf_MUHSE_O24503 (ADH Fam: Plant ADH)	
Anomochloa marantoidea. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Anomochloa.		ADH1tf_ANOMA_O24504 (ADH Fam: Plant ADH) ADH2tf_ANOMA_O24507 (ADH Fam: Plant ADH)	
Lithachne humilis. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Lithachne.		ADH1tf_LITHU_O24506 (ADH Fam: Plant ADH)	
Bambusa glaucescens (Hedge bamboo) (Bambusa multiplex). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Bambusa.		ADH1tf_BAMGL_O24508 (ADH Fam: Plant ADH)	
Rhynchoryza subulata. Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Rhynchoryza.		ADH2tf_RHYSU_AAF37796 (ADH Fam: Plant ADH)	
Eragrostis japonica. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eragrostis.		ADH1tf_ERAJA_O24502 (ADH Fam: Plant ADH)	
Zea diploperennis (Perennial teosinte). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.		ADH1t_ZEADI_Q43238 (ADH Fam: Plant ADH)	

Pennisetum glaucum. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pennisetum.		ADH1t_PENGL_Q43722 (ADH Fam: Plant ADH)	
Saccharum officinarum (sugarcane) Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Saccharum.	CADHt_SACOF_CAA13177 (CAD Fam: CADH)		
Sorghum bicolor milo (Sorghum). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.		ADH1t_SORBI_O82430 (ADH Fam: Plant ADH)	
Leersia perrieri Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Leersia.		ADH2tf_LEEPE_AAF37797 (ADH Fam: Plant ADH)	
Joinvillea ascendens. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Joinvilleaceae; Joinvillea.		ADHtf_JOIAS_O24505 (ADH Fam: Plant ADH)	
Washingtonia robusta. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Arecaceae; Washingtonia.		ADHBt_WASRO_P93623 (ADH Fam: Plant ADH) ADHAtf_WASRO_P93624 (ADH Fam: Plant ADH) ADHCtf_WASRO_P93625 (ADH Fam: Plant ADH; not included: 46 aa)	
Calamus usitatus. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Arecaceae; Calamus.		ADHtf_CALUS_Q96404 (ADH Fam: Plant ADH)	

Phoenix reclinata. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Arecaceae; Phoenix.		ADHAtf_PHORE_O24284 (ADH Fam: Plant ADH)	
Dioscorea tokoro. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Dioscoreales; Dioscoreaceae; Dioscorea.		ADHtf_DIOTO_O24622 (ADH Fam: Plant ADH)	
Dioscorea tenuipes. Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Dioscoreales; Dioscoreaceae; Dioscorea.		ADH1tf_DIOTE_O22063 (ADH Fam: Plant ADH)	
Picea abies (Norway spruce) (Picea excelsa). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.	CADH_PICAB (CAD Fam: CADH) CAD2_PICAB (CAD Fam: CADH)		
Pinus radiata (Monterey pine). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.	CADH_PINRA (CAD Fam: CADH)		
Pinus taeda (Loblolly pine). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.	CADH_PINTA (CAD Fam: CADH)		
Pinus banksiana (Jack pine). Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.		ADHC1tf_PINBN_Q43020 (ADH Fam: Plant ADH) ADHC4tf_PINBN_Q43022 (ADH Fam: Plant ADH) ADHC7tf_PINBN_Q43025 (ADH Fam: Plant ADH) ADHC3tf_PINBN_Q43026 (ADH Fam: Plant ADH) ADHC5tf_PINBN_Q43027 (ADH Fam: Plant ADH) ADHC6tf_PINBN_Q43028 (ADH Fam: Plant ADH) ADHC2tf_PINBN_Q43300 (ADH Fam: Plant ADH)	

Protista	Macrofamily II MDR	Macrofamily I MDR	Macrofamily III MDR
Entamoeba histolytica. Entamoebidae; Entamoeba.		ADH1_ENTHI (PDH Fam: secondary ADH)	
Leishmania amazonensis Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.			QOR_LEIAM (QOR Fam: QOR-bidomain)
Leishmania major Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.	ADHt_LEIMA_CAB58398 (CAD Fam: CADH related)		ORDt_LEIMA_CAB55621 (QOR Fam: PIG3 related) QORt_LEIMA_AAF35929 (QOR Fam: TED2 related) NRBPt_LEIMA_CAC22680 (NRBP Fam: NRBP related) 1302.12tf_LEIMA_CAC00370 (NRBP Fam: NRBP related)
Cryptosporidium parvum. Alveolata; Apicomplexa; Coccidia; Eimeriida; Cryptosporidiidae; Cryptosporidium.			FASt_er-fa_CRYPA_AAC99407 (ER Fam: ER-FAS: alveolata) FASt_er-fb_CRYPA_AAC99407 (ER Fam: ER-FAS: alveolata) FASt_er-fc_CRYPA_AAC99407 (ER Fam: ER-FAS: alveolata)
Crypthecodinium cohnii. Alveolata; Dinophyceae; Gonyaulacales; Crypthecodiniaceae; Crypthecodinium.			DIP3t_CRYCO_AAF05738 (QOR Fam: DINAP1)

Supplementary materials

Additional references for data included in Tables 3 - 8. The references for each subfamily are numbered inside the supplementary Tables S2a - S2f, and listed at the end of the Table S2f.

Table S2a. Main subfamilies that conform PDH family (COG1063) of MDR.

Subfamily
DHSO (sorbitol dehydrogenase)
Ref: [155-157]
BDH (2,3-butanediol dehydrogenase)
Ref: [158-160]
TDH (threonine dehydrogenase)
Ref: [161-164]
BCHC (2-desacetyl-2-hydroxyethyl bacteriochlorophyllide *a* dehydrogenase)
Ref: [165;166]
SORE (L-sorbose-1-phosphate reductase)
Ref: [167-171]
Secondary ADH
Ref: [172-177]
GATD (galactitol 1-phosphate dehydrogenase)
Ref: [178-180]
SSP and related (sensing starvation protein)
Ref: [181;182]
FDEH (5-exo-hydroxycamphor dehydrogenase)
Ref: [183;184]
BDOR (bi-domain oxidoreductase)
Ref: [185]
Archaea GDH (glucose dehydrogenase)
Ref:[186-189]
FADH (formaldehyde dehydrogenase-independent of cofactor-/formaldehyde dismutase)
Ref: [190-196]

References are listed at the end of suplementary material.

Table S2b. Main subfamilies that conform ADH family (COG1062) of MDR.

Subfamily
Aryl/Alkyl ADH: Firmicutes
Ref: [197;198]
Benzyl ADH
Ref: *Acinetobacter calcoaceticus* [199-201]; *Acinetobacter* sp. [202;203]; *Pseudomonas putida* [204-206]; *Pseudomonas* sp.[207;208]; *Sphingomonas aromaticivorans* [209]; *Rhizobium tropici* [210]
HNL (Hydroxynitrile lyase: acetone cyanohydrin lyase)
Ref: [211-214]
FADH: mycothiol-dependent (formaldehyde dehydrogenase dependent of mycothiol)
Ref: [215-217]
Class III ADH (formaldehyde dehydrogenase dependent of glutathion)
Ref: [218-238]
Animal ADH
Ref: *Homo sapiens*: ADHA_HUMAN [239-242], ADHB_HUMAN [239;243-254], ADHG_HUMAN [239;255;256], ADHP_HUMAN [257;258], ADH6_HUMAN [259], ADH7_HUMAN [260-267]; *Papio hamadryas* [268-273]; *Macaca mulatta* [274;275], *Equus caballus*: ADHE_HORSE [276-280], ADHS_HORSE [277-280]; *Oryctolagus cuniculus* [281-284]; *Mus musculus*: ADH7_MOUSE [285;286], ADHA_MOUSE [285;287-290]; *Peromyscus maniculatus* [291]; *Rattus norvegicus*: ADHA_RAT [292;293]; ADH7_RAT [294]; ADH2t_RAT_CAA62241 [295]; *Geomys* [296;297]; *Gallus gallus* [298;299]; *Coturnix coturnix japonica* [300;301]; *Anas platyrhynchos* [302]; *Apterix australis* [303]; *Struthio camelus* [304;305]; *Alligator mississippiensis* [306]; *Uromastix hardwickii* [307;308]; *Naja naja* [309]; *Rana perezi* [310-312]; *Xenopus laevis* [313]; *Gadus callarias* [314;315]
Plant ADH
Ref: *Pinus banksiana* [316]; *Washingtonia robusta* [317]; *Calamus usitatus* and *Phoenix reclinata* [318]; grasses (Poaceae)[319], *Zea mays* [320-322], *Zea luxurians* and *Zea diploperennis* [323], *Sorghum bicolor* [324]; *Tripsacum dactyloides* [325], *Oryza sativa* [326;327]; *Pennisetum americanum* [328]; *Pennisetum glaucum* [329]; *Dioscorea tokoro* and *Dioscorea tenuipes* [330]; *Rhynchoryza subulata* and *Leersia perrieri* [331]; *Hordeum vulgare* [332]; *Vitis vinifera* [333;334]; *Paeonia* [335;336]; *Solanum tuberosum* [337;338]; *Pisum sativum* [339]; *Lactuca sativa* [340]; *Lycopersicon esculentum* [341-344]; *Petunia hybrida* [345;346]; *Phaseolus acutifolius* [347]; *Nicotiana tabacum* [348]; *Pyrus communis* [349]; *Arabis gemmifera* and *Arabidopsis thaliana* [350;351]; *Gossypium* [352-355]

References are listed at the end of suplementary material.

Table S2c. Main subfamilies that conform CADH family and Y-ADH family (COG1064) of MDR.

Subfamily
CADH and related (cinnamyl alcohol dehydrogenase)
Ref: [356-370]
ELI3 (elicitor-inducible defense-related proteins)
Ref: [356;359;371-378]
Yeast ADH and related
Ref: *Kluyveromyces lactis* [379-382]; *Emericella nidulans* [383-385]; *Candida albicans* [386]; *Schizosaccharomyces pombe* [387]; *Saccharomyces cerevisiae* [388-392]; *Bacillus stearothermophilus* [393-395]; *Zymomonas mobilis* [396;397]; *Sinorhizobium meliloti* [398]; *Brucella abortus* [399]
Fungi MTD (mannitol-1-phosphate dehydrogenase)
Ref: [400-403]
Fungi secondary ADH
Ref: [404-406]
Broad ADH (broad substrate specificity ADH)
Ref: [407-415]

References are listed at the end of suplementary material.

Table S2d. Main subfamilies that conform QOR family and NRBP family (COG0504) of MDR.

Subfamily
ζ-crystallin/QOR (quinone oxidoreductase)
Ref: [416-427]
PIG3 and related (animal P53 Induced Gen 3: putative quinone oxidoreductase)
Ref: [428-430]
TED2 and related (quinone oxidoreductase involved in Tracheary Element Differentiation in plants)
Ref: [431-435]
NRBP1 (nuclear receptor binding protein/transcription factor)
Ref:[436-438]
Bifunctional QOR and related
Ref: [439-441]
VAT1
Ref: [442-448]
pER in actinomycetes (probable enoyl reductase in actinomycetes)
Ref: [449;450]
PKS-IAP (polyketide synthase-independent asociated proteins)
Ref: *Aspergillus terreus* [451-453]; *Cochliobolus carbonum* [454-457]
QORL-1 (quinone oxidoreductase-like 1)
Ref: [458]
DINAP (dinoflagellate nuclear associated protein)
Ref: [459;460]
ARP (auxin regulated protein)
Ref: [461;462]
DI-QOR (dark induced-quinone oxidoreductase)
Ref: [463-466]
DI-QOR / ARP related
Unpurified protein; uncharacterized
AL (alginate lyase)
Ref: [467-470]
AST (membrane traffic protein)
Ref: [471]
BRP (bacteriocin-related proteins)
Ref: *Lactococcus lactis* [472]; *Pediococcus pentosaceus*[473]
CCAR (crotonyl-CoA reductase) and related
Ref: *Streptomyces* (Firmicutes) [474-478]; *Pseudomonas* (Proteobacteria: α subdivision) [479;480]; *Methylobacterium* (Proteobacteria: a subdivision) [481;482]

References are listed at the end of suplementary material.

Table S2e. Main subfamilies that conform LTD family (COG2130) of MDR.

Subfamily
LTD (Leucotriene B_4 12-hydroxydehydrogenase)/PGR (15-Oxoprostaglandin 13-reductase)
Ref: [483-490]
AADH (allyl alcohol dehydrogenase)
Ref: [440;491;492]
LTD/AADH related
Uncharacterized proteins

References are listed at the end of suplementary material.

Table S2f. Main subfamilies that conform ER family (COG3321) of MDR.

Subfamily
Enoyl reductase (Fatty acid synthase -*FAS*-)
Ref: [493-505]
Enoyl reductase (modular polyketide synthase -*PKS*-)
Ref: [479;493;506-524]
Enoyl reductase (iterative polyketide synthase -*PKS*-)
Ref: [452;453;525;526]
ER -FAS: alveolata (enoyl reductase from type I fatty acid synthase in alveolata)
Ref: [527]

References are listed at the end of suplementary material.

Additional References

[155] Ng, K., Ye, R., Wu, X.C., & Wong, S.L. (1992) Sorbitol dehydrogenase from Bacillus subtilis. Purification, characterization, and gene cloning. *J. Biol. Chem.* **267,** 24989-24994.

[156] Habenicht, A., Motejadded, H., Kiess, M., Wegerer, A., & Mattes, R. (1999) Xylose utilisation: cloning and characterisation of the xylitol dehydrogenase from Galactocandida mastotermitis. *Biol. Chem.* **380,** 1405-1411.

[157] Johansson, K., El-Ahmad, M., Kaiser, C., Jornvall, H., Eklund, H., Hoog, J., & Ramaswamy, S. (2001) Crystal structure of sorbitol dehydrogenase. *Chem. Biol. Interact.* **130-132,** 351-358.

[158] Gonzalez, E., Fernandez, M.R., Larroy, C., Sola, L., Pericas, M.A., Pares, X., & Biosca, J.A. (2000) Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the Saccharomyces cerevisiae YAL060W gene product. Disruption and induction of the gene. *J. Biol. Chem.* **275,** 35876-35885.

[159] Gonzalez, E., Fernandez, M.R., Larroy, C., Pares, X., & Biosca, J.A. (2001) Characterization and functional role of Saccharomyces cerevisiae 2,3-butanediol dehydrogenase. *Chem. Biol. Interact.* **130-132,** 425-434.

[160] Huang, M., Oppermann, F.B., & Steinbuchel, A. (1994) Molecular characterization of the Pseudomonas putida 2,3-butanediol catabolic pathway. *FEMS Microbiol. Lett.* **124,** 141-150.

[161] Aronson, B.D., Somerville, R.L., Epperly, B.R., & Dekker, E.E. (1989) The primary structure of Escherichia coli L-threonine dehydrogenase. *J. Biol. Chem.* **264,** 5226-5232.

[162] Epperly, B.R. & Dekker, E.E. (1991) L-threonine dehydrogenase from Escherichia coli. Identification of an active site cysteine residue and metal ion studies. *J. Biol. Chem.* **266,** 6086-6092.

[163] Johnson, A.R., Chen, Y.W., & Dekker, E.E. (1998) Investigation of a catalytic zinc binding site in Escherichia coli L-threonine dehydrogenase by site-directed mutagenesis of cysteine-38. *Arch. Biochem. Biophys.* **358,** 211-221.

[164] Weng, S.F., Liu, Y.S., Lin, J.W., & Tseng, Y.H. (1997) Transcriptional analysis of the threonine dehydrogenase gene of Xanthomonas campestris. *Biochem. Biophys. Res. Commun.* **240,** 523-529.

[165] McGlynn, P. & Hunter, C.N. (1993) Genetic analysis of the bchC and bchA genes of Rhodobacter sphaeroides. *Mol. Gen. Genet.* **236,** 227-234.

[166] Wellington, C.L. & Beatty, J.T. (1989) Promoter mapping and nucleotide sequence of the bchC bacteriochlorophyll biosynthesis gene from Rhodobacter capsulatus. *Gene* **83,** 251-261.

[167] Anderson, R.L. & Simkins, R.A. (1982) L-Sorbose-1-phosphate reductase . *Methods Enzymol.* **89** (Pt D)**,** 248-251.

[168] Kelker, N.E., Simkins, R.A., & Anderson, R.L. (1972) Pathway of L-sorbose metabolism in Aerobacter aerogenes. *J. Biol. Chem.* **247,** 1479-1483.

[169] Reizer, J., Reizer, A., & Saier, M.H.J. (1995) Sorbose-1-P reductase (SorE) and the glucitol-6-P dehydrogenase (SorD) of the Klebsiella pneumoniae L-sorbose operon belong to the zinc-dependent dehydrogenase family and the short chain alcohol dehydrogenase family, respectively. *Res. Microbiol.* **146,** 183-184.

[170] Wehmeier, U.F., Nobelmann, B., & Lengeler, J.W. (1992) Cloning of the Escherichia coli sor genes for L-sorbose transport and metabolism and physical mapping of the genes near metH and iclR. *J. Bacteriol.* **174,** 7784-7790.

[171] Wehmeier, U.F. & Lengeler, J.W. (1994) Sequence of the sor-operon for L-sorbose

utilization from Klebsiella pneumoniae KAY2026. *Biochim. Biophys. Acta* **1208,** 348-351.
[172] Burdette, D. & Zeikus, J.G. (1994) Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from Thermoanaerobacter ethanolicus 39E and characterization of the secondary-alcohol dehydrogenase (2 degrees Adh) as a bifunctional alcohol dehydrogenase--acetyl-CoA reductive thioesterase. *Biochem. J.* **302,** 163-170.
[173] Korkhin, Y., Kalb, G., Peretz, M., Bogin, O., Burstein, Y., & Frolow, F. (1998) NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of Clostridium beijerinckii and Thermoanaerobacter brockii. *J. Mol. Biol.* **278,** 967-981.
[174] Li, C., Heatwole, J., Soelaiman, S., & Shoham, M. (1999) Crystal structure of a thermophilic alcohol dehydrogenase substrate complex suggests determinants of substrate specificity and thermostability. *Proteins* **37 ,** 619-627.
[175] Peretz, M. & Burstein, Y. (1989) Amino acid sequence of alcohol dehydrogenase from the thermophilic bacterium Thermoanaerobium brockii. *Biochemistry* **28,** 6549-6555.
[176] Lo, H.S. & Chang, C.J. (1982) Purification and properties of NADP-linked, alcohol dehydrogenase from Entamoeba histolytica. *J. Parasitol.* **68,** 372-377.
[177] Kumar, A., Shen, P.S., Descoteaux, S., Pohl, J. , Bailey, G., & Samuelson, J. (1992) Cloning and expression of an NADP(+)-dependent alcohol dehydrogenase gene of Entamoeba histolytica. *Proc. Natl Acad. Sci. USA* **89,** 10188-10192.
[178] Markwell, J.P. & Anderson, R.L. (1981) Pathway of galactitol catabolism in Klebsiella pneumoniae: oxidation of L-galactitol-1-phosphate by a NAD-specific dehydrogenase. *Arch. Biochem. Biophys.* **209,** 592-597.
[179] Nobelmann, B. & Lengeler, J.W. (1995) Sequence of the gat operon for galactitol utilization from a wild-type strain EC3132 of Escherichia coli. *Biochim. Biophys. Acta* **1262,** 69-72.
[180] Nobelmann, B. & Lengeler, J.W. (1996) Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. *J. Bacteriol.* **178,** 6790-6795.
[181] Huisman, G.W. & Kolter, R. (1994) Sensing starvation: a homoserine lactone--dependent signaling pathway in Escherichia coli. *Science* **265,** 537-539.
[182] Babbit, P.C., Mrachko, G.T., Hasson, M.S., Huisman, G.W., Kolter, R., Ringe, D., Petsko, G.A., Kenyon, G.L., & Gerlt, J.A. (1995) Functionally diverse enzyme superfamily that abstracts the alpha-protons of carboxylic-acids. *Science* **267,** 1159-1161.
[183] Aramaki, H., Koga, H., Sagara, Y., Hosoi, M., & Horiuchi, T. (1993) Complete nucleotide sequence of the 5-exo-hydroxycamphor dehydrogenase gene on the CAM plasmid of Pseudomonas putida (ATCC 17453). *Biochim. Biophys. Acta* **1174,** 91-94.
[184] Koga, H., Aramaki, H., Yamaguchi, E., Takeuchi, K., Horiuchi, T., & Gunsalus, I.C. (1986) camR, a negative regulator locus of the cytochrome P-450cam hydroxylase operon. *J. Bacteriol.* **166,** 1089-1095.
[185] Nakar, D. & Gutnick, D.L. (2001) Analysis of the wee gene cluster responsible for the biosynthesis of the polymeric bioemulsifier from the oil- degrading strain Acinetobacter lwoffii RAG-1. *Microbiology* **147,** 1937-1946.
[186] Bonete, M.J., Pire, C., LLorca, F.I., & Camacho, M.L. (1996) Glucose dehydrogenase from the halophilic Archaeon Haloferax mediterranei: enzyme purification, characterisation and N-terminal sequence. *FEBS Lett.* **383,** 227-229.
[187] Bright, J.R., Byrom, D., Danson, M.J., Hough, D.W., & Towner, P. (1993) Cloning, sequencing and expression of the gene encoding glucose dehydrogenase from the thermophilic archaeon Thermoplasma acidophilum. *Eur. J. Biochem.* **211,** 549-554.
[188] John, J., Crennell, S.J., Hough, D.W., Danson, M.J., & Taylor, G.L. (1994) The crystal

structure of glucose dehydrogenase from Thermoplasma acidophilum. *Structure* **2,** 385-393.
[189] Smith, L.D., Budgen, N., Bungard, S.J., Danson, M.J., & Hough, D.W. (1989) Purification and characterization of glucose dehydrogenase from the thermoacidophilic archaebacterium Thermoplasma acidophilum. *Biochem. J.* **261,** 973-977.
[190] Ito, K., Takahashi, M., Yoshimoto, T., & Tsuru, D. (1994) Cloning and high-level expression of the glutathione-independent formaldehyde dehydrogenase gene from Pseudomonas putida. *J. Bacteriol.* **176,** 2483-2491.
[191] Oppenheimer, N.J., Henehan, G.T., Huete-Perez, J.A., & Ito, K. (1997) P. putida formaldehyde dehydrogenase. An alcohol dehydrogenase masquerading as an aldehyde dehydrogenase. *Adv. Exp. Med. Biol.* **414,** 417-423.
[192] Tsuru, D., Oda, N., Matsuo, Y., Ishikawa, S., Ito, K., & Yoshimoto, T. (1997) Glutathione-independent formaldehyde dehydrogenase from Pseudomons putida: survey of functional groups with special regard for cysteine residues. *Biosci. Biotechnol. Biochem.* **61,** 1354-1357.
[193] Kato, N., Yamagami, T., Shimao, M., & Sakazawa, C. (1986) Formaldehyde dismutase, a novel NAD-binding oxidoreductase from Pseudomonas putida F61. *Eur. J. Biochem.* **156,** 59-64.
[194] Mason, R.P. & Sanders, J.K. (1989) In vivo enzymology: a deuterium NMR study of formaldehyde dismutase in Pseudomonas putida F61a and Staphylococcus aureus. *Biochemistry* **28,** 2160-2168.
[195] Piersma, S.R., de Vries, S., & Duine, J.A. (1997) Nicotinoprotein alcohol/aldehyde oxidoreductases. Enzymes with bound NAD(P) as cofactor. *Adv. Exp. Med. Biol.* **414,** 425-434.
[196] Yanase, H., Noda, H., Aoki, K., Kita, K., & Kato, N. (1995) Cloning, sequence analysis, and expression of the gene encoding formaldehyde dismutase from Pseudomonas putida F61. *Biosci. Biotechnol. Biochem.* **59,** 197-202.
[197] Poelarends, G.J., Kulakov, L.A., Larkin, M.J., van, H., V, & Janssen, D.B. (2000) Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. *J. Bacteriol.* **182,** 2191-2199.
[198] Poelarends, G.J., Zandstra, M., Bosma, T., Kulakov, L.A., Larkin, M.J., Marchesi, J.R., Weightman, A.J., & Janssen, D.B. (2000) Haloalkane-utilizing Rhodococcus strains isolated from geographically distinct locations possess a highly conserved gene cluster encoding haloalkane catabolism. *J. Bacteriol.* **182,** 2725-2731.
[199] MacKintosh, R.W. & Fewson, C.A. (1988) Benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II from Acinetobacter calcoaceticus. Substrate specificities and inhibition studies. *Biochem. J.* **255,** 653-661.
[200] MacKintosh, R.W. & Fewson, C.A. (1988) Benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II from Acinetobacter calcoaceticus. Purification and preliminary characterization. *Biochem. J.* **250,** 743-751.
[201] Gillooly, D.J., Robertson, A.G., & Fewson, C.A. (1998) Molecular characterization of benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II of Acinetobacter calcoaceticus. *Biochem. J.* **330,** 1375-1381.
[202] Jones, R.M., Collier, L.S., Neidle, E.L., & Williams, P.A. (1999) areABC genes determine the catabolism of aryl esters in Acinetobacter sp. Strain ADP1. *J. Bacteriol.* **181,** 4568-4575.
[203] Jones, R.M. & Williams, P.A. (2001) areCBA is an operon in Acinetobacter sp. strain ADP1 and Is controlled by AreR, a sigma(54)-dependent regulator. *J. Bacteriol.* **183,** 405-409.
[204] Shaw, J.P. & Harayama, S. (1990) Purification and characterisation of TOL plasmid-

encoded benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase of Pseudomonas putida. *Eur. J. Biochem.* **191,** 705-714.

[205] Shaw, J.P. & Harayama, S. (1992) Purification and characterisation of the NADH:acceptor reductase component of xylene monooxygenase encoded by the TOL plasmid pWW0 of Pseudomonas putida mt-2. *Eur. J. Biochem.* **209,** 51-61.

[206] Shaw, J.P., Rekik, M., Schwager, F., & Harayama, S. (1993) Kinetic studies on benzyl alcohol dehydrogenase encoded by TOL plasmid pWWO. A member of the zinc-containing long chain alcohol dehydrogenase family. *J. Biol. Chem.* **268,** 10842-10850.

[207] Peterson, J.A., Lu, J.Y., Geisselsoder, J., Graham-Lorence, S., Carmona, C., Witney, F., & Lorence, M.C. (1992) Cytochrome P-450terp. Isolation and purification of the protein and cloning and sequencing of its operon. *J. Biol. Chem.* **267,** 14193-14203.

[208] Iurescia, S., Marconi, A.M., Tofani, D., Gambacorta, A., Paterno, A., Devirgiliis, C., van der Werf, M.J., & Zennaro, E. (1999) Identification and sequencing of beta-myrcene catabolism genes from Pseudomonas sp. strain M1. *Appl. Environ. Microbiol.* **65,** 2871-2876.

[209] Romine, M.F., Stillwell, L.C., Wong, K.K., Thurston, S.J., Sisk, E.C., Sensen, C., Gaasterland, T., Fredrickson, J.K., & Saffer, J.D. (1999) Complete sequence of a 184-kilobase catabolic plasmid from Sphingomonas aromaticivorans F199. *J. Bacteriol.* **181,** 1585-1602.

[210] Rosenblueth, M., Hynes, M.F., & Martinez-Romero, E. (1998) Rhizobium tropici teu genes involved in specific uptake of Phaseolus vulgaris bean-exudate compounds. *Mol. Gen. Genet.* **258,** 587-598.

[211] Xu, L.L., Singh, B.K., & Conn, E.E. (1988) Purification and characterization of acetone cyanohydrin lyase from Linum usitatissimum. *Arch. Biochem. Biophys.* **263,** 256-263.

[212] Trummler, K. & Wajant, H. (1997) Molecular cloning of acetone cyanohydrin lyase from flax (Linum usitatissimum). Definition of a novel class of hydroxynitrile lyases. *J. Biol. Chem.* **272,** 4770-4774.

[213] Trummler, K., Roos, J., Schwaneberg, U., Effenberger, F., Forster, S., Pfizenmaier, K., & Wajant, H. (1998) Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (Linum usitatissimum) in Pichia pastoris - utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. *Plant Sci.* **139,** 19-27.

[214] Breithaupt, H., Pohl, M., Bonigk, W., Heim, P., Schimz, K.L., & Kula, M.R. (1999) Cloning and expression of (R)-hydroxynitrile lyase from Linum usitatissimum (flax). *J. Mol. Catal. B-Enzym.* **6,** 315-332.

[215] Norin, A., Van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A., & Jornvall, H. (1997) Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaroytic alcohol dehydrogenases--primary structure, conformational modelling and functional correlations. *Eur. J. Biochem.* **248,** 282-289.

[216] Van Ophem, P.W., Van Beeumen, J., & Duine, J.A. (1992) NAD-linked, factor-dependent formaldehyde dehydrogenase or trimeric, zinc- containing, long-chain alcohol dehydrogenase from Amycolatopsis methanolica. *Eur. J. Biochem.* **206,** 511-518.

[217] Eggeling, L. & Sahm, H. (1985) The formaldehyde dehydrogenase of Rhodococcus erythropolis, a trimeric enzyme requiring a cofactor and active with alcohols. *Eur. J. Biochem.* **150,** 129-134.

[218] Barber, R.D., Rott, M.A., & Donohue, T.J. (1996) Characterization of a glutathione-dependent formaldehyde dehydrogenase from Rhodobacter sphaeroides. *J. Bacteriol.* **178,** 1386-1393.

[219] Barber, R.D. & Donohue, T.J. (1998) Function of a glutathione-dependent formaldehyde dehydrogenase in Rhodobacter sphaeroides formaldehyde oxidation and assimilation.

Biochemistry **37,** 530-537.
[220] Ras, J., Van Ophem, P.W., Reijnders, W.N., van Spanning, R.J., Duine, J.A., Stouthamer, A.H., & Harms, N. (1995) Isolation, sequencing, and mutagenesis of the gene encoding NAD- and glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) from Paracoccus denitrificans, in which GD-FALDH is essential for methylotrophic growth. *J. Bacteriol.* **177,** 247-251.
[221] Iborra, F.J., Renau-Piqueras, J., Portoles, M., Boleda, M.D., Guerri, C., & Pares, X. (1992) Immunocytochemical and biochemical demonstration of formaldehyde dehydrogenase (class III alcohol dehydrogenase) in the nucleus. *J. Histochem. Cytochem.* **40 ,** 1865-1878.
[222] Persson, B., Zigler, J.S.J., & Jornvall, H. (1994) A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur. J. Biochem.* **226,** 15-22.
[223] Dolferus, R., Osterman, J.C., Peacock, W.J., & Dennis, E.S. (1997) Cloning of the Arabidopsis and rice formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes . *Genetics* **146,** 1131-1141.
[224] Koivusalo, M. & Uotila, L. (1991) Glutathione-dependent formaldehyde dehydrogenase (EC 1.2.1.1): evidence for the identity with class III alcohol dehydrogenase. *Adv. Exp. Med. Biol.* **284,** 305-313.
[225] Fernandez, M.R., Biosca, J.A., Martinez, M.C., Achkor, H., Farres, J., & Pares, X. (1997) Formaldehyde dehydrogenase from yeast and plant. Implications for the general functional and structural significance of class III alcohol dehydrogenase. *Adv. Exp. Med. Biol.* **414,** 373-381.
[226] Yang, Z.N., Bosron, W.F., & Hurley, T.D. (1997) Structure of human chi chi alcohol dehydrogenase: a glutathione-dependent formaldehyde dehydrogenase. *J. Mol. Biol.* **265,** 330-343.
[227] Gutheil, W.G., Holmquist, B., & Vallee, B.L. (1992) Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from Escherichia coli: a class III alcohol dehydrogenase. *Biochemistry* **31,** 475-481.
[228] Shafqat, J., El-Ahmad, M., Danielsson, O., Martinez, M.C., Persson, B., Pares, X., & Jornvall, H. (1996) Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). *Proc. Natl Acad. Sci. USA* **93,** 5595-5599.
[229] Danielsson, O. & Jornvall, H. (1992) Enzymogenesis - classical liver alcohol-dehydrogenase origin from the glutathione-dependent formaldehyde dehydrogenase line. *Proc. Natl Acad. Sci. USA* **89,** 9247-9251.
[230] Gutheil, W.G., Kasimoglu, E., & Nicholson, P.C. (1997) Induction of glutathione-dependent formaldehyde dehydrogenase activity in Escherichia coli and Hemophilus influenza. *Biochem. Biophys. Res. Commun.* **238,** 693-696.
[231] Canestro, C., Hjelmqvist, L., Albalat, R., Garcia-Fernandez, J., Gonzalez-Duarte, R., & Jornvall, H. (2000) Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. *Eur. J. Biochem.* **267,** 6511-6518.
[232] Luque, T., Atrian, S., Danielsson, O., Jornvall, H., & Gonzalez-Duarte, R. (1994) Structure of the Drosophila melanogaster glutathione-dependent formaldehyde dehydrogenase/octanol dehydrogenase gene (class III alcohol dehydrogenase). Evolutionary pathway of the alcohol dehydrogenase genes. *Eur. J. Biochem.* **225,** 985-993.
[233] Keung, W.M., Kunze, L., & Holmquist, B. (1995) Rabbit liver class III alcohol

dehydrogenase: a cathodic isoform with formaldehyde dehydrogenase activity. *Alcohol Clin. Exp. Res.* **19,** 860-866.
[234] Foglio, M.H. & Duester, G. (1996) Characterization of the functional gene encoding mouse class III alcohol dehydrogenase (glutathione-dependent formaldehyde dehydrogenase) and an unexpressed processed pseudogene with an intact open reading frame. *Eur. J. Biochem.* **237,** 496-504.
[235] Barber, R.D. & Donohue, T.J. (1998) Pathways for transcriptional activation of a glutathione-dependent formaldehyde dehydrogenase gene. *J. Mol. Biol.* **280,** 775-784.
[236] Mori, O., Haseba, T., Kameyama, K., Shimizu, H. , Kudoh, M., Ohaki, O., Arai, Y., Yamazaki, M., & Asano, G. (2000) Histological distribution of class III alcohol dehydrogenase in human brain. *Brain Res.* **852,** 186-190.
[237] Beisswenger, T.B., Holmquist, B., & Vallee, B.L. (1985) chi-ADH is the sole alcohol dehydrogenase isozyme of mammalian brains: implications and inferences. *Proc. Natl Acad. Sci. USA* **82,** 8369-8373.
[238] Kaiser, R., Fernandez, M.R., Pares, X., & Jornvall, H. (1993) Origin of the human alcohol dehydrogenase system: implications from the structure and properties of the octopus protein. *Proc. Natl Acad. Sci. USA* **90,** 11222-11226.
[239] Ikuta, T., Szeto, S., & Yoshida, A. (1986) Three human alcohol dehydrogenase subunits: cDNA structure and molecular and evolutionary divergence. *Proc. Natl Acad. Sci. USA* **83,** 634-638.
[240] Matsuo, Y. & Yokoyama, S. (1989) Molecular structure of the human alcohol dehydrogenase 1 gene. *FEBS Lett.* **243,** 57-60.
[241] von Bahr-Lindstrom, H., Hoog, J.O., Heden, L.O., Kaiser, R., Fleetwood, L., Larsson, K., Lake, M., Holmquist, B., Holmgren, A., & Hempel, J. (1986) cDNA and protein structure for the alpha subunit of human liver alcohol dehydrogenase. *Biochemistry* **25,** 2465-2470.
[242] Yasunami, M., Kikuchi, I., Sarapata, D., & Yoshida, A. (1990) The human class I alcohol dehydrogenase gene cluster: three genes are tandemly organized in an 80-kb-long segment of the genome. *Genomics* **7,** 152-158.
[243] Yokoyama, S., Yokoyama, R., & Rotwein, P. (1987) Molecular characterization of cDNA clones encoding the human alcohol-dehydrogenase β_1 and the evolutionary relationship to the other class-I subunit-α and subunit-γ. *Jpn. J. Genet.* **62,** 241-256.
[244] Burnell, J.C., Carr, L.G., Dwulet, F.E., Edenberg, H.J., Li, T.K., & Bosron, W.F. (1987) The human beta 3 alcohol dehydrogenase subunit differs from beta 1 by a Cys for Arg-369 substitution which decreases NAD(H) binding. *Biochem. Biophys. Res. Commun.* **146,** 1127-1133.
[245] Carr, L.G., Xu, Y., Ho, W.H., & Edenberg, H.J. (1989) Nucleotide sequence of the ADH2(3) gene encoding the human alcohol dehydrogenase beta 3 subunit. *Alcohol Clin. Exp. Res.* **13,** 594-596.
[246] Davis, G.J., Bosron, W.F., Stone, C.L., Owusu-Dekyi, K., & Hurley, T.D. (1996) X-ray structure of human beta3beta3 alcohol dehydrogenase. The contribution of ionic interactions to coenzyme binding. *J. Biol. Chem.* **271,** 17057-17061.
[247] Duester, G., Smith, M., Bilanchone, V., & Hatfield, G.W. (1986) Molecular analysis of the human class I alcohol dehydrogenase gene family and nucleotide sequence of the gene encoding the beta subunit. *J. Biol. Chem.* **261,** 2027-2033.
[248] Heden, L.O., Hoog, J.O., Larsson, K., Lake, M., Lagerholm, E., Holmgren, A., Vallee, B.L., Jornvall, H., & von Bahr-Lindstrom, H. (1986) cDNA clones coding for the beta-subunit of human liver alcohol dehydrogenase have differently sized 3'-non-coding regions. *FEBS Lett.* **194,** 327-332.
[249] Hempel, J., Buhler, R., Kaiser, R., Holmquist, B., de Zalenski, C., von Wartburg, J.P., Vallee, B., & Jornvall, H. (1984) Human liver alcohol dehydrogenase. 1. The primary

structure of the beta 1 beta 1 isoenzyme. *Eur. J. Biochem.* **145,** 437-445.
[250] Hurley, T.D., Bosron, W.F., Hamilton, J.A., & Amzel, L.M. (1991) Structure of human beta 1 beta 1 alcohol dehydrogenase: catalytic effects of non-active-site substitutions. *Proc. Natl Acad. Sci. USA* **88,** 8149-8153.
[251] Hurley, T.D., Bosron, W.F., Stone, C.L., & Amzel, L.M. (1994) Structures of three human beta alcohol dehydrogenase variants. Correlations with their functional differences. *J. Mol. Biol.* **239,** 415-429.
[252] Jornvall, H., Hempel, J., Vallee, B.L., Bosron, W.F., & Li, T.K. (1984) Human liver alcohol dehydrogenase: amino acid substitution in the beta 2 beta 2 Oriental isozyme explains functional properties, establishes an active site structure, and parallels mutational exchanges in the yeast enzyme. *Proc. Natl Acad. Sci. USA* **81,** 3024-3028.
[253] Matsuo, Y., Yokoyama, R., & Yokoyama, S. (1989) The genes for human alcohol dehydrogenases beta 1 and beta 2 differ by only one nucleotide. *Eur. J. Biochem.* **183 ,** 317-320.
[254] Xu, Y.L., Carr, L.G., Bosron, W.F., Li, T.K., & Edenberg, H.J. (1988) Genotyping of human alcohol dehydrogenases at the ADH2 and ADH3 loci following DNA sequence amplification. *Genomics* **2,** 209-214.
[255] Buhler, R., Hempel, J., Kaiser, R., de Zalenski, C., von Wartburg, J.P., & Jornvall, H. (1984) Human liver alcohol dehydrogenase. 2. The primary structure of the gamma 1 protein chain. *Eur. J. Biochem.* **145,** 447-453.
[256] Hoog, J.O., Heden, L.O., Larsson, K., Jornvall, H., & von Bahr-Lindstrom, H. (1986) The gamma 1 and gamma 2 subunits of human liver alcohol dehydrogenase. cDNA structures, two amino acid replacements, and compatibility with changes in the enzymatic properties. *Eur. J. Biochem.* **159,** 215-218.
[257] Hoog, J.O., von Bahr-Lindstrom, H., Heden, L.O., Holmquist, B., Larsson, K., Hempel, J., Vallee, B.L., & Jornvall, H. (1987) Structure of the class II enzyme of human liver alcohol dehydrogenase: combined cDNA and protein sequence determination of the pi subunit. *Biochemistry* **26,** 1926-1932.
[258] von Bahr-Lindstrom, H., Jornvall, H., & Hoog, J.O. (1991) Cloning and characterization of the human ADH4 gene. *Gene* **103,** 269-274.
[259] Yasunami, M., Chen, C.S., & Yoshida, A. (1991) A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. *Proc. Natl Acad. Sci. USA* **88,** 7610-7614.
[260] Cheung, B., Anderson, J.K., Holmes, R.S., & Beacham, I.R. (1995) Human stomach class IV alcohol dehydrogenase: molecular genetic analysis. *Alcohol. Clin. Exp. Res.* **19,** 185-186.
[261] Farres, J., Moreno, A., Crosas, B., Peralba, J.M., Allali-Hassani, A., Hjelmqvist, L., Jornvall, H., & Pares, X. (1994) Alcohol dehydrogenase of class IV (sigma sigma-ADH) from human stomach. cDNA sequence and structure/function relationships. *Eur. J. Biochem.* **224,** 549-557.
[262] Kedishvili, N.Y., Bosron, W.F., Stone, C.L., Hurley, T.D., Peggs, C.F., Thomasson, H.R., Popov, K.M., Carr, L.G., Edenberg, H.J., & Li, T.K. (1995) Expression and kinetic characterization of recombinant human stomach alcohol dehydrogenase. Active-site amino acid sequence explains substrate specificity compared with liver isozymes. *J. Biol. Chem.* **270,** 3625-3630.
[263] Parés, X., Cederlund, E., Moreno, A., Saubi, N., Hoog, J.O., & Jornvall, H. (1992) Class-IV alcohol-dehydrogenase (the gastric enzyme) - structural-analysis of human sigma-sigma-ADH reveals class-IV to be variable and confirms the presence of a 5TH mammalian alcohol-dehydrogenase class. *FEBS Lett.* **303,** 69-72.
[264] Satre, M.A., Zgombic-Knight, M., & Duester, G. (1994) The complete structure of human

class IV alcohol dehydrogenase (retinol dehydrogenase) determined from the ADH7 gene. *J. Biol. Chem.* **269,** 15606-15612.

[265] Xie, P., Parsons, S.H., Speckhard, D.C., Bosron, W.F., & Hurley, T.D. (1997) X-ray structure of human class IV sigmasigma alcohol dehydrogenase. Structural basis for substrate specificity. *J. Biol. Chem.* **272,** 18558-18563.

[266] Yokoyama, S., Matsuo, Y., Ramsbotham, R., & Yokoyama, R. (1994) Molecular characterization of a class IV human alcohol dehydrogenase gene (ADH7). *FEBS Lett.* **351,** 411-415.

[267] Zgombic-Knight, M., Foglio, M.H., & Duester, G. (1995) Genomic structure and expression of the ADH7 gene encoding human class IV alcohol dehydrogenase, the form most efficient for retinol metabolism in vitro . *J. Biol. Chem.* **270,** 4305-4311.

[268] Trezise, A.E., Godfrey, E.A., Holmes, R.S., & Beacham, I.R. (1989) Cloning and sequencing of cDNA encoding baboon liver alcohol dehydrogenase: evidence for a common ancestral lineage with the human alcohol dehydrogenase beta subunit and for class I ADH gene duplications predating primate radiation. *Proc. Natl Acad. Sci. USA* **86,** 5454-5458.

[269] Cheung, B., Holmes, R.S., Easteal, S., & Beacham, I.R. (1999) Evolution of class I alcohol dehydrogenase genes in catarrhine primates: gene conversion, substitution rates, and gene regulation. *Mol. Biol. Evol.* **16,** 23-36.

[270] Holmes, R.S. & VandeBerg, J.L. (1987) Baboon alcohol dehydrogenase isozymes: phenotypic changes in liver following chronic consumption of alcohol. *Isozymes Curr. Top. Biol. Med. Res.* **16,** 1-20.

[271] Holmes, R.S., Meyer, J., & VandeBerg, J.L. (1990) Baboon alcohol dehydrogenase isozymes: purification and properties of liver class I ADH. Moderate alcohol consumption reduces liver class I and class II ADH activities. *Prog. Clin. Biol. Res.* **344,** 819-841.

[272] Trezise, A.E., Cheung, B., Holmes, R.S., & Beacham, I.R. (1991) Evidence for three genes encoding class-I alcohol dehydrogenase subunits in baboon and analysis of the 5' region of the gene encoding the ADH beta subunit. *Gene* **103,** 211-218.

[273] Trezise, A.E., Cheung, B., Holmes, R.S., & Beacham, I.R. (1991) Molecular genetics and evolution of baboon liver alcohol dehydrogenase. *Alcohol Alcohol. Suppl* **1,** 85-90.

[274] Dafeldecker, W.P., Meadow, P.E., Pares, X., & Vallee, B.L. (1981) Simian liver alcohol dehydrogenase: isolation and characterization of isoenzymes from Macaca mulatta. *Biochemistry* **20,** 6729-6734.

[275] Light, D.R., Dennis, M.S., Forsythe, I.J., Liu, C.C., Green, D.W., Kratzer, D.A., & Plapp, B.V. (1992) Alpha-isoenzyme of alcohol-dehydrogenase from monkey liver - cloning, expression, mechanism, coenzyme, and substrate- specificity. *J. Biol. Chem.* **267,** 12592-12599.

[276] Alkaradaghi, S., Cedergrenzeppezauer, E.S., Hovmoller, S., Petratos, K., Terry, H., & Wilson, K.S. (1994) Refined crystal-structure of liver alcohol-dehydrogenase NADH complex at 1.8-Angstrom resolution. *Acta Crystall. D-Biol. Crystall.* **50,** 793-807.

[277] Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B.O., Tapia, O., Branden, C.I., & Akeson, A. (1976) Three-dimensional structure of horse liver alcohol dehydrogenase at 2-4 A resolution. *J. Mol. Biol.* **102,** 27-59.

[278] Eklund, H., Samama, J.P., & Jones, T.A. (1984) Crystallographic investigations of nicotinamide adenine dinucleotide binding to horse liver alcohol dehydrogenase. *Biochemistry* **23,** 5982-5996.

[279] Jornvall, H. (1970) Horse liver alcohol dehydrogenase. On the primary structures of the isoenzymes. *Eur. J. Biochem.* **16,** 41-49.

[280] Park, D.H. & Plapp, B.V. (1991) Isoenzymes of horse liver alcohol dehydrogenase active

on ethanol and steroids. cDNA cloning, expression, and comparison of active sites. *J. Biol. Chem.* **266,** 13296-13302.
[281] Hoshino, T., Ishiguro, I., & Ohta, Y. (1985) Rabbit liver alcohol dehydrogenase: purification and properties. *J. Biochem. (Tokyo)* **97,** 1163-1172.
[282] Keung, W.M. & Yip, P.K. (1989) Rabbit liver alcohol dehydrogenase: isolation and characterization of class I isozymes. *Biochem. Biophys. Res. Commun.* **158,** 445-453.
[283] Hoog, J.O., Vagelopoulos, N., Yip, P.K., Keung, W.M., & Jornvall, H. (1993) Isozyme developments in mammalian class-I alcohol dehydrogenase. cDNA cloning, functional correlations, and lack of evidence for genetic isozymes in rabbit. *Eur. J. Biochem.* **213,** 31-38.
[284] Svensson, S., Hedberg, J.J., & Hoog, J.O. (1998) Structural and functional divergence of class II alcohol dehydrogenase--cloning and characterisation of rabbit liver isoforms of the enzyme. *Eur. J. Biochem.* **251,** 236-243.
[285] Zgombic-Knight, M., Ang, H.L., Foglio, M.H., & Duester, G. (1995) Cloning of the mouse class IV alcohol dehydrogenase (retinol dehydrogenase) cDNA and tissue-specific expression patterns of the murine ADH gene family. *J. Biol. Chem.* **270,** 10868-10877.
[286] Zgombic-Knight, M., Deltour, L., Haselbeck, R.J., Foglio, M.H., & Duester, G. (1997) Gene structure and promoter for Adh3 encoding mouse class IV alcohol dehydrogenase (retinol dehydrogenase). *Genomics* **41,** 105-109.
[287] Ceci, J.D., Lawther, R., Duester, G., Hatfield, G.W., Smith, M., O'Malley, M.P., & Felder, M.R. (1986) Androgen induction of alcohol dehydrogenase in mouse kidney. Studies with a cDNA probe confirmed by nucleotide sequence analysis. *Gene* **41,** 217-224.
[288] Ceci, J.D., Zheng, Y.W., & Felder, M.R. (1987) Molecular analysis of mouse alcohol dehydrogenase: nucleotide sequence of the Adh-1 gene and genetic mapping of a related nucleotide sequence to chromosome 3. *Gene* **59,** 171-182.
[289] Edenberg, H.J., Zhang, K., Fong, K., Bosron, W.F., & Li, T.K. (1985) Cloning and sequencing of cDNA encoding the complete mouse liver alcohol dehydrogenase. *Proc. Natl Acad. Sci. USA* **82,** 2262-2266.
[290] Zhang, K., Bosron, W.F., & Edenberg, H.J. (1987) Structure of the mouse Adh-1 gene and identification of a deletion in a long alternating purine-pyrimidine sequence in the first intron of strains expressing low alcohol dehydrogenase activity. *Gene* **57,** 27-36.
[291] Zheng, Y.W., Bey, M., Liu, H., & Felder, M.R. (1993) Molecular basis of the alcohol dehydrogenase-negative deer mouse. Evidence for deletion of the gene for class I enzyme and identification of a possible new enzyme class. *J. Biol. Chem.* **268,** 24933-24939.
[292] Crabb, D.W. & Edenberg, H.J. (1986) Complete amino acid sequence of rat liver alcohol dehydrogenase deduced from the cDNA sequence. *Gene* **48,** 287-291.
[293] Crabb, D.W., Stein, P.M., Dipple, K.M., Hittle, J.B., Sidhu, R., Qulali, M., Zhang, K., & Edenberg, H.J. (1989) Structure and expression of the rat class I alcohol dehydrogenase gene. *Genomics* **5,** 906-914.
[294] Pares, X., Cederlund, E., Moreno, A., Hjelmqvist, L., Farres, J., & Jornvall, H. (1994) Mammalian class IV alcohol dehydrogenase (stomach alcohol dehydrogenase): structure, origin, and correlation with enzymology. *Proc. Natl Acad. Sci. USA* **91,** 1893-1897.
[295] Hoog, J.O. (1995) Cloning and characterization of a novel rat alcohol dehydrogenase of class II type. *FEBS Lett.* **368,** 445-448.
[296] Bradley, R.D., Bull, J.J., Johnson, A.D., & Hillis, D.M. (1993) Origin of a novel allele in a mammalian hybrid zone. *Proc. Natl Acad. Sci. USA* **90,** 8939-8941.
[297] Bradley, R.D., Adkins, R.M., Honeycutt, R.L., & McDonald, J.H. (1998) Nucleotide polymorphism at the alcohol dehydrogenase locus of pocket gophers, genus Geomys.

Mol. Biol. Evol. **15,** 709-717.

[298] Estonius, M., Karlsson, C., Fox, E.A., Hoog, J.O., Holmquist, B., Vallee, B.L., Davidson, W.S., & Jornvall, H. (1990) Avian alcohol dehydrogenase: the chicken liver enzyme. Primary structure, cDNA-cloning, and relationships to other alcohol dehydrogenases. *Eur. J. Biochem.* **194,** 593-602.

[299] Kedishvili, N.Y., Gough, W.H., Chernoff, E.A., Hurley, T.D., Stone, C.L., Bowman, K.D., Popov, K.M., Bosron, W.F., & Li, T.K. (1997) cDNA sequence and catalytic properties of a chick embryo alcohol dehydrogenase that oxidizes retinol and 3beta,5alpha-hydroxysteroids. *J. Biol. Chem.* **272,** 7494-7500.

[300] Kaiser, R., Nussrallah, B., Dam, R., Wagner, F.W., & Jornvall, H. (1990) Avian alcohol dehydrogenase. Characterization of the quail enzyme, functional interpretations, and relationships to the different classes of mammalian alcohol dehydrogenase. *Biochemistry* **29,** 8365-8371.

[301] Nussrallah, B.A., Dam, R., & Wagner, F.W. (1989) Characterization of Coturnix quail liver alcohol dehydrogenase enzymes. *Biochemistry* **28,** 6245-6251.

[302] Hiremath, L.S., Kessler, P.M., Sasaki, G.C., & Kolattukudy, P.E. (1992) Estrogen induction of alcohol-dehydrogenase in the uropygial gland of mallard ducks. *Eur. J. Biochem.* **203,** 449-457.

[303] Hjelmqvist, L., Metsis, M., Persson, H., Hoog, J.O., McLennan, J., & Jornvall, H. (1995) Alcohol dehydrogenase of class I: kiwi liver enzyme, parallel evolution in separate vertebrate lines, and correlation with 12S rRNA patterns. *FEBS Lett.* **367,** 306-310.

[304] Estonius, M., Hjelmqvist, L., & Jornvall, H. (1994) Diversity of vertebrate class I alcohol dehydrogenase. Mammalian and non-mammalian enzyme functions correlated through the structure of a ratite enzyme. *Eur. J. Biochem.* **224,** 373-378.

[305] Hjelmqvist, L., Estonius, M., & Jornvall, H. (1995) The vertebrate alcohol dehydrogenase system: variable class II type form elucidates separate stages of enzymogenesis. *Proc. Natl Acad. Sci. USA* **92,** 10904-10908.

[306] Persson, B., Bergman, T., Keung, W.M., Waldenstrom, U., Holmquist, B., Vallee, B.L., & Jornvall, H. (1993) Basic features of class-I alcohol dehydrogenase: variable and constant segments coordinated by inter-class and intra-class variability. Conclusions from characterization of the alligator enzyme. *Eur. J. Biochem.* **216,** 49-56.

[307] Hjelmqvist, L., Ericsson, M., Shafqat, J., Carlquist, M., Siddiqi, A.R., Hoog, J.O. , & Jornvall, H. (1992) Reptilian alcohol-dehydrogenase - heterogeneity relevant to class multiplicity of the mammalian enzyme. *FEBS Lett.* **298,** 297-300.

[308] Hjelmqvist, L., Shafqat, J., Siddiqi, A.R., & Jornvall, H. (1996) Linking of isozyme and class variability patterns in the emergence of novel alcohol dehydrogenase functions. Characterization of isozymes in Uromastix hardwickii. *Eur. J. Biochem.* **236,** 563-570.

[309] Shafqat, J., Hjelmqvist, L., & Jornvall, H. (1996) Liver class-I alcohol dehydrogenase isozyme relationships and constant patterns in a variable basic structure. Distinctions from characterization of an ethanol dehydrogenase in cobra, Naja naja. *Eur. J. Biochem*. **236,** 571-578.

[310] Cederlund, E., Peralba, J.M., Pares, X., & Jornvall, H. (1991) Amphibian alcohol dehydrogenase, the major frog liver enzyme. Relationships to other forms and assessment of an early gene duplication separating vertebrate class I and class III alcohol dehydrogenases. *Biochemistry* **30,** 2811-2816.

[311] Peralba, J.M., Cederlund, E., Crosas, B., Moreno, A., Julia, P., Martinez, S.E., Persson, B., Farres, J., Pares, X., & Jornvall, H. (1999) Structural and enzymatic properties of a gastric NADP(H)- dependent and retinal-active alcohol dehydrogenase. *J. Biol. Chem*. **274,** 26021-26026.

[312] Peralba, J.M., Crosas, B., Martinez, S.E., Julia, P., Farres, J., & Pares, X. (1999)

Amphibian alcohol dehydrogenase - Purification and characterization of classes I and III from Rana perezi. *Adv. Exp. Med. Biol.* **463,** 343-350.
[313] Hoffmann, I., Ang, H.L., & Duester, G. (1998) Alcohol dehydrogenases in Xenopus development: Conserved expression of ADH1 and ADH4 in epithelial retinoid target tissues. *Develop. Dynam.* **213,** 261-270.
[314] Danielsson, O., Eklund, H., & Jornvall, H. (1992) The major piscine liver alcohol-dehydrogenase has class-mixed properties in relation to mammalian alcohol dehydrogenases of class-I and class-III. *Biochemistry* **31,** 3751-3759.
[315] Ramaswamy, S., el Ahmad, M., Danielsson, O., Jornvall, H., & Eklund, H. (1996) Crystal structure of cod liver class I alcohol dehydrogenase: substrate pocket and structurally variable segments. *Protein Sci.* **5,** 663-671.
[316] Perry, D.J. & Furnier, G.R. (1996) Pinus banksiana has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl Acad. Sci. USA* **93,** 13020-13023.
[317] Morton, B.R., Gaut, B.S., & Clegg, M.T. (1996) Evolution of alcohol dehydrogenase genes in the palm and grass families. *Proc. Natl Acad. Sci. USA* **93,** 11735-11739.
[318] Gaut, B.S., Morton, B.R., McCaig, B.C., & Clegg, M.T. (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl Acad. Sci. USA* **93,** 10274-10279.
[319] Gaut, B.S., Peek, A.S., Morton, B.R., & Clegg, M.T. (1999) Patterns of genetic diversification within the Adh gene family in the grasses (Poaceae). *Mol. Biol. Evol.* **16,** 1086-1097.
[320] Gerlach, W.L., Pryor, A.J., Dennis, E.S., Ferl, R.J., Sachs, M.M., & Peacock, W.J. (1982) cDNA cloning and induction of the alcohol dehydrogenase gene (*Adh1*) of maize. *Proc. Natl Acad. Sci. USA* **79,** 2981-2985.
[321] Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M.M., Ferl, R.J., & Peacock, W.J. (1984) Molecular analysis of the alcohol dehydrogenase (Adh1) gene of maize. *Nucleic Acids Res.* **12,** 3983-4000.
[322] Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J., & Peacock, W.J. (1985) Molecular analysis of the alcohol dehydrogenase 2 (Adh2) gene of maize. *Nucleic Acids Res.* **13,** 727-743.
[323] Gaut, B.S. & Clegg, M.T. (1993) Molecular evolution of the Adh1 locus in the genus Zea. *Proc. Natl Acad. Sci. USA* **90,** 5095-5099.
[324] SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., & Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20,** 43-45.
[325] Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., & Gaut, B.S. (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA* **95,** 4441-4446.
[326] Xie, Y. & Wu, R. (1989) Rice alcohol dehydrogenase genes: anaerobic induction, organ specific expression and characterization of cDNA clones. *Plant Mol. Biol.* **13,** 53-68.
[327] Xie, Y. & Wu, R. (1990) Molecular analysis of an alcohol dehydrogenase-encoding genomic clone (adh2) from rice. *Gene* **87,** 185-191.
[328] Ha, D.B., Buffard, D., Berger, F., Breda, C., & Esnault, R. (1990) Nucleotide sequence encoding a slow allele of Adh1 in pearl millet. *Plant Mol. Biol.* **14,** 453-455.
[329] Gaut, B.S. & Clegg, M.T. (1993) Nucleotide polymorphism in the Adh1 locus of pearl millet (Pennisetum glaucum) (Poaceae). *Genetics* **135,** 1091-1097.
[330] Terauchi, R., Terachi, T., & Miyashita, N.T. (1997) DNA polymorphism at the Pgi locus of a wild yam, Dioscorea tokoro. *Genetics* **147,** 1899-1914.
[331] Ge, S., Sang, T., Lu, B.R., & Hong, D.Y. (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl Acad. Sci. USA* **96,** 14400-

14405.
[332] Trick, M., Dennis, E.S., Edwards, K.J.R., & Peacock, W.J. (1988) Molecular analysis of the alcohol dehydrogenase gene family of barley. *Plant Mol. Biol.* **11,** 147-160.
[333] SarniManchado, P., Verries, C., & Tesniere, C. (1997) Molecular characterization and structural analysis of one alcohol dehydrogenase gene (GV-Adh1) expressed during ripening of grapevine (Vitis vinifera L) berry. *Plant Sci.* **125,** 177-187.
[334] Tesniere, C. & Verries, C. (2000) Molecular cloning and expression of cDNAs encoding alcohol dehydrogenases from Vitis vinifera L. during berry development. *Plant Sci.* **157,** 77-88.
[335] Sang, T., Donoghue, M.J., & Zhang, D. (1997) Evolution of alcohol dehydrogenase genes in peonies (Paeonia): phylogenetic relationships of putative nonhybrid species. *Mol. Biol. Evol.* **14,** 994-1007.
[336] Sang, T. & Zhang, D.M. (1999) Reconstructing hybrid speciation using sequences of low copy nuclear genes: Hybrid origins of five Paeonia species based on Adh gene phylogenies. *Syst. Bot.* **24,** 148-163.
[337] Matton, D.P., Constabel, P., & Brisson, N. (1990) Alcohol dehydrogenase gene expression in potato following elicitor and stress treatment. *Plant Mol. Biol.* **14,** 775-783.
[338] Van Eldik, G.J., Ruiter, R.K., Van Herpen, M.M.A., Schrauwen, J.A.M., & Wullems, G.J. (1998) An alcohol dehydrogenase-like gene is specifically expressed in potato pistils. *J. Exp. Bot.* **49,** 1453-1454.
[339] Llewellyn, D.J., Finnegan, E.J., Ellis, J.G., Dennis, E.S., & Peacock, W.J. (1987) Structure and expression of an alcohol dehydrogenase 1 gene from Pisum sativum (cv. "Greenfeast"). *J. Mol. Biol.* **195,** 115-123.
[340] Toyomasu, T., Yamauchi, T., Yamane, H., Murofushi, N., & Inoue, Y. (1995) cDNA cloning and characterization of gibberellin-responsive genes in photoblastic lettuce seeds. *Biosci. Biotechnol. Biochem.* **59,** 1846-1849.
[341] Ingersoll, J.C., Rothenberg, M., Liedl, B.E., Folkerts, K., Garvin, D., Hanson, M.R., Doyle, J.J., & Mutschler, M.A. (1994) A novel anther-expressed adh-homologous gene in Lycopersicon esculentum. *Plant Mol. Biol.* **26,** 1875-1891.
[342] Van Der Straeten, D., Rodrigues, P.R., Gielen, J., & Van Montagu, M. (1991) Tomato alcohol dehydrogenase. Expression during fruit ripening and under hypoxic conditions. *FEBS Lett.* **295,** 39-42.
[343] Longhurst, T., Lee, E., Hinde, R., Brady, C., & Speirs, J. (1994) Structure of the tomato Adh2 gene and Adh2 pseudogenes, and a study of Adh2 gene expression in fruit. *Plant Mol. Biol.* **26,** 1073-1084.
[344] Bicsak, T.A., Kann, L.R., Reiter, A., & Chase, T.J. (1982) Tomato alcohol dehydrogenase: purification and substrate specificity. *Arch. Biochem. Biophys.* **216,** 605-615.
[345] Gregerson, R., McLean, M., Beld, M., Gerats, A.G., & Strommer, J. (1991) Structure, expression, chromosomal location and product of the gene encoding ADH1 in Petunia. *Plant Mol. Biol.* **17,** 37-48.
[346] Gregerson, R.G., Cameron, L., McLean, M., Dennis, P., & Strommer, J. (1993) Structure, expression, chromosomal location and product of the gene encoding Adh2 in Petunia. *Genetics* **133,** 999-1007.
[347] Garvin, D.F., Weeden, N.F., & Doyle, J.J. (1994) The reduced stability of a plant alcohol dehydrogenase is due to the substitution of serine for a highly conserved phenylalanine residue. *Plant Mol. Biol.* **26,** 643-655.
[348] Bucher, M., Brander, K.A., Sbicego, S., Mandel, T., & Kuhlemeier, C. (1995) Aerobic fermentation in tobacco pollen. *Plant Mol. Biol.* **28,** 739-750.
[349] Chervin, C., Truett, J.K., & Speirs, J. (1999) Alcohol dehydrogenase expression and

alcohol production during pear ripening. *J. Amer. Soc. Hort. Sci.* **124,** 71-75.
[350] Miyashita, N.T., Innan, H., & Terauchi, R. (1996) Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants Arabis gemmifera and Arabidopsis thaliana [letter]. *Mol. Biol. Evol.* **13,** 433-436.
[351] Miyashita, N.T. (2001) DNA variation in the 5' upstream region of the Adh locus of the wild plants Arabidopsis thaliana and Arabis gemmifera. *Mol. Biol. Evol.* **18,** 164-171.
[352] Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., & Wendel, J.F. (1998) The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *Amer. J. Bot.* **85,** 1301-1315.
[353] Small, R.L., Ryburn, J.A., & Wendel, J.F. (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (Gossypium L.). *Mol. Biol. Evol.* **16,** 491-501.
[354] Small, R.L. & Wendel, J.F. (2000) Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (Gossypium). *Genetics* **155,** 1913-1926.
[355] Small, R.L. & Wendel, J.F. (2000) Phylogeny, duplication, and intraspecific variation of Adh sequences in New World diploid cottons (Gossypium l., malvaceae). *Mol. Phylogenet. Evol.* **16,** 73-84.
[356] Brill, E.M., Abrahams, S., Hayes, C.M., Jenkins, C.L., & Watson, J.M. (1999) Molecular characterisation and expression of a wound-inducible cDNA encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne (*Medicago sativa* L.). *Plant Mol. Biol.* **41,** 279-291.
[357] Hibino, T., Shibata, D., Chen, J.Q., & Higuchi, T. (1993) Cinnamyl alcohol dehydrogenase from *Aralia cordata*: Cloning of the cDNA and expression of the gene in lignified tissues. *Plant Cell Physiol.* **34,** 659-665.
[358] De Melis, L.E., Whiteman, P.H., & Stevenson, T.W. (1999) Isolation and characterization of a cDNA clone encoding cinnamyl alcohol dehydrogenase in *Eucalyptus globulus* Labill. *Plant Sci.* **143,** 173-182.
[359] Williamson, J.D., Stoop, J.M., Massel, M.O., Conkling, M.A., & Pharr, D.M. (1995) Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein ELI3. *Proc. Natl Acad. Sci. USA* **92,** 7148-7152.
[360] Campbell, M.M. & Ellis, B.E. (1992) Fungal elicitor-mediated responses in pine cell cultures. I. Induction of phenylpropanoid metabolism. *Planta* **186,** 409-417.
[361] Zinser, C., Ernst, D., & Sandermann, H.J. (1998) Induction of stilbene synthase and cinnamyl alcohol dehydrogenase mRNAs in Scots pine (Pinus sylvestris L.) seedlings. *Planta* **204,** 169-176.
[362] Walter, M. (1992) Regulation of lignification in defense. In *Plant Gene Research: Genes Involved in Plant Defense* (Boller, T. & Meins, F., eds), pp. 327-352. Springer-Verlag, Vienna.
[363] Sederoff, R.R., MacKay, J.J., Ralph, J., & Hatfield, R.D. (1999) Unexpected variation in lignin. *Curr. Opin. Plant. Biol.* **2,** 145-152.
[364] Russell, W.R., Provan, G.J., Burkitt, M.J., & Chesson, A. (2000) Extent of incorporation of hydroxycinnamaldehydes into lignin in cinnamyl alcohol dehydrogenase-downregulated plants. *J. Biotechnol.* **79,** 73-85.
[365] Kim, H., Ralph, J., Yahiaoui, N., Pean, M., & Boudet, A.M. (2000) Cross-coupling of hydroxycinnamyl aldehydes into lignins. *Org. Lett.* **2,** 2197-2200.
[366] MacKay, J.J., O'Malley, D.M., Presnell, T., Booker, F.L., Campbell, M.M., Whetten, R.W., & Sederoff, R.R. (1997) Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proc. Natl Acad. Sci. USA* **94,** 8255-8260.

[367] Vailhe, M.A.B., Besle, J.M., Maillot, M.P., Cornu, A., Halpin, C., & Knight, M. (1998) Effect of down regulation of cinnamyl alcohol dehydrogenase on cell wall composition and on degradability of tobacco stems. *J. Sci. Food Agric.* **76,** 505-514.

[368] Baucher, M., Bernard-Vailhe, M.A., Chabbert, B., Besle, J.M., Opsomer, C., Van Montagu, M., & Botterman, J. (1999) Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (Medicago sativa L.) and the effect on lignin composition and digestibility. *Plant Mol. Biol.* **39,** 437-447.

[369] McKie, J.H., Jaouhari, R., Douglas, K.T., Goffner, D., Feuillet, C., Grima-Pettenati, J., Boudet, A.M., Baltas, M., & Gorrichon, L. (1993) A molecular model for cinnamyl alcohol dehydrogenase, a plant aromatic alcohol dehydrogenase involved in lignification. *Biochim. Biophys. Acta* **1202,** 61-69.

[370] Lauvergeat, V., Kennedy, K., Feuillet, C., McKie, J.H., Gorrichon, L., Baltas, M., Boudet, A.M., Grima-Pettenati, J., & Douglas, K.T. (1995) Site-directed mutagenesis of a serine residue in cinnamyl alcohol dehydrogenase, a plant NADPH-dependent dehydrogenase, affects the specificity for the coenzyme. *Biochemistry* **34,** 12426-12434.

[371] Logemann, E., Reinold, S., Somssich, I.E., & Hahlbrock, K. (1997) A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. *Biol. Chem.* **378,** 909-913.

[372] Quirino, B.F., Normanly, J., & Amasino, R.M. (1999) Diverse range of gene activity during Arabidopsis thaliana leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Mol. Biol.* **40,** 267-278.

[373] Prata, R.T.N., Williamson, J.D., Conkling, M.A., & Pharr, D.M. (1997) Sugar repression of mannitol dehydrogenase activity in celery cells. *Plant Physiol.* **114,** 307-314.

[374] Stoop, J.M.H., Williamson, J.D., Conkling, M.A., MacKay, J.J., & Pharr, D.M. (1998) Characterization of NAD-dependent mannitol dehydrogenase from celery as affected by ions, chelators, reducing agents and metabolites. *Plant Sci.* **131,** 43-51.

[375] Pharr, D.M., Prata, R.T.N., Jennings, D.B., Williamson, J.D., Zamski, E., Yamamoto, Y.T., & Conkling, M.A. (1999) Regulation of mannitol dehydrogenase: Relationship to plant growth and stress tolerance. *HortScience* **34,** 1027-1032.

[376] Somssich, I.E., Wernert, P., Kiedrowski, S., & Hahlbrock, K. (1996) Arabidopsis thaliana defense-related protein ELI3 is an aromatic alcohol:NADP(+) oxidoreductase. *Proc. Natl Acad. Sci. USA* **93,** 14199-14203.

[377] Stoop, J.M. & Pharr, D.M. (1992) Partial purification and characterization of mannitol: mannose 1-oxidoreductase from celeriac (Apium graveolens var. rapaceum) roots. *Arch. Biochem. Biophys.* **298,** 612-619.

[378] Stoop, J.M., Williamson, J.D., Conkling, M.A., & Pharr, D.M. (1995) Purification of NAD-dependent mannitol dehydrogenase from celery suspension cultures. *Plant Physiol.* **108,** 1219-1225.

[379] Saliola, M., Gonnella, R., Mazzoni, C., & Falcone, C. (1991) Two genes encoding putative mitochondrial alcohol dehydrogenases are present in the yeast Kluyveromyces lactis. *Yeast* **7,** 391-400.

[380] Shain, D.H., Salvadore, C., & Denis, C.L. (1992) Evolution of the alcohol dehydrogenase (ADH) genes in yeast: characterization of a fourth ADH in Kluyveromyces lactis. *Mol. Gen. Genet.* **232,** 479-488.

[381] Zeeman, A.M., Kuyper, M., Pronk, J.T., Van Dijken, J.P., & Steensma, H.Y. (2000) Regulation of pyruvate metabolism in chemostat cultures of Kluyveromyces lactis CBS 2359. *Yeast* **16,** 611-620.

[382] Bozzi, A., Saliola, M., Falcone, C., Bossa, F., & Martini, F. (1997) Structural and biochemical studies of alcohol dehydrogenase isozymes from Kluyveromyces lactis . *Biochim. Biophys. Acta* **1339,** 133-142.

[383] Hunter, G.D., Jones, I.G., & Sealy-Lewis, H.M. (1996) The cloning and sequencing of the

alcB gene, coding for alcohol dehydrogenase II, in Aspergillus nidulans. *Curr. Genet.* **29,** 122-129.

[384] Gwynne, D.I., Buxton, F.P., Sibley, S., Davies, R.W., Lockington, R.A., Scazzocchio, C., & Sealy-Lewis, H.M. (1987) Comparison of the cis-acting control regions of two coordinately controlled genes involved in ethanol utilization in Aspergillus nidulans. *Gene* **51,** 205-216.

[385] McKnight, G.L., Kato, H., Upshall, A., Parker, M.D., Saari, G., & O'Hara, P.J. (1985) Identification and molecular analysis of a third Aspergillus nidulans alcohol dehydrogenase gene. *EMBO J.* **4,** 2093-2099.

[386] Bertram, G., Swoboda, R.K., Gooday, G.W., Gow, N.A., & Brown, A.J. (1996) Structure and regulation of the Candida albicans ADH1 gene encoding an immunogenic alcohol dehydrogenase. *Yeast* **12,** 115-127.

[387] Russell, P.R. & Hall, B.D. (1983) The primary structure of the alcohol dehydrogenase gene from the fission yeast Schizosaccharomyces pombe. *J. Biol. Chem.* **258,** 143-149.

[388] Denis, C.L., Ferguson, J., & Young, E.T. (1983) mRNA levels for the fermentative alcohol dehydrogenase of Saccharomyces cerevisiae decrease upon growth on a nonfermentable carbon source. *J. Biol. Chem.* **258,** 1165-1171.

[389] Ciriacy, M. (1975) Genetics of alcohol dehydrogenase in Saccharomyces cerevisiae. II. Two loci controlling synthesis of the glucose-repressible ADH II. *Mol. Gen. Genet.* **138,** 157-164.

[390] Dombek, K.M., Voronkova, V., Raney, A., & Young, E.T. (1999) Functional analysis of the yeast Glc7-binding protein Reg1 identifies a protein phosphatase type 1-binding motif as essential for repression of ADH2 expression. *Mol. Cell. Biol.* **19,** 6029-6040.

[391] Young, E.T. & Pilgrim, D. (1985) Isolation and DNA sequence of ADH3, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in Saccharomyces cerevisiae [published erratum appears in *Mol. Cell. Biol.* 1986; 6:2284]. *Mol. Cell. Biol.* **5,** 3024-3034.

[392] Bakker, B.M., Bro, C., Kotter, P., Luttik, M.A. , Van Dijken, J.P., & Pronk, J.T. (2000) The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in Saccharomyces cerevisiae. *J. Bacteriol.* **182,** 4730-4737.

[393] Sakoda, H. & Imanaka, T. (1992) Cloning and sequencing of the gene coding for alcohol dehydrogenase of Bacillus stearothermophilus and rational shift of the optimum pH. *J. Bacteriol.* **174,** 1397-1402.

[394] Cannio, R., Rossi, M., & Bartolucci, S. (1994) A few amino acid substitutions are responsible for the higher thermostability of a novel NAD(+)-dependent bacillar alcohol dehydrogenase. *Eur. J. Biochem.* **222,** 345-352.

[395] Robinson, G.A., Bailey, C.J., & Dowds, B.C. (1994) Gene structure and amino acid sequences of alcohol dehydrogenases of Bacillus stearothermophilus. *Biochim. Biophys. Acta* **1218,** 432-434.

[396] Neale, A.D., Scopes, R.K., Kelly, J.M., & Wettenhall, R.E. (1986) The two alcohol dehydrogenases of Zymomonas mobilis. Purification by differential dye ligand chromatography, molecular characterisation and physiological roles. *Eur. J. Biochem*. **154,** 119-124.

[397] Keshav, K.F., Yomano, L.P., An, H.J., & Ingram, L.O. (1990) Cloning of the Zymomonas mobilis structural gene encoding alcohol dehydrogenase I (adhA): sequence comparison and expression in Escherichia coli. *J. Bacteriol.* **172,** 2491-2497.

[398] Willis, L.B. & Walker, G.C. (1998) Identification of the Rhizobium meliloti alcohol dehydrogenase gene (adhA) and heterologous expression in Alcaligenes eutrophus. *Biochim. Biophys. Acta* **1384,** 197-203.

[399] Scupham, A.J. & Triplett, E.W. (1997) Isolation and characterization of the UDP-glucose

4'-epimerase-encoding gene, galE, from Brucella abortus 2308. *Gene* **202,** 53-59.
[400] Suvarna, K., Bartiss, A., & Wong, B. (2000) Mannitol-1-phosphate dehydrogenase from cryptococcus neoformans is a zinc-containing long-chain alcohol/polyol dehydrogenase. *Microbiology* **146,** 2705-2713.
[401] Chaturvedi, V., Flynn, T., Niehaus, W.G., & Wong, B. (1996) Stress tolerance and pathogenic potential of a mannitol mutant of Cryptococcus neoformans. *Microbiology* **142,** 937-943.
[402] Chaturvedi, V., Wong, B., & Newman, S.L. (1996) Oxidative killing of Cryptococcus neoformans by human neutrophils. Evidence that fungal mannitol protects by scavenging reactive oxygen intermediates. *J. Immunol.* **156,** 3836-3840.
[403] Niehaus, W.G. & Flynn, T. (1994) Regulation of mannitol biosynthesis and degradation by Cryptococcus neoformans. *J. Bacteriol.* **176,** 651-655.
[404] Yamamoto, H., Matsuyama, A., Kobayashi, Y., & Kawada, N. (1995) Purification and characterization of (S)-1,3-butanediol dehydrogenase from Candida parapsilosis. *Biosci. Biotechnol. Biochem.* **59,** 1769-1770.
[405] Yamamoto, H., Kawada, N., Matsuyama, A., & Kobayashi, Y. (1999) Cloning and expression in Escherichia coli of a gene coding for a secondary alcohol dehydrogenase from *Candida parapsilosis*. *Biosci. Biotechnol. Biochem.* **63,** 1051-1055.
[406] Matsuyama, A., Yamamoto, H., Kawada, N., & Kobayashi, Y. (2001) Industrial production of (R)-1,3-butanediol by new biocatalysts. *J. Mol. Catal. B-Enzymatic* **11,** 513-521.
[407] Rella, R., Raia, C.A., Pensa, M., Pisani, F.M., Gambacorta, A., De Rosa, M., & Rossi, M. (1987) A novel archaebacterial NAD+-dependent alcohol dehydrogenase. Purification and properties. *Eur. J. Biochem.* **167,** 475-479.
[408] Ammendola, S., Raia, C.A., Caruso, C., Camardella, L., D'Auria, S., De Rosa, M., & Rossi, M. (1992) Thermostable NAD(+)-dependent alcohol dehydrogenase from Sulfolobus solfataricus: gene and protein sequence determination and relationship to other alcohol dehydrogenases. *Biochemistry* **31,** 12514-12523.
[409] Cannio, R., Fiorentino, G., Carpinelli, P., Rossi, M., & Bartolucci, S. (1996) Cloning and overexpression in Escherichia coli of the genes encoding NAD-dependent alcohol dehydrogenase from two Sulfolobus species. *J. Bacteriol.* **178,** 301-305.
[410] Raia, C.A., Caruso, C., Marino, M., Vespa, N., & Rossi, M. (1996) Activation of Sulfolobus solfataricus alcohol dehydrogenase by modification of cysteine residue 38 with iodoacetic acid. *Biochemistry* **35,** 638-647.
[411] Cannio, R., Fiorentino, G., Rossi, M., & Bartolucci, S. (1999) The alcohol dehydrogenase gene: distribution among Sulfolobales and regulation in Sulfolobus solfataricus. *FEMS Microbiol. Lett.* **170,** 31-39.
[412] Itoh, N., Morihama, R., Wang, J., Okada, K., & Mizuguchi, N. (1997) Purification and characterization of phenylacetaldehyde reductase from a styrene-assimilating Corynebacterium strain, ST-10. *Appl. Environ. Microbiol.* **63,** 3783-3788.
[413] Wang, J.C., Sakakibara, M., Liu, J.Q., Dairi, T., & Itoh, N. (1999) Cloning, sequence analysis, and expression in Escherichia coli of the gene encoding phenylacetaldehyde reductase from styrene-assimilating Corynebacterium sp. strain ST-10 . *Appl. Microbiol. Biotechnol.* **52 ,** 386-392.
[414] Giordano, A., Cannio, R., La Cara, F., Bartolucci, S., Rossi, M., & Raia, C.A. (1999) Asn249Tyr substitution at the coenzyme binding domain activates Sulfolobus solfataricus alcohol dehydrogenase and increases its thermal stability. *Biochemistry* **38,** 3043-3054.
[415] Cheng, Q., Thomas, S.M., Kostichka, K., Valentine, J.R., & Nagarajan, V. (2000) Genetic analysis of a gene cluster for cyclohexanol oxidation in Acinetobacter sp. Strain SE19 by in vitro transposition. *J. Bacteriol.* **182,** 4744-4751.
[416] Rao, P.V. & Zigler, J.S.J. (1992) Purification and characterization of zeta-

crystallin/quinone reductase from guinea pig liver. *Biochim. Biophys. Acta* **1117,** 315-320.
[417] Gonzalez, P., Hernandez-Calzadilla, C., Rao, P.V., Rodriguez, I.R., Zigler, J.S.J., & Borras, T. (1994) Comparative analysis of the zeta-crystallin/quinone reductase gene in guinea pig and mouse. *Mol. Biol. Evol.* **11,** 305-315.
[418] Gonzalez, P., Rao, P.V., Nunez, S.B., & Zigler, J.S.J. (1995) Evidence for independent recruitment of zeta-crystallin/quinone reductase (CRYZ) as a crystallin in camelids and hystricomorph rodents. *Mol. Biol. Evol.* **12 ,** 773-781.
[419] Rao, P.V., Gonzalez, P., Persson, B., Jornvall, H., Garland, D., & Zigler, J.S.J. (1997) Guinea pig and bovine zeta-crystallins have distinct functional characteristics highlighting replacements in otherwise similar structures. *Biochemistry* **36,** 5353-5362.
[420] Fujii, Y., Kimoto, H., Ishikawa, K., Watanabe, K., Yokota, Y., Nakai, N., & Taketo, A. (2001) Taxon-specific zeta -crystallin in Japanese tree frog (Hyla japonica) lens. *J. Biol. Chem.* **276,** 28134-28139.
[421] Garland, D., Rao, P.V., Del Corso, A., Mura, U. , & Zigler, J.S.J. (1991) zeta-Crystallin is a major protein in the lens of Camelus dromedarius. *Arch. Biochem. Biophys.* **285,** 134-136.
[422] Rao, P.V., Krishna, C.M., & Zigler, J.S.J. (1992) Identification and characterization of the enzymatic activity of zeta-crystallin from guinea pig lens. A novel NADPH:quinone oxidoreductase. *J. Biol. Chem.* **267,** 96-102.
[423] Duhaiman, A.S., Rabbani, N., AlJafari, A.A., & Alhomida, A.S. (1995) Purification and characterization of zeta-crystallin from the camel lens. *Biochem. Biophys. Res. Commun.* **215,** 632-640.
[424] Kang, L.Y., Zheng, S.H., Liu, Y.X., Shen, D.F., & Chen, J.H. (1985) Isolation and characterization of a novel nucleic acid binding protein from calf lenses. *Curr. Eye Res.* **4,** 199-206.
[425] Gagna, C.E., Chen, J.H., Kuo, H.R., & Lambert, W.C. (1998) Binding properties of bovine ocular lens zeta-crystallin to right-handed B-DNA, left-handed Z-DNA, and single-stranded DNA. *Cell. Biol. Int.* **22,** 217-225.
[426] Chen, J.H., Tong, T.C., & Zhang, L. (1991) Rapid enhancement of "MIP" 26kDa protein phosphorylation by RF-36 nucleic acid binding protein in lens cells. *Lens Eye Toxic Res*. **8,** 469-487.
[427] Tang, A. & Curthoys, N.P. (2001) Identification of zeta-crystallin/NADPH:quinone reductase as a renal glutaminase mRNA pH response element-binding protein. *J. Biol. Chem.* **276,** 21375-21380.
[428] Flatt, P.M., Polyak, K., Tang, L.J., Scatena, C.D., Westfall, M.D., Rubinstein, L.A., Yu, J., Kinzler, K.W., Vogelstein, B., Hill, D.E., & Pietenpol, J.A. (2000) p53-dependent expression of PIG3 during proliferation, genotoxic stress, and reversible growth arrest. *Cancer Lett.* **156,** 63-72.
[429] Kostic, C. & Shaw, P.H. (2000) Isolation and characterization of sixteen novel p53 response genes. *Oncogene* **19,** 3978-3987.
[430] Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W., & Vogelstein, B. (1997) A model for p53-induced apoptosis. *Nature* **389,** 300-305.
[431] Paoli, G.C., Vichivanives, P., & Tabita, F.R. (1998) Physiological control and regulation of the Rhodobacter capsulatus cbb operons. *J. Bacteriol.* **180,** 4258-4269.
[432] Suzuki, K., Hattori, Y., Uraji, M., Ohta, N., Iwata, K., Murata, K., Kato, A., & Yoshida, K. (2000) Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. *Gene* **242,** 331-336.
[433] Thorn, J.M., Barton, J.D., Dixon, N.E., Ollis, D.L., & Edwards, K.J. (1995) Crystal structure of Escherichia coli QOR quinone oxidoreductase complexed with NADPH. *J.*

Mol. Biol. **249,** 785-799.
[434] Edwards, K.J., Barton, J.D., Rossjohn, J., Thorn, J.M., Taylor, G.L., & Ollis, D.L. (1996) Structural and sequence comparisons of quinone oxidoreductase, zeta-crystallin, and glucose and alcohol dehydrogenases. *Arch. Biochem. Biophys.* **328,** 173-183.
[435] Demura, T. & Fukuda, H. (1994) Novel vascular cell-specific genes whose expression is regulated temporally and spatially during vascular system development. *Plant Cell* **6,** 967-981.
[436] Masuda, N., Yasumo, H., Furusawa, T., Tsukamoto, T., Sadano, H., & Osumi, T. (1998) Nuclear receptor binding factor-1 (NRBF-1), a protein interacting with a wide spectrum of nuclear hormone receptors. *Gene* **221,** 225-233.
[437] Yamazoe, M., Shirahige, K., Rashid, M.B., Kaneko, Y., Nakayama, T., Ogasawara, N., & Yoshikawa, H. (1994) A protein which binds preferentially to single-stranded core sequence of autonomously replicating sequence is essential for respiratory function in mitochondria of Saccharomyces cerevisiae. *J. Biol. Chem.* **269,** 15244-15252.
[438] Owen, G.I. & Zelent, A. (2000) Origins and evolutionary diversification of the nuclear receptor superfamily. *Cell. Mol. Life Sci.* **57,** 809-827.
[439] Liu, X. & Chang, K.P. (1994) Identification by extrachromosomal amplification and overexpression of a zeta-crystallin/NADPH-oxidoreductase homologue constitutively expressed in Leishmania spp. *Mol. Biochem. Parasitol.* **66,** 201-210.
[440] Babiychuk, E., Kushnir, S., Belles-Boix, E., Van Montagu, M., & Inze, D. (1995) Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. *J. Biol. Chem.* **270,** 26224-26231.
[441] Ichinose, Y., Tiemann, K., Schwenger-Erger, C., Toyoda, K., Hein, F., Hanselle, T., Cornels, H., & Barz, W. (2000) Genes expressed in Ascochyta rabiei-inoculated chickpea plants and elicited cell cultures as detected by differential cDNA- hybridization. *Z. Naturforsch C* **55,** 44-54.
[442] Linial, M., Miller, K., & Scheller, R.H. (1989) VAT-1: an abundant membrane protein from Torpedo cholinergic synaptic vesicles. *Neuron* **2,** 1265-1273.
[443] Linial, M. (1993) VAT-1 from Torpedo electric organ forms a high-molecular-mass protein complex within the synaptic vesicle membrane. *Eur. J. Biochem.* **216,** 189-197.
[444] Linial, M. & Levius, O. (1993) The protein VAT-1 from Torpedo electric organ exhibits an ATPase activity. *Neurosci Lett.* **152,** 155-157.
[445] Linial, M. & Levius, O. (1993) VAT-1 from Torpedo is a membranous homologue of zeta crystallin. *FEBS Lett.* **315,** 91-94.
[446] Hayess, K., Kraft, R., Sachsinger, J., Janke, J., Beckmann, G., Rohde, K., Jandrig, B., & Benndorf, R. (1998) Mammalian protein homologous to VAT-1 of Torpedo californica: isolation from Ehrlich ascites tumor cells, biochemical characterization, and organization of its gene. *J. Cell. Biochem.* **69,** 304-315.
[447] Levius, O. & Linial, M. (1993) VAT-1 from Torpedo synaptic vesicles is a calcium binding protein: a study in bacterial expression systems. *Cell. Mol. Neurobiol.* **13,** 483-492.
[448] Linial, M., Levius, O., Ilouz, N., & Parnas, D. (1995) The effect of calcium levels on synaptic proteins. A study on VAT-1 from Torpedo. *J. Physiol. Paris* **89,** 103-112.
[449] Fernandez-Moreno, M.A., Martinez, E., Caballero, J.L., Ichinose, K., Hopwood, D.A., & Malpartida, F. (1994) DNA sequence and functions of the actVI region of the actinorhodin biosynthetic gene cluster of Streptomyces coelicolor A3(2). *J. Biol. Chem.* **269,** 24854-24863.
[450] Taguchi, T., Itou, K., Ebizuka, Y., Malpartida, F., Hopwood, D.A., Surti, C.M., Booker-Milburn, K.I., Stephenson, G.R., & Ichinose, K. (2000) Chemical characterisation of disruptants of the Streptomyces coelicolor A3(2) actVI genes involved in actinorhodin biosynthesis. *J. Antibiot. (Tokyo)* **53,** 144-152.

[451] Auclair, K., Kennedy, J., Hutchinson, C.R., & Vederas, J.C. (2001) Conversion of cyclic nonaketides to lovastatin and compactin by a lovc deficient mutant of Aspergillus terreus. *Bioorg. Med. Chem. Lett.* **11,** 1527-1531.

[452] Hutchinson, C.R., Kennedy, J., Park, C., Kendrew, S., Auclair, K., & Vederas, J. (2000) Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. *Antonie Van Leeuwenhoek* **78,** 287-295.

[453] Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C., & Hutchinson, C.R. (1999) Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284,** 1368-1372.

[454] Ahn, J.H. & Walton, J.D. (1998) Regulation of cyclic peptide biosynthesis and pathogenicity in *Cochliobolus carbonum* by TOXEp, a novel protein with a bZIP basic DNA-binding motif and four ankyrin repeats. *Mol. Gen. Genet.* **260,** 462-469.

[455] Cheng, Y.Q., Ahn, J.H., & Walton, J.D. (1999) A putative branched-chain-amino-acid transaminase gene required for HC-toxin biosynthesis and pathogenicity in Cochliobolus carbonum. *Microbiology-Uk* **145,** 3539-3546.

[456] Pitkin, J.W., Nikolskaya, A., Ahn, J.H., & Walton, J.D. (2000) Reduced virulence caused by meiotic instability of the TOX2 chromosome of the maize pathogen *Cochliobolus carbonum*. *Mol. Plant-Microbe Interact.* **13,** 80-87.

[457] Pedley, K.F. & Walton, J.D. (2001) Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor. *Proc. Natl Acad. Sci. USA* **98,** 14174-14179.

[458] Kim, M.Y., Lee, H.K., Park, J.S., Park, S.H., Kwon, H.B., & Soh, J. (1999) Identification of a zeta-crystallin (quinone reductase)-like 1 gene (CRYZL1) mapped to human chromosome 21q22.1. *Genomics* **57,** 156-159.

[459] Bhaud, Y., Geraud, M.L., Ausseil, J., Soyer-Gobillard, M.O., & Moreau, H. (1999) Cyclic expression of a nuclear protein in a dinoflagellate. *J. Eukaryot. Microbiol.* **46,** 259-267.

[460] Guillebault, D., Derelle, E., Bhaud, Y., & Moreau, H. (2001) Role of nuclear WW domains and proline-rich proteins in dinoflagellate transcription. *Protist* **152,** 127-138.

[461] Chen, J., Wu, D.R., Witham, F.H., Heuser, C.W. , & Arteca, R.N. (1996) Molecular cloning and characterization of auxin-regulated genes from mungbean hypocotyls during adventitious root formation. *J. Amer. Soc. Hort. Sci.* **121,** 393-398.

[462] Nam, Y.W., Tichit, L., Leperlier, M., Cuerq, B. , Marty, I., & Lelievre, J.M. (1999) Isolation and characterization of mRNAs differentially expressed during ripening of wild strawberry (Fragaria vesca L.) fruits. *Plant Mol. Biol.* **39,** 629-636.

[463] Matvienko, M., Wojtowicz, A., Wrobel, R., Jamison, D., Goldwasser, Y., & Yoder, J.I. (2001) Quinone oxidoreductase message levels are differentially regulated in parasitic and non-parasitic plants exposed to allelopathic quinones. *Plant J.* **25,** 375-387.

[464] Matvienko, M., Torres, M.J., & Yoder, J.I. (2001) Transcriptional Responses in the Hemiparasitic Plant Triphysaria versicolor to Host Plant Signals. *Plant Physiol.* **127,** 272-282.

[465] Yazaki, K., Matsuoka, H., Ujihara, T., & Sato, F. (1999) Shikonin biosynthesis in *Lithospermum erythrorhizon*: Light-induced negative regulation of secondary metabolism. *Plant Biotechnol.* **16,** 335-342.

[466] Yazaki, K. (2001) Root-specific production of secondary metabolites: Regulation of shikonin biosynthesis by light in *Lithospermum erythrorhizon*. *Natural Medicines* **55,** 49-54.

[467] Kong, I.S., Kim, Y.O., Kim, J.M., KIM, S.K., OH, D.H., Yu, J.H., & KONG, J.Y. (1995) Alginate lyase production of halophilic *Pseudomonas* sp by recombinant *Escherichia coli*. *J. Microbiol. Biotechnol.* **5,** 92-95.

[468] Lee, J.H., Kang, J.H., Kim, Y.O., Kim, J.M., & Kong, I.S. (1998) Nucleotide sequence

analysis and expression of the alginate lyase gene from Pseudomonas sp. W7 in Escherichia coli. *J. Microbiol. Biotechnol.* **8,** 531-535.

[469] Wong, T.Y., Preston, L.A., & Schiller, N.L. (2000) ALGINATE LYASE: review of major sources and enzyme characteristics, structure-function analysis, biological roles, and applications. *Annu. Rev. Microbiol.* **54,** 289-340.

[470] Kim, G.T., Kim, J.M., Yu, J.H., & Kong, I.S. (1996) Characterization of catabolite repression and the promoter of the alginate lyase gene (aly) from Pseudomonas sp. *Biotechnol. Lett.* **18,** 1271-1276.

[471] Chang, A. & Fink, G.R. (1995) Targeting of the yeast plasma membrane [H+]ATPase: a novel gene AST1 prevents mislocalization of mutant ATPase to the vacuole. *J. Cell. Biol.* **128,** 39-49.

[472] Dougherty, B.A., Hill, C., Weidman, J.F., Richardson, D.R., Venter, J.C., & Ross, R.P. (1998) Sequence and analysis of the 60 kb conjugative, bacteriocin- producing plasmid pMRC01 from Lactococcus lactis DPC3147. *Mol. Microbiol.* **29,** 1029-1038.

[473] Giacomini, A., Squartini, A., & Nuti, M.P. (2000) Nucleotide sequence and analysis of plasmid pMD136 from *Pediococcus pentosaceus* FBB61 (ATCC43200) involved in Pediocin A production. *Plasmid* **43,** 111-122.

[474] Wallace, K.K., Bao, Z.Y., Dai, H., Digate, R., Schuler, G., Speedie, M.K., & Reynolds, K.A. (1995) Purification of crotonyl-CoA reductase from Streptomyces collinus and cloning, sequencing and expression of the corresponding gene in Escherichia coli. *Eur. J. Biochem.* **233,** 954-962.

[475] Gandecha, A.R., Large, S.L., & Cundliffe, E. (1997) Analysis of four tylosin biosynthetic genes from the tylLM region of the Streptomyces fradiae genome. *Gene* **184,** 197-203.

[476] Liu, H. & Reynolds, K.A. (1999) Role of crotonyl coenzyme A reductase in determining the ratio of polyketides monensin A and monensin B produced by Streptomyces cinnamonensis. *J. Bacteriol.* **181,** 6806-6813.

[477] Wu, K., Chung, L., Revill, W.P., Katz, L., & Reeves, C.D. (2000) The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251,** 81-90.

[478] Liu, H. & Reynolds, K.A. (2001) Precursor supply for polyketide biosynthesis: the role of crotonyl-CoA reductase. *Metab. Eng.* **3,** 40-48.

[479] Rangaswamy, V., Jiralerspong, S., Parry, R., & Bender, C.L. (1998) Biosynthesis of the Pseudomonas polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. *Proc. Natl Acad. Sci. USA* **95,** 15469-15474.

[480] Rangaswamy, V., Mitchell, R., Ullrich, M., & Bender, C. (1998) Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *J. Bacteriol.* **180,** 3330-3338.

[481] Chistoserdova, L.V. & Lidstrom, M.E. (1996) Molecular characterization of a chromosomal region involved in the oxidation of acetyl-CoA to glyoxylate in the isocitrate-lyase-negative methylotroph Methylobacterium extorquens AM1. *Microbiology* **142,** 1459-1468.

[482] Korotkova, N. & Lidstrom, M.E. (2001) Connection between poly-beta-hydroxybutyrate biosynthesis and growth on C(1) and C(2) compounds in the methylotroph Methylobacterium extorquens AM1. *J. Bacteriol.* **183,** 1038-1046.

[483] Ensor, C.M., Zhang, H., & Tai, H.H. (1998) Purification, cDNA cloning and expression of 15-oxoprostaglandin 13-reductase from pig lung. *Biochem. J.* **330 (Pt 1),** 103-108.

[484] Primiano, T., Li, Y., Kensler, T.W., Trush, M.A., & Sutter, T.R. (1998) Identification of dithiolethione-inducible gene-1 as a leukotriene B4 12-hydroxydehydrogenase: implications for chemoprevention. *Carcinogenesis* **19,** 999-1005.

[485] Wiens, M., Krasko, A., Blumbach, B., Muller, I.M., & Muller, W.E. (2000) Increased

expression of the potential proapoptotic molecule DD2 and increased synthesis of leukotriene B4 during allograft rejection in a marine sponge. *Cell Death Differ.* **7,** 461-469.
[486] Yokomizo, T., Izumi, T., Takahashi, T., Kasama, T., Kobayashi, Y., Sato, F., Taketani, Y., & Shimizu, T. (1993) Enzymatic inactivation of leukotriene B4 by a novel enzyme found in the porcine kidney. Purification and properties of leukotriene B4 12-hydroxydehydrogenase. *J. Biol. Chem.* **268,** 18128-18135.
[487] Yokomizo, T., Ogawa, Y., Uozumi, N., Kume, K., Izumi, T., & Shimizu, T. (1996) cDNA cloning, expression, and mutagenesis study of leukotriene B4 12-hydroxydehydrogenase. *J. Biol. Chem.* **271,** 2844-2850.
[488] Yokomizo, T., Ogawa, Y., Uozumi, N., Kume, K., Izumi, T., & Shimizu, T. (1997) cDNA cloning and mutagenesis study of leukotriene B4 12-hydroxydehydrogenase. *Adv. Exp. Med. Biol.* **407,** 151-156.
[489] Clish, C.B., Levy, B.D., Chiang, N., Tai, H.H., & Serhan, C.N. (2000) Oxidoreductases in lipoxin A4 metabolic inactivation: a novel role for 15-oxoprostaglandin 13-reductase/leukotriene B4 12-hydroxydehydrogenase in inflammation. *J. Biol. Chem.* **275,** 25372-25380.
[490] Kitamura, S., Katsura, H., & Tatsumi, K. (1993) Purification of a 15-ketoprostaglandin delta 13-reductase from rat liver and its ability to reduce the double bond of xenobiotics. *Biochem. Mol. Biol. Int.* **30,** 839-847.
[491] Hirata, T., Tamura, Y., Yokobatake, N., Shimoda, K., & Ashida, Y. (2000) A 38 kDa allylic alcohol dehydrogenase from the cultured cells of Nicotiana tabacum. *Phytochemistry* **55,** 297-303.
[492] Mano, J., Babiychuk, E., Belles-Boix, E., Hiratake, J., Kimura, A., Inze, D., Kushnir, S., & Asada, K. (2000) A novel NADPH:diamide oxidoreductase activity in arabidopsis thaliana P1 zeta-crystallin. *Eur. J. Biochem.* **267,** 3661-3671.
[493] Smith, S. (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* **8,** 1248-1259.
[494] Jayakumar, A., Tai, M.H., Huang, W.Y., al-Feel, W., Hsu, M., Abu-Elheiga, L., Chirala, S.S., & Wakil, S.J. (1995) Human fatty acid synthase: properties and molecular cloning. *Proc. Natl Acad. Sci. USA* **92,** 8695-8699.
[495] Jayakumar, A., Huang, W.Y., Raetz, B., Chirala, S.S., & Wakil, S.J. (1996) Cloning and expression of the multifunctional human fatty acid synthase and its subdomains in Escherichia coli. *Proc. Natl Acad. Sci. USA* **93,** 14509-14514.
[496] Amy, C.M., Witkowski, A., Naggert, J., Williams, B., Randhawa, Z., & Smith, S. (1989) Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. *Proc. Natl Acad. Sci. USA* **86,** 3114-3118.
[497] Beck, K.F., Schreglmann, R., Stathopulos, I., Klein, H., Hoch, J., & Schweizer, M. (1992) The fatty acid synthase (FAS) gene and its promoter in Rattus norvegicus. *DNA Seq.* **2,** 359-386.
[498] Holzer, K.P., Liu, W., & Hammes, G.G. (1989) Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA. *Proc. Natl Acad. Sci. USA* **86,** 4387-4391.
[499] Huang, W.Y., Chirala, S.S., & Wakil, S.J. (1994) Amino-terminal blocking group and sequence of the animal fatty acid synthase. *Arch. Biochem. Biophys.* **314,** 45-49.
[500] Wakil, S.J., Stoops, J.K., & Joshi, V.C. (1983) Fatty acid synthesis and its regulation. *Annu. Rev. Biochem.* **52,** 537-579.
[501] Ueno, K. & Suzuki, Y. (1997) p260/270 expressed in embryonic abdominal leg cells of Bombyx mori can transfer palmitate to peptides. *J. Biol. Chem.* **272,** 13519-13526.
[502] Ueno, K. (2000) Involvement of fatty acid synthase in axonal development in mouse embryos. *Genes Cells* **5,** 859-869.
[503] Kuhajda, F.P., Jenner, K., Wood, F.D., Hennigar, R.A., Jacobs, L.B., Dick, J.D., &

Pasternack, G.R. (1994) Fatty acid synthesis: a potential selective target for antineoplastic therapy. *Proc. Natl Acad. Sci. USA* **91,** 6379-6383.

[504] Kuhajda, F.P. (2000) Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology. *Nutrition* **16,** 202-208.

[505] Kuhajda, F.P., Pizer, E.S., Li, J.N., Mani, N.S., Frehywot, G.L., & Townsend, C.A. (2000) Synthesis and antitumor activity of an inhibitor of fatty acid synthase. *Proc. Natl Acad. Sci. USA* **97,** 3450-3454.

[506] Mathur, M. & Kolattukudy, P.E. (1992) Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. bovis Bacillus Calmette-Guerin. *J. Biol. Chem.* **267,** 19388-19395.

[507] Rainwater, D.L. & Kolattukudy, P.E. (1985) Fatty acid biosynthesis in Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin. Purification and characterization of a novel fatty acid synthase, mycocerosic acid synthase, which elongates n-fatty acyl-CoA with methylmalonyl-CoA. *J. Biol. Chem.* **260,** 616-623.

[508] Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., & Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* **252,** 675-679.

[509] Ruan, X., Stassi, D., Lax, S.A., & Katz, L. (1997) A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 29253, a rapamycin-producing strain. *Gene* **203,** 1-9.

[510] Xue, Y., Zhao, L., Liu, H.W., & Sherman, D.H. (1998) A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. *Proc. Natl Acad. Sci. USA* **95,** 12111-12116.

[511] Xue, Y. & Sherman, D.H. (2000) Alternative modular polyketide synthase expression controls macrolactone structure. *Nature* **403,** 571-575.

[512] Staunton, J., Caffrey, P., Aparicio, J.F., Roberts, G.A., Bethell, S.S., & Leadlay, P.F. (1996) Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3,** 188-192.

[513] Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., & Julien, B. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* **287,** 640-642.

[514] Julien, B., Shah, S., Ziermann, R., Goldman, R. , Katz, L., & Khosla, C. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. *Gene* **249,** 153-160.

[515] Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J., & Ligon, J.M. (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chem. Biol.* **7,** 97-109.

[516] Kakavas, S.J., Katz, L., & Stassi, D. (1997) Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. *J. Bacteriol.* **179,** 7515-7522.

[517] Motamedi, H., Cai, S.J., Shafiee, A., & Elliston, K.O. (1997) Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506. *Eur. J. Biochem.* **244,** 74-80.

[518] Motamedi, H. & Shafiee, A. (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* **256,** 528-534.

[519] Schwecke, T., Aparicio, J.F., Molnar, I., Konig, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortes, J., & Lester, J.B. (1995) The biosynthetic gene cluster for the

polyketide immunosuppressant rapamycin. *Proc. Natl Acad. Sci. USA* **92,** 7839-7843.
[520] Petrovics, G., Putnoky, P., Reuhs, B., Kim, J., Thorp, T.A., Noel, K.D., Carlson, R.W., & Kondorosi, A. (1993) The presence of a novel type of surface polysaccharide in Rhizobium meliloti requires a new fatty acid synthase-like gene cluster involved in symbiotic nodule development. *Mol. Microbiol.* **8,** 1083-1094.
[521] Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Puhler, A., Purnelle, B., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S., & Galibert, F. (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl Acad. Sci. USA* **98,** 9877-9882.
[522] Jiralerspong, S., Rangaswamy, V., Bender, C.L., & Parry, R.J. (2001) Analysis of the enzymatic domains in the modular portion of the coronafacic acid polyketide synthase. *Gene* **270,** 191-200.
[523] Azad, A.K., Sirakova, T.D., Fernandes, N.D., & Kolattukudy, P.E. (1997) Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J. Biol. Chem.* **272,** 16741-16745.
[524] Sirakova, T.D., Thirumala, A.K., Dubey, V.S., Sprecher, H., & Kolattukudy, P.E. (2001) The Mycobacterium tuberculosis pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276,** 16833-16839.
[525] Proctor, R.H., Desjardins, A.E., Plattner, R.D., & Hohn, T.M. (1999) A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. *Fungal Genet. Biol.* **27,** 100-112.
[526] Yang, G., Rose, M.S., Turgeon, B.G., & Yoder, O.C. (1996) A polyketide synthase is required for fungal virulence and production of the polyketide T-toxin. *Plant Cell* **8,** 2139-2150.
[527] Zhu, G., Marchewka, M.J., Woods, K.M., Upton, S.J., & Keithly, J.S. (2000) Molecular analysis of a Type I fatty acid synthase in *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* **105,** 253-260.