



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Oxigenasas tipo Rieske/mononuclear en el metabolismo de colina en espinaca y *Pseudomonas aeruginosa*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M.C. JAVIER CARRILLO CAMPOS

TUTOR PRINCIPAL:

DRA. ROSARIO ADELAIDA MUÑOZ CLARES
FACULTAD DE QUÍMICA

MIEMBROS DEL COMITÉ TUTOR:
DR. JUAN PABLO PARDO VÁZQUEZ
FACULTAD DE MEDICINA

DRA. ADELA RODRÍGUEZ ROMERO
INSTITUTO DE QUÍMICA

Ciudad de México. Junio, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

A la Dra. Rosario Muñoz Clares y a los miembros del comité tutor, la Dra. Adela Rodríguez Romero y el Dr. Juan Pablo Pardo Vázquez

Al maestro Carlos Mújica Jiménez por todo su apoyo en la parte experimental de este trabajo de tesis

A la Dra. Lilian González Segura por todo su apoyo en los intentos de cristalización de las enzimas estudiadas en este trabajo de tesis

A la Facultad de Química de la Universidad Nacional Autónoma de México (UNAM)

Al CONACYT por el financiamiento a través del proyecto INFR-252123 cuya responsable es la Dra. Rosario A. Muñoz Clares y por la beca de posgrado otorgada

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) por el apoyo económico otorgado para la asistencia a congresos y a una estancia en el extranjero

A PAPIIT por el financiamiento a través de los proyectos IN217814 y IN220317 cuya responsable es la Dra. Rosario A. Muñoz Clares

A PAPIIT por la beca otorgada para conclusión del doctorado a través del proyecto IN218819 cuyo responsable es el Dr. Héctor Riveros Rosas

Al proyecto de supercómputo LANCAD-UNAM-DGTIC-215

ÍNDICE

RESUMEN	1
ABSTRACT.....	4
I. INTRODUCCIÓN.....	6
I.1. Oxigenasas	6
I.2. Oxigenasas tipo Rieske/mononuclear	7
I.2.1. Características estructurales de las oxigenasas tipo Rieske/mononuclear.....	9
I.2.2. Clasificación, distribución y filogenia	12
I.3. Participación de oxigenasas Rieske/mononuclear en el metabolismo de colina en plantas	15
I.4. Oxigenasas tipo Rieske/mononuclear en el catabolismo de colina en bacterias.....	18
II. PLANTEAMIENTO DEL PROBLEMA	21
III. OBJETIVO GENERAL	23
IV. OBJETIVOS PARTICULARES	23
V. MATERIALES Y MÉTODOS.....	24
V.1. Sub-clonación del gen <i>CMO</i> de <i>S. oleraceae</i> (espinaca) y clonación de los genes <i>gbcA</i> y <i>gbcB</i> de <i>P. aeruginosa</i> PAO1.....	24
V.2. Expresión y purificación de las proteínas CMO, GbcA y GbcB recombinantes.	26
V.3. Determinación del estado oligomérico de las enzimas recombinantes	27
V.4. Comprobación de la integridad del centro Rieske de las enzimas recombinantes	28

V.5. Ensayos de actividad enzimática de un único ciclo catalítico de la CMO	29
V.6. Ensayos de actividad enzimática multiciclo de GbcA y GbcB	30
V.7. Comparación de secuencias de aminoácidos y análisis filogenético de las CMOs	31
V.8. Generación de modelos tridimensionales por homología de las enzimas CMO y GbcA y acoplamiento molecular de posibles sustratos.....	32
VI. RESULTADOS Y DISCUSIÓN	34
VI.1. Colina monooxigenasa (CMO).....	34
VI.1.1. Aspectos evolutivos de las proteínas CMO.....	34
VI.1.2. Obtención de la enzima <i>SoCMO2</i> recombinante y determinación del estado oligomérico	41
VI.1.3. Comprobación de integridad de los centros Rieske de la <i>SoCMO2</i> madura sin etiqueta de histidinas.....	42
VI.1.4. Cinética de un único ciclo catalítico de la <i>SoCMO2</i>	43
VI.1.5. Modelos por homología de las dos isozimas CMO de espinaca (<i>SoCMO1</i> y <i>SoCMO2</i>).....	46
VI.2. Glicina betaína monooxigenasa (GbcA) y glicina betaína monooxigenasa reductasa (GbcB) de <i>P. aeruginosa</i>	52
VI.2.1. Obtención de las proteínas recombinantes GbcA y GbcB y determinación del estado oligomérico de las enzimas.	52
VI.2.2. Caracterización cinética de la enzima recombinante GbcB	54

VI.2.3. Caracterización cinética de la enzima recombinante GbcA	56
VI.2.4. Modelo por homología de la enzima GbcA	58
VII. CONCLUSIONES	61
VIII. PERSPECTIVAS	63
IX. REFERENCIAS.....	65
X.ANEXO.....	80

ÍNDICE DE FIGURAS

Figura 1. Representación esquemática de los tipos de oxigenasas Rieske/mononuclear.	8
Figura 2. Coordinación de los centros metálicos en las oxigenasas Rieske/mononuclear.	10
Figura 3. Representación en lazos del heterohexámero de la naftaleno dioxigenasa (PDB: 2DE6).	11
Figura 4. Transferencia de electrones entre los centros metálicos de las oxigenasas Rieske/mononuclear.	12
Figura 5. Biosíntesis de GB en plantas.	16
Figura 6. Representación esquemática de la transferencia de electrones entre la ferredoxina y la colina monooxigenasa.	17
Figura 7. Catabolismo de GB en bacterias.	18
Figura 8. Reacción catalizada por GbcA.	20
Figura 9. Representación esquemática de la transferencia de electrones entre GbcB y GbcA.	20
Figura 10. Árbol filogenético de las CMO.	37
Figura 11. Contexto genómico de microorganismos con proteínas CMO-like.	39
Figura 12. Árboles filogenéticos de los dos dominios de las CMO.	40
Figura 13. Purificación de <i>SoCMO2</i> madura sin etiqueta de histidinas por cromatografía de intercambio iónico.	41
Figura 14. Perfil de elución de la enzima <i>SoCMO2</i> de la cromatografía de exclusión molecular.	42
Figura 15. Espectro UV-visible de la <i>SoCMO2</i>	43
Figura 16. Cinética de un solo ciclo catalítico de <i>SoCMO2</i>	46

Figura 17. Modelo por homología de las <i>SoCMO</i>	50
Figura 18. Predicción del sitio activo de <i>SoCMO</i> en los modelos por homología.	51
Figura 19. Purificación de <i>GbcA</i>	52
Figura 20. Purificación de <i>GbcB</i>	53
Figura 21. Perfil de elución de la enzima <i>GbcA</i> de la cromatografía de exclusión molecular.	54
Figura 22. Curva de saturación de <i>GbcB</i> con NADH.....	55
Figura 23. Curva de saturación de <i>GbcA</i> con GB.	57
Figura 24. Especificidad por el sustrato de <i>GbcA</i>	58
Figura 25. Modelo por homología de <i>GbcA</i>	60

ÍNDICE DE TABLAS

Tabla 1. Clasificación de las oxigenasas Rieske/mononuclear en la base de datos de dominios conservados (CDD) del NCBI.	13
Tabla 2. Evaluación de los modelos por homología de <i>SoCMO1</i> y <i>SoCMO2</i> con el protocolo Rd.HMM	48
Tabla 3. Parámetros cinéticos aparentes de la enzima GbcB con NADH.....	55
Tabla 4. Parámetros cinéticos aparentes de la enzima GbcA con GB.	57

ABREVIATURAS

Å	Armstrongs
Da	Daltones
FAD/FMN	Dinucleótido de flavina y adenina/Mononucleótido de flavina y adenina
HEPES	Ácido 4-(2-hidroxiethyl)piperazin-1-iletanosulfónico
IPTG	Isopropil- β -D-tiogalactopiranosido
KCl	Cloruro de Potasio
kDa	KiloDaltones
KOH	Hidróxido de Potasio
L-DOPA	L-3,4-dihidroxifenilalanina
NADH	Dinucleótido de nicotinamida y adenina (forma reducida)
NADPH	Dinucleótido de nicotinamida y adenina fosfato (forma reducida)
NCBI	Centro Nacional para la Información Biotecnológica
Ni-NTA	Agarosa de ácido nitrilotriacético y níquel
PVDF	Fluoruro de polivinilideno
SDS-PAGE	Electroforesis desnaturalizante en gel de prolicarilamida en presencia de dodecil sulfato de sodio

RESUMEN

Las oxigenasas Rieske/mononuclear son enzimas que poseen en su estructura un centro tipo Rieske [2Fe-2S] y un átomo de Fe mononuclear no hémico. Estas enzimas catalizan la oxidación de compuestos incorporando uno o dos átomos de oxígeno, dependiendo de la enzima. En este trabajo de tesis se estudiaron dos enzimas Rieske/mononuclear involucradas en el catabolismo de colina: la colina monooxigenasa de espinaca (*SoCMO2*) que participa en la ruta de biosíntesis del osmoprotector glicina betaína (GB) y la glicina betaína monooxigenasa de *Pseudomonas aeruginosa* PAO1 (*GbcA*) que participa en el catabolismo de GB, la cual en este organismo es un intermediario en la degradación de colina. La primera juega un papel muy importante en la supervivencia de ciertas plantas en ambientes con estrés osmótico, es decir sequía, suelos salinos o frío, mientras que la segunda probablemente juegue un papel importante en la fisiopatología de *P. aeruginosa*, un patógeno oportunista altamente resistente a antibióticos que constituye un grave problema de salud debido a las infecciones que produce especialmente en pacientes con fibrosis quística y en pacientes inmunocomprometidos. *P. aeruginosa* es capaz de utilizar a colina como única fuente de carbono y nitrógeno y se sabe que mantiene un equilibrio entre la síntesis y catabolismo de GB a partir de colina que le permite establecerse y prevalecer en los sitios de infección.

En los estudios con las CMO, encontramos que la isozima a la que llamamos CMO1 está presentes en todas las plantas de genoma conocido, a pesar de que solo algunas plantas son capaces de sintetizar GB. En forma interesante, una segunda isozima, a la que llamamos CMO2, solo está presente en la familia Amaranthaceae, a la cual pertenece la espinaca. Esta isozima (CMO2) es el resultado de un evento de duplicación génica y de una evolución acelerada del gen duplicado, probablemente debido a una fuerte presión de selección en los ambientes naturales en

los que crecen las amarantáceas, lo que hizo aparecer en esta isozima la actividad colina monooxigenasa detectada en la enzima de espinaca (*SoCMO2*) y que es la que participa en la síntesis de GB. Para profundizar en el conocimiento de estas enzimas, obtuvimos la *SoCMO2* recombinante pura y activa, lo que hasta la fecha no se ha reportado por ningún otro grupo de investigación. En estudios de cinética rápida encontramos que además de colina puede oxidar a la prolina betaína y al ácido salicílico. Estos hallazgos son importantes no solo porque aportan datos novedosos sobre una enzima de gran relevancia en la respuesta de ciertas plantas al estrés osmótico, sino además porque pensamos que ayudarán a descubrir el sustrato y la función fisiológica de las CMO1 de las que derivaron. Nuestros numerosos intentos de cristalizar la enzima *SoCMO2* no dieron resultado, pero mediante modelos por homología logramos identificar residuos del sitio activo tanto de las CMO1 como de las CMO2, encontrando diferencias importantes entre estos residuos que pensamos determinan su especificidad por el sustrato. La naturaleza de estos residuos es consistente con el uso de colina por las CMO2 y con el uso de otro sustrato, probablemente un compuesto aromático, por las CMO1, lo que apoya nuestra hipótesis de que las CMO1 participan en una ruta metabólica aún desconocida oxidando un compuesto distinto a colina, ya que están presentes en plantas que no son capaces de sintetizar GB. La actividad de las CMO1 también debe ser relevante para la fisiología de las plantas dado que está presente en todas. Sin embargo, hasta la fecha no se ha caracterizado ni bioquímica ni estructuralmente ninguna de las CMO1, por lo que en este trabajo obtuvimos la enzima de espinaca (*SoCMO1*) recombinante pura y se inició su caracterización bioquímica.

Respecto a la *GbcA* obtuvimos por primera vez la enzima recombinante y la caracterizamos cinéticamente, corroborándose su función propuesta y encontrándose que es muy específica para GB y no puede usar como sustrato otros compuestos de amonio cuaternario. Como donador de

electrones, esta enzima utiliza una reductasa (GbcB), la que también se obtuvo por primera vez recombinante y activa en este trabajo, lo que permitió realizar los estudios cinéticos de GbcA.

ABSTRACT

Rieske/mononuclear oxygenases are enzymes containing a Rieske [2Fe-2S] type center and a non-heme mononuclear iron atom, these enzymes catalyze the oxidation of their substrates incorporating one or two oxygen atoms to the product, depending on the type of enzyme. In this thesis work, two Rieske/mononuclear oxygenases that are involved in the catabolism of choline were studied: choline monooxygenase from spinach (*SoCMO2*) that participates in the biosynthesis of the osmoprotectant glycine betaine (GB) and the enzyme glycine betaine monooxygenase from *Pseudomonas aeruginosa* PAO1 (*GbcA*) known to take part in the catabolism of GB which is an intermediate of choline catabolism in this microorganism. Choline monooxygenase (CMO) plays a very important role in the survival of certain plants that must cope with osmotic stress caused by drought, saline soils and cold, *GbcA* probably plays an important role in the physiopathology of *P. aeruginosa*, which is an opportunistic pathogen highly resistant to antibiotics. This bacteria constitutes a serious health problem causing infections to patients with cystic fibrosis and immunocompromised patients. *P. aeruginosa* is able to use choline as a sole source of carbon and nitrogen and it is known to be able to maintain a balance between the synthesis and the catabolism of GB, this balance allows this microorganism to establish and prevail.

From the study of the CMO enzymes, the isozyme here designated as *CMO1* was found in the genome of all plants that have been sequenced, even though only some plants can synthesize GB. We also found a second isozyme, which we called *CMO2*, only present in the Amaranthaceae, the spinach family. This second isozyme is the result of a genic duplication event and of a high rate evolution of the duplicated gene, this was probably the result of a strong selection pressure caused by the natural habitats in which the plants from the Amaranthaceae family live in, this

caused the emergence of the choline monooxygenase activity in this second isozyme, the one that participates in the biosynthesis of GB. Here, for the first time, a pure and active recombinant *SoCMO2* is reported. We performed fast kinetics measurements on this enzyme catalyzed reaction and found that beside choline it can oxidize both proline betaine and salicylic acid. These findings are, not only novel, but help us to understand better the plant responses to osmotic stress. This information could also help to elucidate the substrate and the physiological function of the CMO1 proteins. Crystallization trials of the *SoCMO2* protein were unsuccessful, and homology models of both enzymes CMO1 and CMO2 were produced and analyzed to compare their potential active sites, where important differences were found. The nature of the active site of the CMO2 enzymes is consistent with the use of choline as substrate and the one of CMO1 is consistent with the probable use of an aromatic compound, this support our hypothesis that CMO1 enzymes participate in a metabolic route yet unknown, oxidizing a compound, different from choline. The activity of the CMO1 enzymes is probably also relevant for the physiology of plants because is present in all plants with a known genome. Nevertheless, to date no CMO1 enzyme has been biochemically or structurally characterized, for that reason we obtained the pure and recombinant *SoCMO1* enzyme and we started its biochemical characterization.

Regarding GbcA, we also obtained for the first time the recombinant and pure enzyme and we performed a kinetics characterization, we confirmed its proposed function and found it to be highly specific for GB, as no other quaternary ammonium compound was oxidized by GbcA. As an electron donor, GbcA uses a reductase (GbcB) which we also obtained for the first time as a recombinant active enzyme, this allow us to perform the studies with GbcA.

I. INTRODUCCIÓN.

I.1. Oxigenasas

Las oxigenasas son enzimas oxidoreductasas (EC 1.13 o EC 1.14) que oxidan sus sustratos introduciendo uno (monooxigenasas) o dos (dioxigenasas) átomos de oxígeno a partir del oxígeno diatómico molecular. Los átomos de oxígeno pueden ser transferidos directamente al sustrato (EC 1.13) o bien mediante un intermediario (EC 1.14), usualmente un elemento metálico, el cual, a través de cambios en su estado de oxidación, activa al oxígeno para que ocurra la reacción. Estas enzimas fueron descritas por primera vez en un estudio sobre la fenolasa (Mason *et al.*, 1955) y otro sobre la pirocatecasa (Hayaishi *et al.*, 1955), enzimas que en microorganismos y animales participan en el metabolismo de ciertos compuestos como el catecol (Kojima *et al.*, 1967) y la L-DOPA (Prabhakaran *et al.*, 1968). Desde entonces se ha reportado su participación en diversas rutas metabólicas, como por ejemplo la oxidación de compuestos aromáticos en bacterias (Mallick *et al.*, 2011) y la síntesis y catabolismo de diversos metabolitos en plantas (Mitchell & Weng, 2019).

Las oxigenasas se pueden clasificar en función del cofactor que utilizan. A continuación, se enlistan los distintos tipos de cofactores utilizados por estas enzimas:

1. Dependientes de FAD/FMN. Este grupo de enzimas es muy amplio al igual que el número de reacciones enzimáticas que catalizan. El número de este tipo de oxigenasas es tal que se subclasifican en 7 grupos, atendiendo a criterios como cuál es la molécula donadora de electrones, el tipo de reacción que catalizan o el plegamiento de la enzima (van Berkel *et al.*, 2006; Huijbers *et al.*, 2014).
2. Dependientes de cobre. Entre ellas se encuentran enzimas como la quercetinasasa, que cataliza la oxidación del compuesto flavonoide quercetina (Oka & Simpson, 1971), la dopamina- β -

monooxigenasa, que participa en la biosíntesis de la norepinefrina (Levin *et al.*, 1960; Yoshizawa *et al.*, 2006) y la tirosinasa que participa en la biosíntesis de melanina y en el metabolismo de compuestos fenólicos (Mayer & Harel, 1979; Iozumi *et al.*, 1993).

3. Dependientes de hemo: Estas oxigenasas tienen una función biológica muy diversa, por ejemplo, participan en sistemas de transporte de electrones, hidroxilación de alcanos, metabolismo del grupo hemo y síntesis de prostaglandinas (Sono *et al.*, 1996).

4. Dependientes de Fe no hémico: Forman dos grupos; las enzimas que contienen dos átomos de Fe en su sitio activo y aquellas que solo tienen uno. En aquellas con dos átomos de Fe, estos están coordinados por cuatro residuos de histidinas y un residuo de aspártico, y a su vez se dividen en cuatro grupos acorde a su estructura (Nordlund & Ekland, 1995; Leahy *et al.*, 2003). Las enzimas que solo tienen un átomo de Fe (Fe mononuclear no hémico) se clasifican acorde al donador de electrones que usan, el cual puede ser un centro Rieske [2Fe-2S], una molécula de α -acetoglutarato o una molécula de pterina (Bruijninx *et al.*, 2008).

I.2. Oxigenasas tipo Rieske/mononuclear

Las monooxigenasas y dioxigenasas Rieske/mononuclear son enzimas bidominio que tienen en su dominio N-terminal un centro Rieske [2Fe-2S] y en su dominio C-terminal un Fe mononuclear no hémico. Estas enzimas fueron descritas por primera vez como parte del sistema de degradación de benceno en bacterias del género *Pseudomonas* (Axcell & Geary, 1973 y 1975). Estas oxigenasas dependen de un sistema de transporte de electrones y en función del número de componentes de este sistema se pueden clasificar como de dos o tres componentes (Mason & Cammack, 1992; Butler & Mason, 1997). En el caso de tres componentes, estos son una ferredoxina reductasa, una ferredoxina tipo Rieske y la oxigenasa (Fig. 1A), mientras que, en el caso de dos componentes, solo hay una reductasa y la oxigenasa (Fig. 1B). En cualquiera de los

dos sistemas, el NAD(P)H es el que aporta los electrones (Mason & Cammack, 1992; Butler & Mason, 1997). El proceso de transferencia de electrones empieza con la reducción del FAD de la ferredoxina reductasa, seguida de la reducción del centro Rieske [2Fe-2S] de la oxigenasa o el centro Rieske [2Fe-2S] de la ferredoxina en los sistemas de tres componentes. En estos últimos, la ferredoxina reduce el centro Rieske [2Fe-2S] de la oxigenasa.

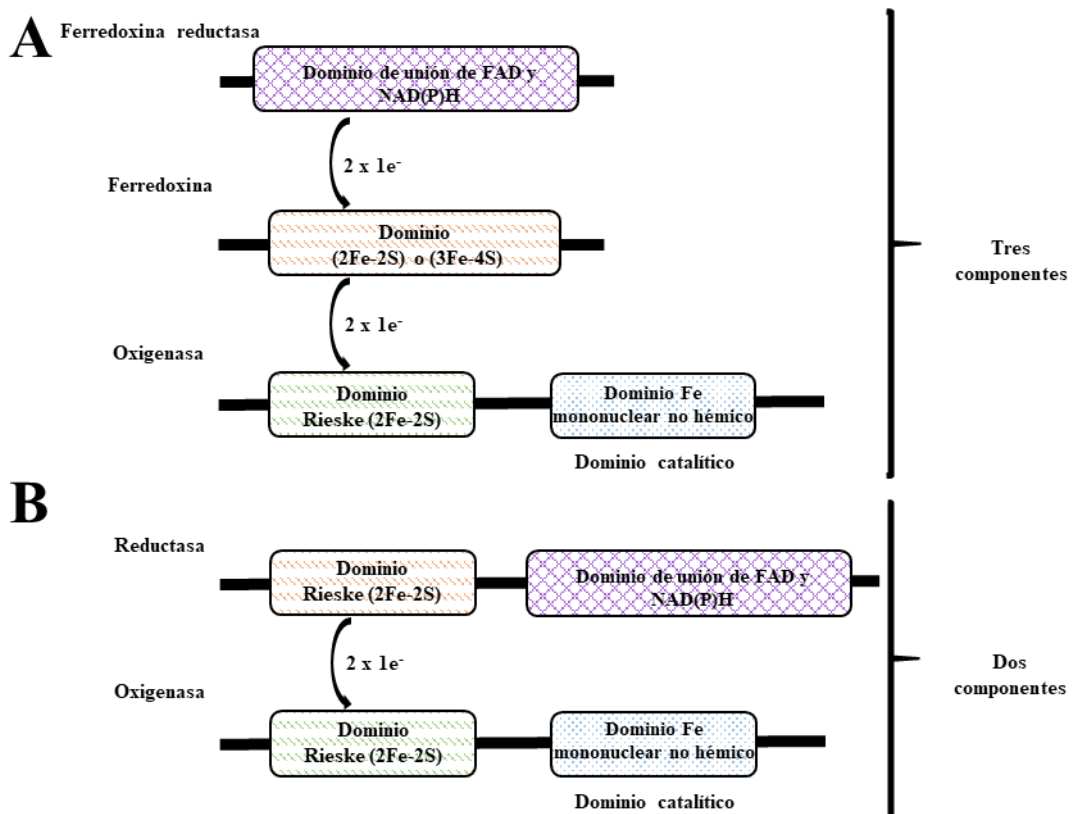


Figura 1. Representación esquemática de los tipos de oxigenasas Rieske/mononuclear. (A) de tres componentes y (B) de dos componentes. Los electrones llegan a los centros Rieske [2Fe-2S] y al Fe mononuclear de la oxigenasa de uno en uno, pero se necesitan dos para que ocurra un ciclo catalítico.

I.2.1. Características estructurales de las oxigenasas tipo Rieske/mononuclear.

El centro Rieske [2Fe-2S] fue descrito inicialmente como parte del citocromo *bc* de la mitocondria de células de res (Rieske *et al.*, 1964) y consiste en una variante de los centros [2Fe-2S] clásicos en el que un átomo de Fe está coordinado por dos residuos de cisteína y el otro átomo de Fe lo está por dos residuos de histidina (Fig. 2A), a diferencia de los clásicos en los que los cuatro residuos que coordinan a los dos átomos de Fe son cisteínas. Como en el resto de los centros Fe-S, los átomos de azufre no están coordinados por ningún residuo de la proteína. Este centro se encuentra en el dominio que forma el amino terminal de la proteína. Los cuatro residuos de coordinación de los átomos de Fe están totalmente conservados en estas enzimas y son esenciales para la unión del Fe y, por lo tanto, para que pueda ocurrir la transferencia de electrones (Davidson *et al.*, 1992; Van Doren *et al.*, 1993). La secuencia de la región de unión del centro Rieske, CisXHis_nCisXXHis, está conservada y es característica de estos.

Los centros Rieske poseen un potencial de óxido-reducción que varía dependiendo de las interacciones que tienen lugar entre los residuos que estructuralmente están cercanos a los residuos que los coordinan (Colbert *et al.*, 2000). Existen centros Rieske de alto potencial redox como el del citocromo *bc* (+300 mV) y otros de menor potencial como el de las ferredoxinas tipo Rieske (-150 mV) y el de las oxigenasas Rieske/mononuclear (~ -100 mV) (Colbert *et al.*, 2000). A pesar de tener dos átomos de Fe, estos centros reciben y transfieren los electrones de uno en uno. El Fe mononuclear no hémico está coordinado por dos residuos de histidina y un residuo de aspártico totalmente conservados (Fig. 2B), este átomo se encuentra en el dominio que forma el carboxilo terminal de la proteína. A este arreglo se le conoce como la tríada 2-His-1-carboxilato. En el dominio del Fe mononuclear es donde tiene lugar la activación del oxígeno y la reacción enzimática, no solo de las oxigenasas Rieske/mononuclear sino de los otros tipos de oxigenasas

que poseen un Fe mononuclear no hémico (Koehtop *et al.*, 2005). El átomo de Fe se coordina con una geometría de bipirámide cuadrada con los tres residuos previamente mencionados y tres moléculas de agua. En algunos casos la coordinación con el residuo aspártico es bidentada, uniendo solo dos moléculas de agua, pero conservando la misma geometría (Hegg & Que, 1997; Que, 2000).

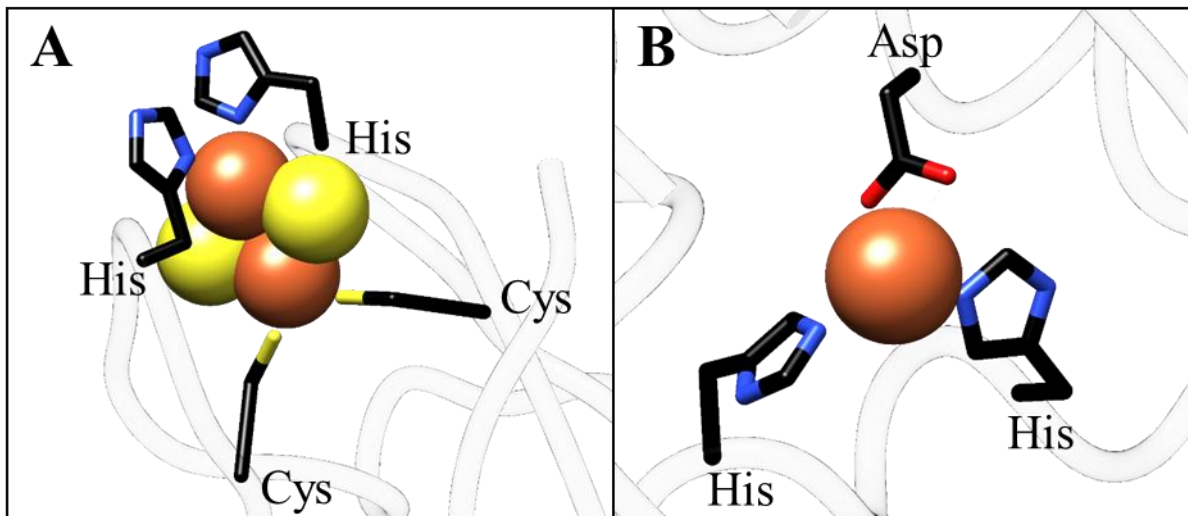


Figura 2. Coordinación de los centros metálicos en las oxigenasas Rieske/mononuclear. A) Centro Rieske [2Fe-2S]. B) Fe mononuclear no hémico. Los átomos de azufre y hierro se muestran como esferas de color amarillo y naranja respectivamente. Los residuos de aminoácidos de la proteína que coordinan a los Fe se representan como varillas, con carbonos en color negro, nitrógenos en azul y oxígenos en rojo. La figura se realizó con el software UCSF Chimera (Pettersen *et al.*, 2004).

La estructura de las oxigenasas Rieske/mononuclear consiste en un homotrímero de subunidades α catalíticas (α_3) (Fig. 3A) o bien en un heterohexámero formado por tres subunidades α y tres subunidades β ($\alpha_3\beta_3$) (Fig. 3B). La función precisa de la subunidad β se desconoce hasta la fecha, pero se sabe que no es esencial para la actividad catalítica por lo que se ha propuesto que puede ayudar a la estabilidad estructural del heterohexámero, o bien regular la actividad enzimática de las subunidades α (Hurtubise *et al.*, 1998; Kauppi *et al.*, 1998).

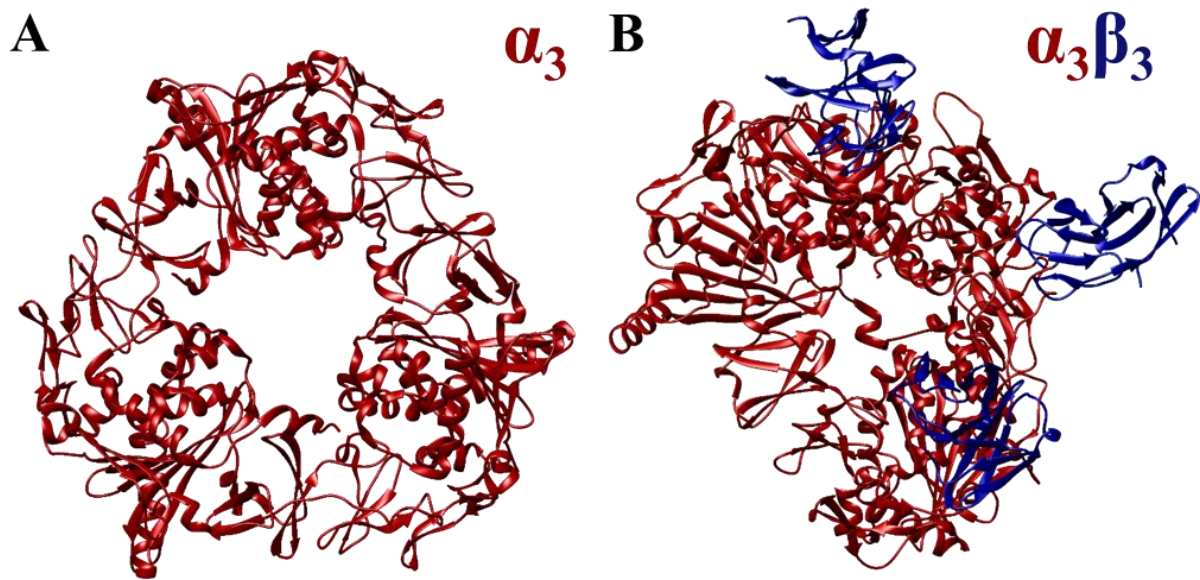


Figura 3. Representación en lazos del heterohexámero de la naftaleno dioxigenasa (PDB: 2DE6). (A) Homotrímero formado por subunidades α . (B) Heterohexámero formado por las subunidades α y β . La figura se realizó con el software UCSF Chimera (Pettersen *et al.*, 2004).

Como parte del ciclo catalítico, el centro Rieske transfiere electrones al Fe mononuclear para la activación del oxígeno, por lo que dada la gran separación de estos centros metálicos dentro de un mismo monómero (Fig. 4A) la transferencia de electrones tiene lugar desde el centro Rieske de una subunidad al Fe mononuclear de la subunidad vecina que está mucho más cercano (Fig. 4B). Esta transferencia está mediada por un residuo de aspartato o glutamato de la subunidad a la que pertenece el centro Fe mononuclear que recibe el electrón (Parales *et al.*, 1999). La mutación de cualquiera de los residuos que coordinan a los átomos de Fe del centro Rieske, los que coordinan al Fe mononuclear o del residuo ácido que participa en la transferencia de electrones entre los centros metálicos produce la pérdida de actividad de estas enzimas (Van Doren *et al.*, 1993; Jiang *et al.*, 1996; Parales *et al.*, 1999).

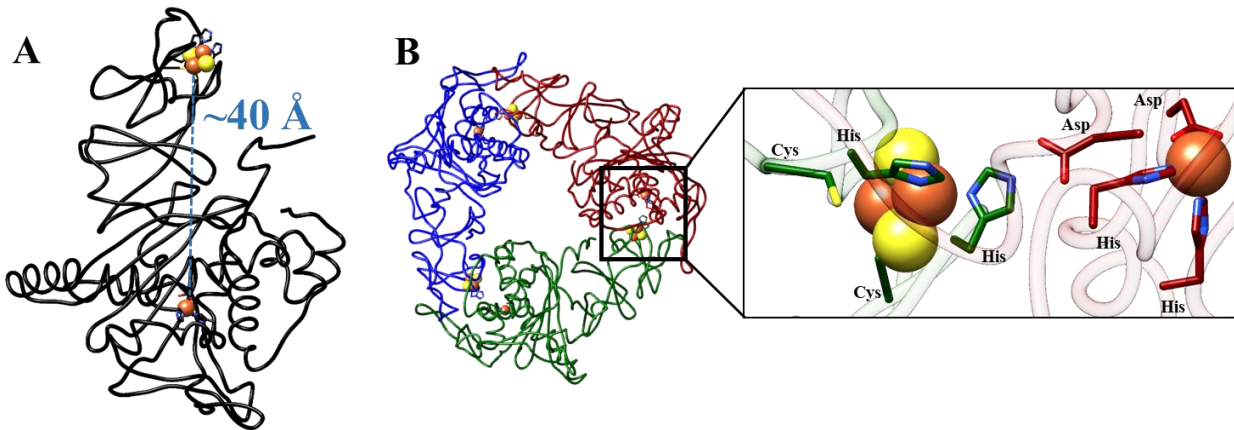


Figura 4. Transferencia de electrones entre los centros metálicos de las oxigenasas Rieske/mononuclear. **A)** Representación en lazos de un monómero de la oxigenasa de prolina betaína (PDB: 3VCA); con una línea punteada se indica la distancia (en angstroms) entre sus dos centros metálicos. **B)** Representación en lazos del trímero de la oxigenasa de prolina betaína y acercamiento mostrando los residuos de aminoácidos que coordinan a los centros metálicos y del residuo de aspártico que asiste la transferencia de electrones desde el centro Rieske de una subunidad al Fe mononuclear no hémico de su subunidad vecina. Los residuos de aminoácidos se representan como varillas. Los carbonos se muestran en colores verde, azul o rojo cobrizo, dependiendo de la subunidad, los nitrógenos en azul y los oxígenos en rojo. Los átomos de azufre y hierro se muestran como esferas de color amarillo y naranja respectivamente. La figura se realizó con el software UCSF Chimera (Pettersen *et al.*, 2004).

I.2.2. Clasificación, distribución y filogenia

La base de datos de dominios conservados CDD (Conserved Domain Database, <https://www.ncbi.nlm.nih.gov/cdd>; Marchler-Bauer *et al.*, 2017) hace una clasificación por dominio de las oxigenasas Rieske/mononuclear que permite relacionar la información que se tiene acerca de su estructura, función y evolución. Filogenéticamente, el dominio del centro Rieske pertenece a la familia denominada cd03469_Rieske_RO_alpha_N, que hasta la fecha contiene 16 subfamilias (Tabla 1) y el dominio del Fe mononuclear no hémico pertenece a la familia denominada cd00680_RHO_alpha_C que hasta la fecha contiene 10 subfamilias (Tabla 1).

Tabla 1. Clasificación de las oxigenasas Rieske/mononuclear en la base de datos de dominios conservados (CDD) del NCBI. (A) Dominio Rieske, perteneciente a la familia cd03469_Rieske_RO_alpha_N (B) Dominio Fe mononuclear no hémico, perteneciente a la familia cd00680_RHO_alpha_C.

Familias	
cd03469 Rieske_RO_Alpha_N	cd00680 RHO_alpha_C
Subfamilias	
cd03472 Rieske_RO_Alpha_BPDO_like	cd08878 RHO_alpha_C_DMO-like
cd03479 Rieske_RO_Alpha_PhDO_like	cd08879 RHO_alpha_C_AntDO-like
cd03480 Rieske_RO_Alpha_PaO	cd08880 RHO_alpha_C_ahdA1c-like
cd03531 Rieske_RO_Alpha_KSH	cd08881 RHO_alpha_C_NDO-like
cd03532 Rieske_RO_Alpha_VanA_DdmcC	cd08882 RHO_alpha_C_MupW-like
cd03535 Rieske_RO_Alpha__NDO	cd08883 RHO_alpha_C_CMO-like
cd03536 Rieske_RO_Alpha_DTDO	cd08884 RHO_alpha_C_GbcA-like
cd03537 Rieske_RO_Alpha_PrnD	cd08885 RHO_alpha_C_1
cd03538 Rieske_RO_Alpha_AntDO	cd08886 RHO_alpha_C_2
cd03539 Rieske_RO_Alpha_S5H	cd08887 RHO_alpha_C_3
cd03541 Rieske_RO_Alpha_CMO	
cd03542 Rieske_RO_Alpha_HBDO	
cd03545 Rieske_RO_Alpha_OHBDO_like	
cd04337 Rieske_RO_Alpha_Cao	
cd04338 Rieske_RO_Alpha_Tic55	

Se ha propuesto otro método de clasificación de estas enzimas que toma como criterio las características estructurales del sistema de transferencia de electrones que alimenta a la oxigenasa y de esta misma. Así, se clasifican según el tipo de reductasa o de ferredoxina con el que trabajan, por ejemplo, si en la reductasa el dominio de unión del FAD/FMN se encuentra en el amino o en el carboxilo terminal, o en el caso de la ferredoxina si el centro metálico es de tipo Rieske [2Fe-2S] o del tipo [3Fe-4S], y si la oxigenasa posee o no una subunidad β (Kweon *et al.*, 2008; Chakraborty *et al.*, 2012).

Las oxigenasas Rieske/mononuclear están ampliamente distribuidas en bacterias (proteobacterias, actinobacterias, algunas cianobacterias y firmicutes) e incluso en algunas arqueas (Chakraborty *et al.*, 2012). Se ha descrito que estas enzimas están presentes también en bacterias termófilas, sin embargo, hasta el momento no se ha reportado la función de alguna de ellas (Chakraborty *et al.*, 2017). La mayoría de las oxigenasas Rieske/mononuclear presentes en bacterias participan en rutas de degradación de compuestos aromáticos policíclicos, como el naftaleno (Ensley *et al.*, 1982; Ensley & Gibson, 1983), carbazol (Sato *et al.*, 1997), bifenilo (Haddock *et al.*, 1995), benzoato (Wolfe *et al.*, 2002), cumeno (Dong *et al.*, 2005), oxoquinolina (Rosche *et al.*, 1995), ftalato (Batie *et al.*, 1987), benzeno (Mason *et al.*, 1997) y tolueno (Jiang *et al.*, 1999). Se han encontrado oxigenasas Rieske/mononuclear en bacterias que degradan compuestos no aromáticos que contienen un amonio cuaternario, como la glicina betaína (GB) (Shao *et al.*, 2018), la prolina betaína (Daughtry *et al.*, 2012), la carnitina (Zhu *et al.*, 2014) y el cloruro de benzalconio (Ertekin *et al.*, 2017). Otros compuestos que son oxidados por oxigenasas Rieske/mononuclear en bacterias son el herbicida dicamba (Chakraborty *et al.*, 2005) y la cafeína (Summers *et al.*, 2012). En eucariontes, estas enzimas no se encuentran tan ampliamente distribuidas. En plantas, hasta la fecha se ha encontrado que participan en la síntesis del osmoprotector GB a partir de colina (Rathinasabapathi *et al.*, 1997), en la formación de clorofila b a partir de la clorofila a (Tanaka *et al.*, 1998) y en el catabolismo de esta (Pruzinska *et al.*, 2003). En animales, participan en la degradación del colesterol (Yoshiyama-Yanagawa *et al.*, 2011) y en algunas levaduras en el catabolismo de colina (Linder, 2014).

Encontrar el mecanismo específico por el cual se originó una proteína con dos o más dominios no es sencillo. En el caso particular de las enzimas Rieske/mononuclear la fusión de dos genes que codificaban a proteínas de función distinta pudiera ser el origen de este tipo de enzimas. El

dominio Rieske pertenece a la superfamilia de proteínas que contienen hierro-azufre (ISP por sus siglas en inglés) cuyos miembros tienen un plegamiento similar y por lo tanto este dominio pudo haber surgido de una proteína ancestral cuya función estuviera relacionada con el transporte de electrones (Schmidt & Shaw, 2001; Chakraborty & Dutta, 2011). En el caso del dominio de unión al Fe mononuclear no hémico, su plegamiento es parecido al de las proteínas de la superfamilia denominada Bet v1, cuyos miembros tienen funciones diversas (Radauer *et al.*, 2008). En particular se han encontrado similitudes estructurales entre la naftaleno dioxigenasa y la proteína PITP, que no tiene actividad catalítica pero media el intercambio de fosfatidilinositol y fosfatidilcolina entre diferentes membranas (Yoder *et al.*, 2002). Tanto el dominio del Fe mononuclear no hémico de la naftaleno dioxigenasa como la PITP poseen una cavidad central, que en el caso de la PITP es donde se unen los fosfolípidos y en el de la dioxigenasa es donde se une el naftaleno (Bashton & Chothia, 2007). Se ha propuesto que la proteína PA1206 de *P. aeruginosa*, de función desconocida, es muy parecida estructuralmente al dominio de unión del Fe mononuclear no hémico de las oxigenasas Rieske/mononuclear (Chakraborty & Dutta, 2011). Sin embargo, ni PITP ni PA1206 unen iones metálicos y tampoco tienen conservados los residuos de unión del Fe mononuclear no hémico, totalmente conservados en las oxigenasas Rieske/mononuclear.

I.3. Participación de oxigenasas Rieske/mononuclear en el metabolismo de colina en plantas

La colina es un alcohol que contiene un amonio cuaternario y es precursor de varios compuestos de importancia biológica. En algunas plantas es utilizada como precursor del compuesto osmoprotector GB, un compuesto de amonio cuaternario cuyas características de carga neutra y

alta solubilidad permiten que se acumule en el citosol hasta elevadas concentraciones, contribuyendo al mantenimiento del potencial osmótico de las células bajo condiciones de deshidratación producidas por la sequía, concentraciones altas de sales en el suelo o el frío (Rhodes & Hanson, 1993; Chen & Murata, 2008). En plantas, GB es sintetizada en dos pasos a partir de la colina: el primero catalizado por la enzima Rieske/mononuclear colina monooxigenasa (CMO) (EC 1.14.15.7) produciendo hidrato de betaína aldehído (Brouquisse *et al.*, 1989) que en solución está en equilibrio con la betaína aldehído (BAL) (Fig. 5), la cual es oxidada en el segundo paso por la enzima betaína aldehído deshidrogenasa (BADH) (EC 1.2.1.8) dando como producto de la reacción a GB (Hanson *et al.*, 1985) (Fig. 5). Esta síntesis tiene lugar en cloroplasto, aunque existen reportes de que CMO es peroxisomal y BADH citosólica en plantas como mangles y cebada (Hibino *et al.*, 2001; Fujiwara *et al.*, 2008).

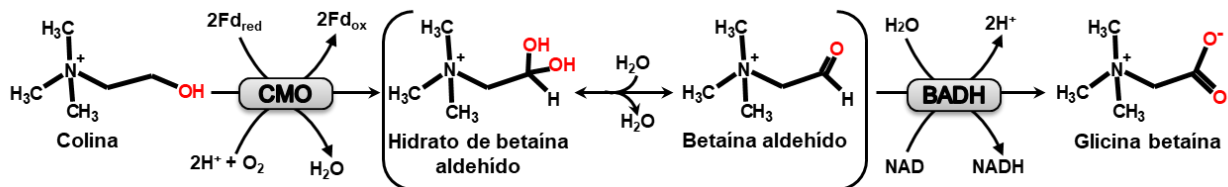


Figura 5. Biosíntesis de GB en plantas. Representación esquemática de la ruta de biosíntesis de la GB, desde la hidroxilación de la colina por la colina monooxigenasa (CMO) hasta la oxidación de la betaína aldehído por la betaína aldehído deshidrogenasa (BADH).

La CMO fue descrita por primera vez en *Spinacia oleracea* (espinaca, SoCMO) (Brouquisse *et al.*, 1989), que es una especie de la familia Amaranthaceae y que se sabe que posee una gran capacidad de acumulación de GB. De hecho, el conocimiento que se tiene hasta ahora de las CMO proviene principalmente del estudio de la SoCMO (Hanson *et al.*, 1985; Burnet *et al.*, 1995; Rathinasabapathi *et al.*, 1997). La expresión de CMO se incrementa en respuesta a estrés

osmótico no sólo en espinaca (Weretilnyk & Hanson, 1989; Burnet *et al.*, 1995) sino también en otras plantas acumuladoras de GB (Meng *et al.*, 2001; Joseph *et al.*, 2013).

La CMO cataliza la hidroxilación de la colina usando oxígeno molecular y utilizando como donador de electrones a la ferredoxina del cloroplasto reducida por el Fotosistema I (Hanson *et al.*, 1985; Burnet *et al.*, 1995) (Fig. 6). La SoCMO posee un péptido señal para cloroplasto de 60 residuos (Cline & Henry, 1996; Rathinasabapathi *et al.*, 1997), lo cual es consistente con el hecho de utilizar a la ferredoxina cloroplástica como donador de electrones. En el caso de las CMO peroxisomales se desconoce cuál es el donador de electrones.

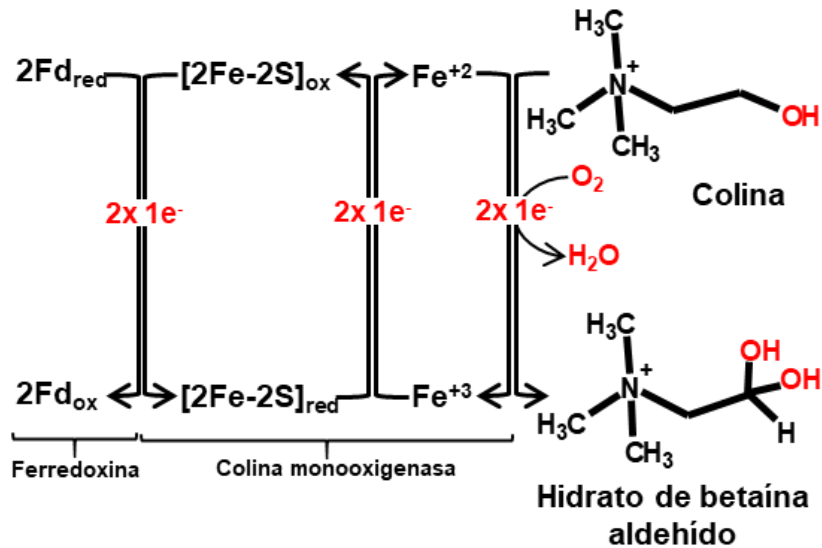


Figura 6. Representación esquemática de la transferencia de electrones entre la ferredoxina y la colina monooxigenasa.

I.4. Oxigenasas tipo Rieske/mononuclear en el catabolismo de colina en bacterias.

En bacterias, la colina es degradada de dos maneras: la primera hasta piruvato, en una ruta que tiene como intermediario a la GB (Kortstee, 1970) (Fig. 7A) y la segunda hasta trimetilamina (Craciun & Balskus, 2012) (Fig. 7B).

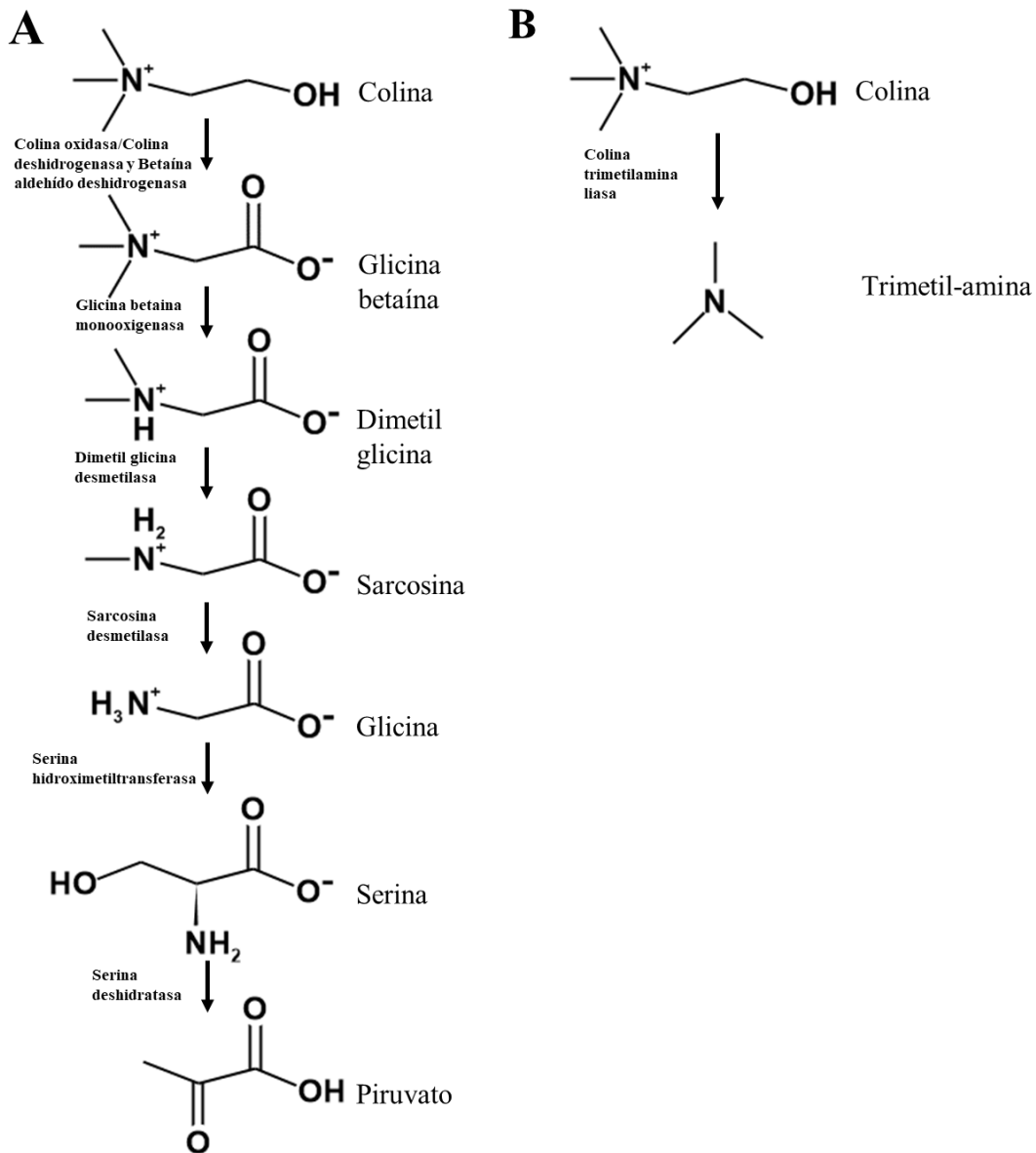


Figura 7. Catabolismo de GB en bacterias. Representación esquemática de las dos rutas de catabolismo de GB.

En bacterias, la GB producida a partir de colina puede no degradarse y acumularse funcionando como un osmoprotector, como en el caso de *Escherichia coli*, o puede ser degradada hasta piruvato (Fig. 7A), funcionando como una fuente de carbono y nitrógeno para la bacteria, lo que ocurre en bacterias de suelo (Kortstee, 1970). El catabolismo de la GB se estudió inicialmente en *Sinorhizobium* (Smith *et al.*, 1988) y si bien se logró demostrar su catabolismo hasta glicina, la enzima responsable del primer paso de esta ruta metabólica no se conocía con certeza. En *Pseudomonas denitrificans* ya se había identificado que la reacción de desmetilación de GB a dimetilglicina en la ruta de producción de la vitamina B12 es catalizada por la enzima betaína homocisteína metiltransferasa (White *et al.*, 1973). En *P. aeruginosa*, la glicina betaína monooxigenasa (GbcA) es la enzima que cataliza esta primera reacción de desmetilación (Fig. 8) en la ruta de degradación de colina (Wargo *et al.*, 2008).

GbcA es una oxigenasa tipo Rieske/mononuclear que forma parte de un sistema de dos componentes en el que una reductasa (GbcB), que tiene FAD y un centro [2Fe-2S] como grupos prostéticos, recibe los electrones del NADH y los transfiere hasta el centro Rieske [2Fe-2S] de GbcA (Fig. 9). Aunque el mecanismo catalítico de GbcA se desconoce, se sabe que pertenece a la misma subfamilia (cd03469/cd08884) que la enzima prolina betaína (estaquidrina) desmetilasa del microorganismo *Sinorhizobium meliloti*. Ambas enzimas catalizan una reacción de hidroxilación que conduce a una desmetilación. En el caso de la prolina betaína desmetilasa de *S. meliloti*, se ha propuesto que el producto de la hidroxilación es inestable y espontáneamente se desmetila (Daughtry *et al.*, 2012). En el caso de GbcA, el mecanismo de reacción probablemente sea muy parecido y la hidroxilación de GB da lugar a un compuesto inestable que pierde un metilo de manera espontánea formando la dimetilglicina.

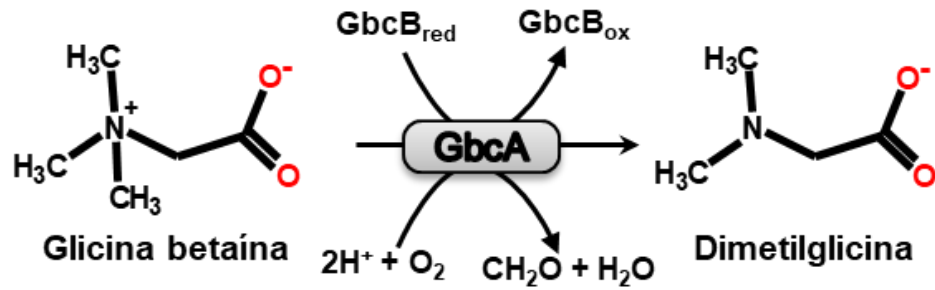


Figura 8. Reacción catalizada por GbcA. Representación esquemática de la reacción de desmetilación llevada a cabo por GbcA. La enzima depende de los electrones donados por GbcB.

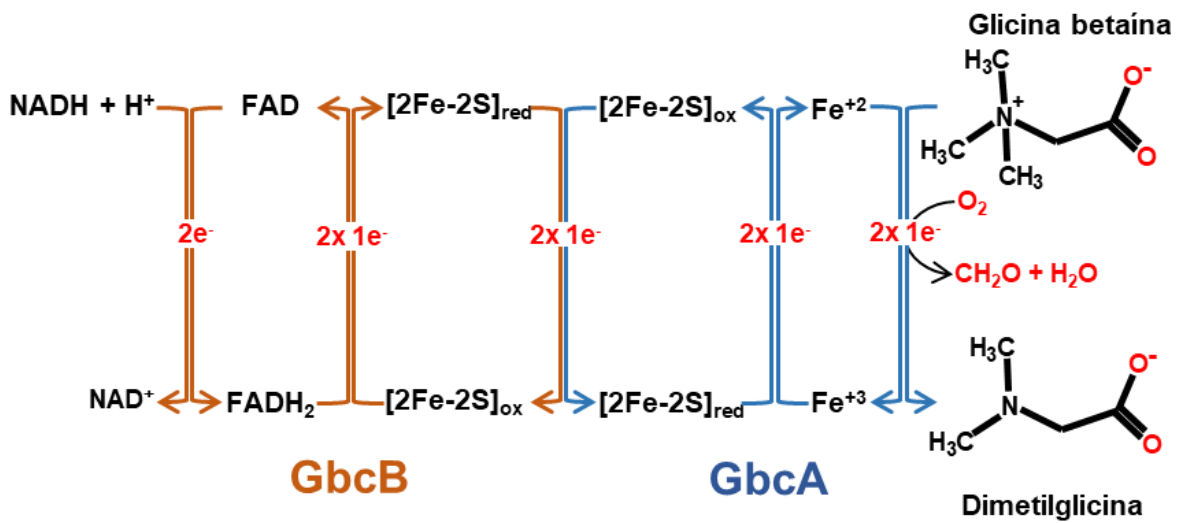


Figura 9. Representación esquemática de la transferencia de electrones entre GbcB y GbcA.

II. PLANTEAMIENTO DEL PROBLEMA

Como se mencionó, la participación de oxigenasas Rieske/mononuclear en el metabolismo de compuestos no aromáticos que poseen un grupo trimetilamonio tiene lugar en plantas y bacterias. En plantas, la colina monooxigenasa (CMO) oxida a la colina en la ruta biosintética del osmoprotector GB y en algunas bacterias, en particular en *P. aeruginosa*, la glicina betaína monooxigenasa llamada también GbcA, participa en el catabolismo de GB. Los sustratos de ambas enzimas (colina y GB) son estructuralmente análogos y están relacionados metabólicamente. Pensamos por ello que el estudio de las dos enzimas se complementaría, es importante mencionar que a pesar de que la actividad enzimática es distinta, ambas enzimas son oxigenasas Rieske/mononuclear. A continuación, se discuten otras razones por las que consideramos de interés el estudiar a estas enzimas en sus respectivos organismos modelo.

El conocimiento funcional y estructural que se tiene hasta el momento de las CMO es muy limitado. Se sabe que en plantas acumuladoras de GB en respuesta a un estrés osmótico participa en la síntesis de este osmoprotector pero se desconoce cuál es su función en aquellas plantas que no acumulan GB en las que pudiera estar catalizando otra reacción. Hasta la fecha, no se conoce a cabalidad la especificidad por el sustrato de las CMO, con excepción de aquellas que usan a colina en la ruta de biosíntesis de GB, tampoco se ha obtenido la estructura cristalográfica de ninguna de estas enzimas.

P. aeruginosa es un importante patógeno oportunista de plantas y animales que sintetiza al osmoprotector GB a partir de colina para lidiar con el estrés osmótico que encuentra frecuentemente en los tejidos que infecta. Además, puede utilizar tanto a colina como a GB como únicas fuentes de carbono, nitrógeno y energía, GB es un intermediario del catabolismo de colina. GbcA forma parte de la ruta catabólica de colina en este microorganismo, la cual no ha

sido estudiada a detalle ya que de las siete enzimas que la forman, hasta la fecha, solo se ha caracterizado bioquímica y estructuralmente a la enzima betaína aldehído deshidrogenasa, que es la que produce la GB, sustrato de GbcA. El estudio de esta enzima permitiría ampliar el conocimiento del catabolismo de colina en *P. aeruginosa*, lo que es de gran interés dada la importancia de este compuesto y sus precursores en la fisiopatología de este patógeno oportunista.

Finalmente, la oxidación de compuestos con amonio cuaternario catalizada por oxigenasas Rieske/mononuclear ha sido escasamente estudiada hasta la fecha (Barry & Challis, 2013), por lo que el estudio de las enzimas CMO y GbcA contribuye también al conocimiento de la diversidad catalítica de las oxigenasas Rieske/mononuclear.

III. OBJETIVO GENERAL

Determinar aspectos cinéticos, estructurales y evolutivos de las oxigenasas Rieske/mononuclear que participan en el metabolismo de colina en plantas y en bacterias.

IV. OBJETIVOS PARTICULARES

Colina monooxigenasa

1. Determinar su actividad enzimática *in vitro* y revisar si es capaz de oxidar otros compuestos distintos a colina.
2. Caracterizarla estructuralmente mediante cristalografía de proteínas y difracción de rayos X, o bien mediante la generación de modelos obtenidos con base en su homología.
3. Establecer las relaciones evolutivas de las CMO de plantas.

Glicina betaína monooxigenasa

1. Determinar su actividad enzimática *in vitro* y revisar si es capaz de oxidar otros compuestos distintos a glicina betaína.
2. Caracterizarla estructuralmente mediante cristalografía de proteínas y difracción de rayos X, o bien mediante la generación de modelos obtenidos con base en su homología.

V. MATERIALES Y MÉTODOS

V.1. Sub-clonación del gen *CMO* de *S. oleraceae* (espinaca) y clonación de los genes *gbcA* y *gbcB* de *P. aeruginosa* PAO1.

Se adquirió el gen sintético *CMO* de espinaca (NCBI RefSeq: XM_021993817.1) de la compañía Genscript, con codones optimizados para su expresión en *E. coli* y sin los 60 residuos correspondientes al péptido señal para cloroplasto (enzima madura), la enzima con el péptido señal tiene 439 aminoácidos la enzima madura corresponde a la secuencia del residuo 61-439. Debido a los problemas que se tuvieron para obtener la proteína soluble y ver su actividad enzimática, se generaron diversas construcciones para expresar la enzima que se enlistan a continuación:

- Sub-clonación del gene que codifica para la enzima madura en el plásmido de sobreexpresión pET22b(+) (Novagen, Alemania), que incorpora etiqueta de histidinas en el C-terminal.
- Sub-clonación del gene que codifica para la enzima truncada (19 residuos del N-terminal de la enzima madura, 80-439 aa) en el plásmido de sobreexpresión pET22b(+).
- Sub-clonación del gene que codifica para la enzima madura en el plásmido de sobreexpresión pET22b(+), incorporando un codón de paro para obtener la enzima sin etiqueta de histidinas
- Sub-clonación del gene que codifica para la enzima truncada en el plásmido de sobreexpresión pET22b(+), incorporando un codón de paro para expresar la enzima sin etiquetas de histidinas.
- Sub-clonación del gene que codifica para la enzima madura en el plásmido de sobreexpresión pET28b(+) (Novagen, Alemania), que incorpora etiqueta de histidinas en el N-terminal.

- Sub-clonación del gene que codifica para la enzima truncada en el plásmido de sobreexpresión pET28b(+).

De estas construcciones, la enzima madura sin etiqueta de histidinas proveniente del plásmido de sobreexpresión pET-22b(+) fue con la que se trabajó en la presente tesis.

En todos los casos, el gen adquirido se subclonó utilizando oligonucleótidos que incorporaran los sitios de reconocimiento de las enzimas de restricción *NdeI* en el oligonucleótido sentido (CMOF1: 5' GGTGGTCATATGGCCGTCGCTGCTCCGTCTTTTC 3') y de la enzima *NotI* en el antisentido (5' GGTGGTGCGGCCGCTTTCAGGGTCTGTT 3'), en aquellas construcciones sin etiqueta de histidinas, se incluyó un codón de paro previo al sitio de reconocimiento de la enzima *NotI*. El gen se amplificó utilizando la reacción en cadena de la DNA polimerasa (PCR) y tanto el producto de la amplificación como el plásmido se digirieron con las enzimas de restricción previamente mencionadas. El gen y plásmido digeridos se ligaron utilizando la enzima T4 DNA ligasa. El producto de ligado se utilizó para transformar células de clonación quimiocompetentes XL10-Gold de *E. coli*.

Se utilizó DNA genómico de la cepa PAO1 de *P. aeruginosa* para amplificar los genes *gbcA* (NCBI RefSeq genomic: NC_002516.2) y *gbcB* (NCBI RefSeq genomic: NC_002516.2) mediante PCR. El DNA genómico se obtuvo utilizando el método del bromuro de hexadeciltrimetilamonio (CTAB) (Park, 2007). El gen *gbcA* se clonó en el vector de sobreexpresión pET-22b(+) utilizando oligonucleótidos que incorporaron los sitios de reconocimiento de las enzimas de restricción *NdeI* en el oligonucleótido sentido (GbcAF1: 5' CATATGGACGTCACCTCCACCCTGAGTC 3') y de la enzima *NotI* en el antisentido (GbcAR1: 5' TATTCTTAGGCGGCCGGCTGCCGGCCACCTTGCG 3'). El gen *gbcB* se clonó en el vector de sobreexpresión pET-28b(+) utilizando oligonucleótidos que incorporaran los sitios de reconocimiento de las enzimas de restricción *NdeI* en el oligonucleótido sentido

(GbcBF1: 5' CGGGAATTCCCATATGTCGAGCAACTTCCTCAATCC 3') y de la enzima *Bam*HI en el antisentido (GbcBR1: 5' CGGGATCCTCAGTAGTCGATGACCACGTCG 3'). Ambos genes se amplificaron utilizando PCR y los productos de la amplificación y los plásmidos se digirieron con las enzimas de restricción previamente mencionadas. Los genes y plásmidos digeridos se ligaron utilizando la enzima T4 DNA ligasa. Los productos de ligado se utilizaron para transformar células de clonación quimiocompetentes XL10-Gold de *E. coli*.

De las células transformadas se extrajo DNA plasmídico utilizando el kit comercial StrataPrep Plasmid Miniprep Kit (Stratagene, EUA) siguiendo el protocolo sugerido por la casa comercial. Este DNA plasmídico se secuenció en la unidad de secuenciación del Instituto de Biotecnología de la UNAM y se confirmó que el marco de lectura y la secuencia nucleotídica fueran correctos. Los plásmidos pET22-CMO, pET22-GbcA y pET28-GbcB se utilizaron para transformar células de expresión quimiocompetentes BL21(DE3) de *E. coli*.

V.2. Expresión y purificación de las proteínas CMO, GbcA y GbcB recombinantes.

Para la obtención de cualquiera de las tres proteínas se utilizaron 4 L de medio LB (Lysogeny Broth) adicionado con el antibiótico correspondiente (kanamicina 0.1 mg/ml para las células transformadas con el plásmido pET28-GbcB y ampicilina 0.1 mg/ml para las células transformadas con los plásmidos pET22-GbcA o pET22-CMO). Cada litro se inoculó con 10 ml de un cultivo de 12 h en medio LB con el antibiótico correspondiente. Los 4 litros ya inoculados, se incubaron a 37 °C con una agitación constante de 200 rpm hasta alcanzar una densidad óptica de 0.6-0.8 a 600 nm. Pasado este tiempo se realizó una inducción con IPTG 0.1 mM y se le añadió también 0.1 mM de sulfato de hierro amonio (sal de Mohr), se continuó el cultivo a 18 °C y agitación constante a 150 rpm durante 18 h. Las células se recolectaron mediante

centrifugación a 4,000 x g, y la pastilla celular de 2 L (2 pastillas celulares por cada 4 L de cultivo procesados) se resuspendió en 50 ml de amortiguador de lisis (HEPES-KOH 50 mM, pH 7.5, KCl 50 mM, glicerol 10% (v/v)) y se almacenó a -70 °C, hasta su uso.

Cuando se llevó a cabo la purificación, las células resuspendidas en amortiguador de lisis se descongelaron y se lisaron utilizando el homogenizador de alta presión Emulsiflex (Avestin, Canadá), y una vez lisadas, se centrifugaron a 12,000 x g por 30 min. En el caso de las proteínas GbcA y GbcB el sobrenadante obtenido se aplicó a una columna empacada con la resina Protino Ni-NTA Agarose (Macherey-Nagel, Alemania) para su purificación mediante cromatografía de afinidad. En el caso de la proteína CMO el sobrenadante se aplicó a una columna empacada con resina Q Sepharose Fast Flow (GE Life sciences, EUA) para su purificación mediante cromatografía de intercambio iónico. La concentración de proteína se determinó por espectrofotometría a 280 nm usando un valor de $\epsilon_{280 \text{ nm}} = 75,290 \text{ M}^{-1} \cdot \text{cm}^{-1}$ para CMO, $\epsilon_{280 \text{ nm}} = 75,860 \text{ M}^{-1} \cdot \text{cm}^{-1}$ para GbcA y $\epsilon_{280 \text{ nm}} = 42,400 \text{ M}^{-1} \cdot \text{cm}^{-1}$ para GbcB. Estos valores son aproximaciones teóricas, calculadas a través del servidor ProtParam (<https://web.expasy.org/protparam/>).

V.3. Determinación del estado oligomérico de las enzimas recombinantes

El estado de asociación de las enzimas recombinantes se determinó por exclusión molecular (SEC) en una columna Superdex 200 HR-10/30 (GE Life sciences, EUA) acoplada a un equipo de cromatografía ÄKTA pure (GE Life sciences, EUA). La columna se equilibró y se eluyó a temperatura ambiente con un amortiguador (HEPES-KOH 50 mM, pH 7.5, KCl 50 mM, glicerol 10% (v/v)) en el caso de la CMO y un amortiguador (HEPES-KOH 50 mM, pH 7.0, KCl 50 mM, glicerol 10% (v/v)) en el caso de la enzima GbcA a una velocidad de flujo de 1.0 ml/min.

Previamente a la cromatografía, las soluciones de proteína se pasaron por un filtro de membrana hidrofílica de PVDF no estéril para jeringa con un poro de 0.22 μm de diámetro (Millipore, Alemania). La elución se monitoreó siguiendo la absorbancia a 280 nm. El estándar Gel Filtration Standard (Bio-Rad, EUA) utilizado para la calibración de la columna de filtración en gel contenía las proteínas: triglobulina de bovino (670,000 Da), γ -globulina de bovino (158,000 Da), ovoalbúmina de pollo (44,000 Da), mioglobina de caballo (17,000 Da) y vitamina B12 (1,350 Da).

V.4. Comprobación de la integridad del centro Rieske de las enzimas recombinantes

Debido al centro tipo Rieske, las oxigenasas Rieske/mononuclear tienen una absorbancia característica en la región de los 400 nm, esta absorbancia es menor en la enzima reducida que en la oxidada (Karlsson *et al.*, 2000). Esta propiedad se aprovechó para comprobar la integridad estructural de la enzima, es decir, que en la proteína recombinante se hubiera incorporado el centro Rieske. Para ello se redujo a la CMO (a una concentración de sitios activos 90 μM) con ditionita de sodio (5 mM) y se compararon los espectros de la enzima antes y después de la reducción con ditionita. A la enzima reducida se le agregó colina (30 mM) en una solución que contenía el oxígeno ambiental y se registró de nueva cuenta su espectro para comprobar que se había oxidado por el sustrato y, por tanto, se trataba de una enzima activa. El procedimiento de reducción de la enzima con ditionita se realizó a temperatura ambiente en condiciones de anaerobiosis obtenidas con atmósfera de nitrógeno y en un amortiguador (HEPES-KOH 50 mM, pH 7.5, KCl 50 mM, glicerol 10% (v/v)) burbujeado con nitrógeno durante 30 min previamente a su uso.

V.5. Ensayos de actividad enzimática de un único ciclo catalítico de la CMO

Con la proteína CMO reducida se realizaron ensayos de actividad de un único ciclo catalítico, utilizando un equipo de cinética rápida de flujo detenido (Applied Photophysics, UK). Estos ensayos se realizaron a 25 °C utilizando un baño termostático acoplado al equipo para controlar la temperatura. En el ensayo, la enzima reducida y bajo condiciones de anaerobiosis se cargó en una de las jeringas del equipo y en la segunda jeringa se cargó una solución 60 mM de colina; el equipo mezcló ambas soluciones y se registró el cambio en la absorbancia a 464 nm. El amortiguador en el que se disolvió la colina (HEPES-KOH 50 mM, pH 7.5, KCl 50 mM, glicerol 10% (v/v)) no se burbujeó con nitrógeno dado que el oxígeno es necesario para que ocurra la reacción enzimática. El mismo ensayo se repitió utilizando otros posibles sustratos de la CMO. Se siguió el curso temporal del aumento de absorbancia a 464 nm hasta que concluyó la reacción, es decir, hasta que todas las moléculas de enzima reducida ya habían sido oxidadas. Los datos obtenidos se ajustaron por regresión no lineal (mediante Microcal Origin 5.0) a una exponencial creciente sencilla.

$$A = A_{\infty} (1 - e^{-kt})$$

Ecuación 1

en donde A es la absorbancia a 460 nm al tiempo t , A_{∞} es la absorbancia a $t = \infty$ y corresponde a la amplitud del cambio de absorbancia cuando ha finalizado el primer ciclo catalítico y k es la constante de velocidad observada de orden uno (en unidades de s^{-1}).

V.6. Ensayos de actividad enzimática multiciclo de GbcA y GbcB

En primer lugar, se realizó una saturación de GbcB por NADH para conocer la concentración saturante de NADH que deberíamos usar en los ensayos de GbcA acoplados con GbcB. En la saturación de GbcB por NADH se utilizó el diclorofenol indolfenol (DCPIP) como aceptor final de los electrones que el NADH transfiere a GbcB. Este reactivo absorbe a 600 nm solo cuando está oxidado y se torna incoloro al reducirse, por lo que su reducción se puede seguir a una longitud de 600 nm. El ensayo se llevó a cabo en 50 mM de HEPES-KOH pH 6.9, 50 mM KCl, 10% (v/v) glicerol, 2 mM FAD, La concentración de proteína de GbcB fue 50 nM y de DCPIP 2 μ M.

La saturación de GbcA por su sustrato GB se determinó mediante un ensayo acoplado a GbcB, monitoreando el consumo de NADH mediante el cambio en la absorbancia a 340 nm, que es la longitud de onda del pico máximo de absorción de este compuesto. La reacción se llevó a cabo en 50 mM de HEPES-KOH, pH 6.9, 50 mM KCl, 10% (v/v) glicerol, 2 mM FAD y 0.1 mM NADH. En estos ensayos se utilizó una concentración de sitios activos de la enzima GbcA 2 μ M y 4 μ M de la enzima GbcB.

El valor de los parámetros cinéticos se estimó mediante regresión no lineal ajustando los datos de velocidad inicial a las ecuaciones de Michaelis-Menten (Ecuación 2) o de Hill (Ecuación 3). El análisis se hizo con los gráficos directo mediante las siguientes ecuaciones.

$$v_0 = V[S]/(K_m + [S]) \quad \text{Ecuación 2}$$

En la que v_0 es la velocidad inicial, V es la velocidad máxima, $[S]$ es la concentración del sustrato variable y K_m es la concentración del sustrato a la que se obtiene la mitad de la velocidad máxima.

$$v_0 = V[S]^h/(S_{0.5}^h + [S]^h) \quad \text{Ecuación 3}$$

En la que $S_{0.5}$ es la concentración del sustrato a la que se obtiene la mitad de la velocidad máxima, h es el coeficiente de Hill y los demás parámetros y variables como se definen para la Ecuación 2.

Se exploró la especificidad por el sustrato de GbcA monitoreando el consumo de NADH mediante el cambio en la absorbancia a 340 nm, que es la longitud de onda del pico máximo de absorción de este compuesto. La reacción se llevó a cabo en 50 mM de HEPES-KOH, pH 6.9, 50 mM KCl, 10% (v/v) glicerol, 2 mM FAD y 0.1 mM NADH. En estos ensayos se utilizó una concentración de sitios activos de la enzima GbcA de 2 μ M y 4 μ M de la enzima GbcB. Se utilizó una concentración fija de los posibles sustratos de 10 mM, los cuales fueron, colina, prolina betaína y betaína aldehído.

V.7. Comparación de secuencias de aminoácidos y análisis filogenético de las CMOs

Se recuperaron las secuencias de CMO de plantas disponibles en la base de datos del NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Benson *et al.*, 2013) y de Phytozome (<https://phytozome.jgi.doe.gov>) (Goodstein *et al.*, 2012) haciendo uso de la herramienta BLASTP (Boratin *et al.*, 2013). Se utilizó como semilla la secuencia de la CMO de espinaca (SoCMO) pero dado que en este trabajo encontramos que existe un duplicado, se utilizó la secuencia con número de acceso XP_021849509 que corresponde a la enzima con actividad colina monooxigenasa comprobada (Rathinasabapathi *et al.*, 1997). Se recuperaron 500 secuencias y se llevó a cabo un proceso de análisis y edición de las secuencias obtenidas con el programa BioEdit (Hall, 1999), asegurándonos de que todas las secuencias tuvieran un tamaño promedio (~400 aa) y de que tuvieran los dos dominios correspondientes a las colina monooxigenasa, esto último apoyado por la base de datos de dominios conservados del NCBI

(CDD; <http://www.ncbi.nlm.nih.gov/cdd>; Marchler-Bauer *et al.*, 2017). De las secuencias con un tamaño distinto al promedio, se recuperó la secuencia de nucleótidos y se revisó la predicción de los respectivos genes utilizando el programa Softberry FGENESH+ (<http://www.softberry.com/>) (Solovyev *et al.*, 2006) para corregir posibles errores introducidos por los sistemas de predicción automatizados. Después se realizó un alineamiento con el programa CLUSTALX (Larkin *et al.*, 2007) y los análisis filogenéticos se realizaron utilizando el programa MEGA7 (Kumar *et al.*, 2016). El árbol filogenético se obtuvo con 167 secuencias CMO utilizando el método de máxima verosimilitud (Whelan & Goldman, 2001). Se obtuvieron también árboles filogenéticos de los dos dominios por separado, es decir, del dominio Rieske y del dominio del Fe mononuclear no hémico, utilizando igualmente el método de máxima verosimilitud (Whelan & Goldman, 2001).

V.8. Generación de modelos tridimensionales por homología de las enzimas CMO y GbcA y acoplamiento molecular de posibles sustratos

Para la generación de los modelos por homología, se realizó en primer lugar un alineamiento de secuencias utilizando la herramienta MAFFT (Kato & Standley, 2013) para seleccionar el mejor templado. Los modelos se obtuvieron con el programa MODELLER v9.14 (Eswar *et al.*, 2008). Primero se obtuvieron los modelos de los centros metálicos como se ha hecho para otras enzimas (Mitra *et al.*, 2013) y luego se incorporaron al modelo. Para evaluar los modelos, se utilizó el protocolo Rd.HMM (Martínez-Castilla & Rodríguez-Sotres, 2010), un método que permite saber si el modelo generado realmente representa la manera en que la secuencia de aminoácidos se pliega. Este método consiste en eliminar las cadenas laterales de los residuos de aminoácidos del modelo generado dejando solo el esqueleto polipeptídico, tras lo cual mediante Rosetta design se le asignan aminoácidos al azar reconstruyendo nuevos modelos ahora con las

cadenas laterales de los residuos que predijo Rosetta, pero conservando intactas las coordenadas tridimensionales del esqueleto, es decir el patrón de plegamiento. De estos nuevos modelos se extraen las secuencias y mediante un modelo Markoviano oculto se busca en las bases de datos si secuencias de aminoácidos que respondan al patrón de frecuencias de los restos de aminoácido asignados por Rosetta, para cada posición, entre ellas, se espera encontrar la secuencia de la proteína que inicialmente se intentó modelar.

Para obtener el modelo por homología de GbcA se realizó un alineamiento de secuencias de la misma forma que en el caso de la CMO para seleccionar el mejor molde. Las simulaciones de acoplamiento molecular (“docking”) se realizaron con el programa AutoDock Vina (Trott & Olson, 2010) utilizando una caja de 10 \AA^3 centrada en el Fe mononuclear no hémico.

VI. RESULTADOS Y DISCUSIÓN

VI.1. Colina monooxigenasa (CMO)

VI.1.1. Aspectos evolutivos de las proteínas CMO

Todas las proteínas CMO son oxigenasas del tipo Rieske/mononuclear y, acorde a la base de datos de dominios conservados del NCBI, poseen un dominio Rieske, que corresponde al cd03541 de la subfamilia denominada Rieske_Ro_Alpha_N_CMO, y un dominio Fe mononuclear no hémico, que corresponde al cd08883 de la subfamilia denominada Rho_Alpha_C_CMO-like. Encontramos que las secuencias CMO están presentes en todas las plantas de genoma conocido y que todas ellas poseen los dos dominios que son específicos para las CMO. Además, se encontraron 10 secuencias CMO en cordados (3 en *Saccoglossus kowalevskii*, 3 en *Branchiostoma floridae* y 4 en *B. belcheri*) y 2 secuencias CMO en protistas (1 en *Nannochloropsis gaditana* y 1 en *Acanthamoeba castellanii*). El genoma de *A. castellanii* ha sufrido una extensiva transferencia horizontal de genes a través de su evolución (Clarke *et al.*, 2013) y esta pudiera ser la razón de la presencia de las secuencias CMO en los dos protistas. La transferencia horizontal de genes en cordados ya ha sido descrita (Sagane *et al.*, 2010) y pudiera ser también la explicación de que existan secuencias CMO en cordados. En ambos tipos de organismos, se desconoce la función de estas proteínas. Además, encontramos secuencias en bacterias y hongos que poseen un dominio del Fe mononuclear no hémico perteneciente al cd08883 que es el de las CMO de plantas, pero su dominio Rieske pertenece al cd03469 y no al cd03541 al que pertenece el dominio Rieske de las CMO. A estas proteínas bacterianas y de hongos se decidió llamarlas CMO-like y hasta el momento no se conoce la función fisiológica de ellas.

El árbol filogenético de las secuencias CMO recuperadas de las bases de datos y seleccionadas con el criterio de que poseían la combinación de dominios cd03541/cd08883, mostró que las CMO de angiospermas forman un grupo monofilético bien soportado (Fig. 10B). En forma interesante, el árbol filogenético mostró dos clados principales. Uno de ellos contiene secuencias de proteínas CMO solo de la familia Amaranthaceae, con la excepción de *Ophiopogon japonicus* que es una asparagácea y que también presenta una mayor identidad a nivel secuencia de su proteína betaína aldehído deshidrogenasa con los miembros de la familia Amaranthaceae (Muñoz-Clares *et al*, 2014). Este es un hallazgo interesante porque estas son las dos enzimas que en plantas catalizan la síntesis de GB y pareciera entonces que *O. japonicus* adquirió la ruta completa por transferencia horizontal a partir de alguna amarantácea. El otro clado contiene las secuencias de las proteínas CMO de todas las familias de plantas (incluida Amaranthaceae), lo que sugiere que las proteínas del primer clado probablemente surgieron de un proceso de duplicación génica que ocurrió en un ancestro común de las amarantáceas. Por esta razón, a las proteínas CMO que están presentes en todas las plantas decidimos llamarlas CMO1 y CMO2 a las que están presentes solo en las amarantáceas. La duplicación génica e incluso la duplicación completa de genomas en plantas es bastante común y después de un evento de duplicación génica se espera que la tasa de evolución de uno de los duplicados aumente, dada la disminución de la presión por evitar mutaciones que pudieran ser deletéreas cuando hay una sola copia de un gen (Scanell & Wolf, 2008). De acuerdo con esto, las CMO2 evolucionaron a una tasa mucho mayor que las CMO1, como se observa al comparar el grado de divergencia de este clado respecto al otro (Fig. 10B). La neo-funcionalización es un mecanismo por el cual uno de los duplicados adquiere una nueva función y si esta nueva función le confiere mejores propiedades adaptativas a la especie, el proceso de selección natural permite que se mantengan ambos

duplicados (Flagel & Wendel, 2009). En el caso de las amarantáceas, que poseen CMO1 y CMO2, la adquisición de la función de colina monooxigenasa por la CMO2 probablemente les haya conferido a las especies de esta familia mejores capacidades adaptativas a su hábitat, que predominantemente consiste en zonas áridas o semiáridas con suelos salinos (Kadereit *et al.*, 2004). A pesar de que se encontraron algunas secuencias CMO duplicadas en otras familias de plantas diferentes a Amaranthaceae, estos duplicados se agruparon en el mismo clado dejando claro que son recientes y no han divergido de las CMO1 originales. Siguen por tanto siendo proteínas CMO1. Las proteínas CMO1 evolutivamente están más cercanas a las bacterianas y probablemente derivaron de éstas, por lo que pensamos que la función de las proteínas CMO1, aún desconocida, probablemente sea muy parecida a la del ancestro bacteriano, también desconocida. De hecho, intentando encontrar alguna pista sobre la posible actividad fisiológica de las isozimas CMO1, se analizó el contexto genómico de los microorganismos *Gemmata obscuriglobus* y *Pseudoxanthomonas* sp. CF125 (Fig. 11), que son los dos que ya tienen su genoma secuenciado y cuentan con una enzima CMO-like. Sin embargo, los genes contiguos al gen *cmo-like* no tienen función conocida y las proteínas que codifican dichos genes tienen dominios conservados con diversas posibles funciones asociadas. En el caso de *G. obscuriglobus* ninguna de estas funciones está relacionada con el metabolismo de compuestos como la colina o algún compuesto aromático. En el caso de *Pseudoxanthomonas* hay dos funciones interesantes, una asociada con el metabolismo de compuestos aromáticos xenobióticos y la otra con la del ácido fenilacético, pero dado que ninguna de estas proteínas ha sido caracterizada solo podemos hipotetizar que sí podrían estar relacionadas con la degradación de compuestos aromáticos.

Para investigar más a fondo la historia evolutiva de las CMO, realizamos análisis filogenéticos de los dos dominios por separado, lo que nos permitió determinar que en las proteínas CMO2 el

dominio del Fe mononuclear no hémico, donde se encuentra el sitio activo y, por lo tanto, donde se determina la especificidad por el sustrato, evolucionó a una tasa mucho mayor que el dominio donde se encuentra el centro Rieske [2Fe-2S] (Fig. 12). Nuestra propuesta es que las enzimas CMO2 han estado sujetas a una presión evolutiva, probablemente derivada de su hábitat, que promovió la adquisición de la actividad colina monooxigenasa, es decir la adquisición de la capacidad de usar a colina como sustrato y así participar en la biosíntesis de GB; creemos que la especificidad por el sustrato de las enzimas CMO1 es probablemente diferente a las de las enzimas CMO2. Pudiera ser que las CMO2 conserven al menos como una actividad vestigial la capacidad de usar el sustrato fisiológico que usen las CMO1, que por el momento desconocemos cual es, y/o que las CMO1 posean alguna capacidad de usar colina, a partir de la cual evolucionó la especificidad de las CMO2.

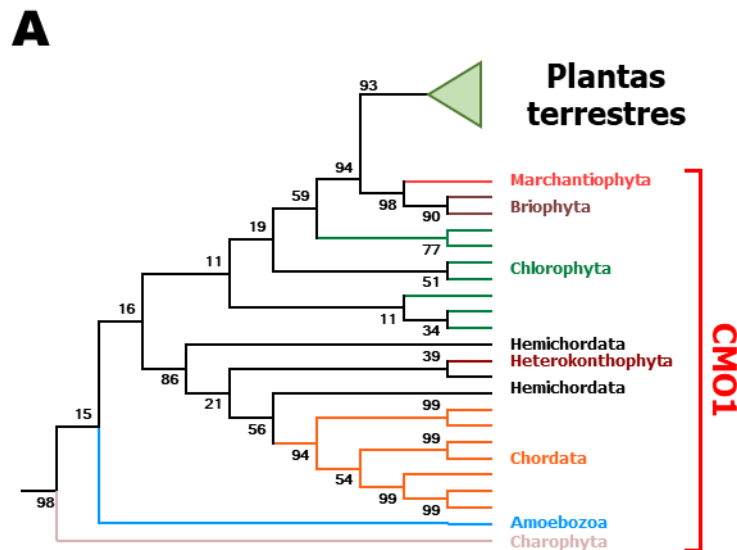
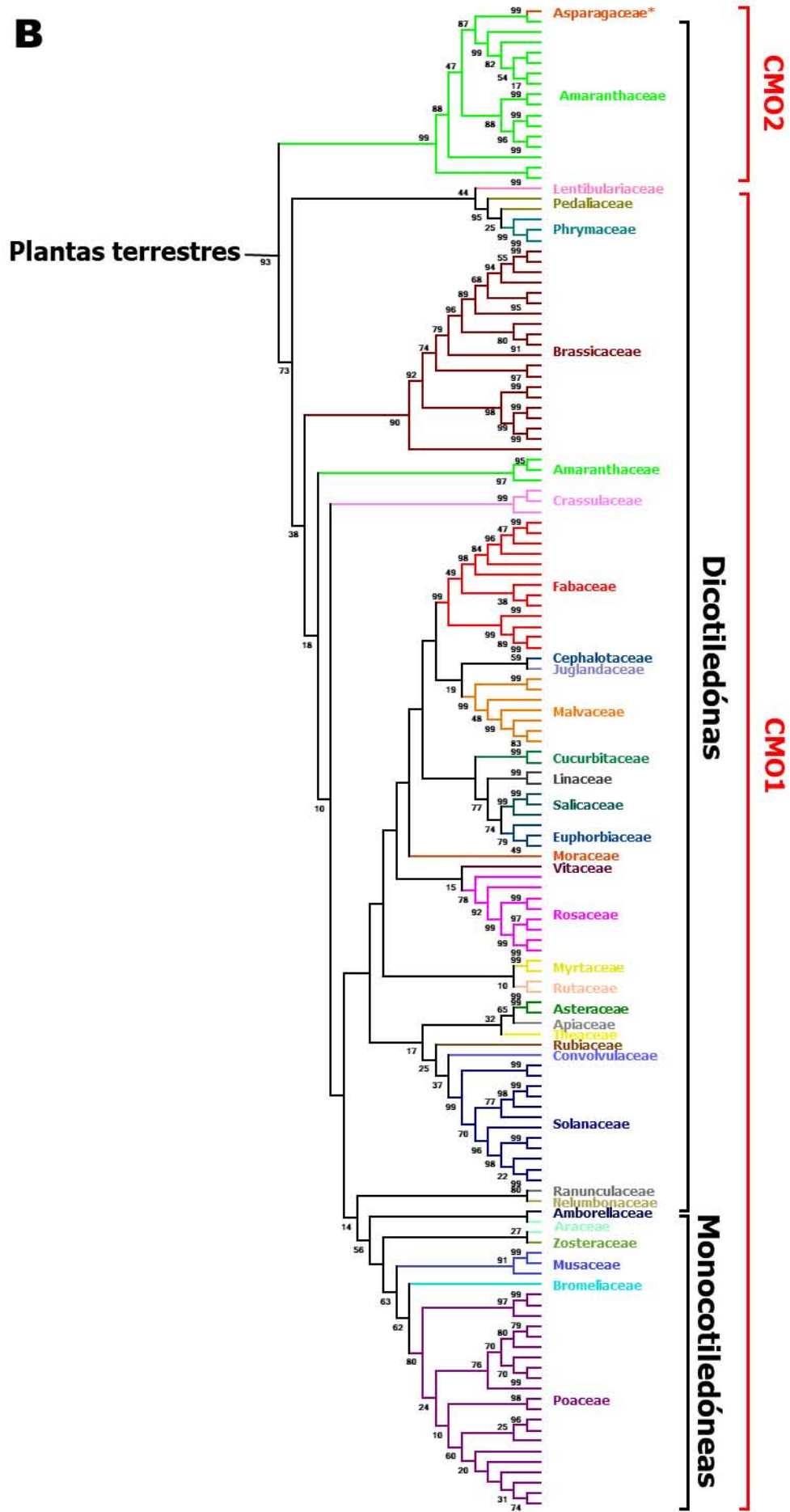
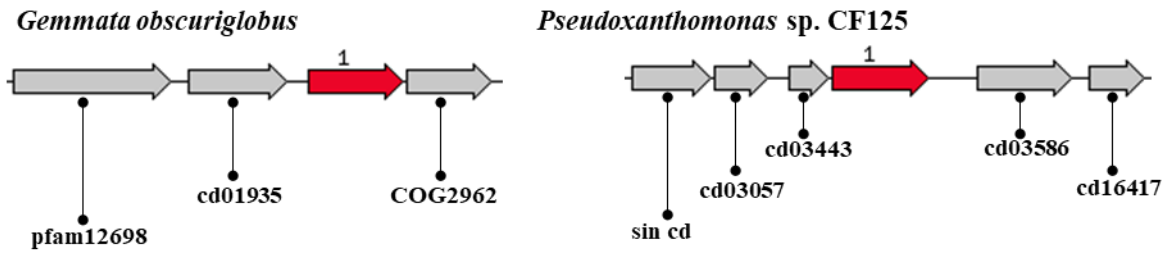


Figura 10. Árbol filogenético de las CMO. El árbol incluye las 181 secuencias de plantas (**B**) y secuencias de otros organismos eucariontes directamente relacionadas con las CMO de plantas (**A**). Las familias de plantas se encuentran señaladas y separadas por color. El árbol es el resultado de 500 réplicas, se muestran los valores de soporte en cada rama.

B



A**B**

pfam12698	Familia de transportadores tipo ABC
cd01935	Proteínas parecidas a la colilglicina hidrolasa
COG2962	Proteínas denominadas RarD de función desconocida
cd03057	Función asociada a la detoxificación a través de la conjugación de glutatión con agentes xenobióticos.
cd03443	Función asociada a la degradación de ácido fenilacético
cd03586	DNA polimerasa IV/Kappa
cd16417	Fosfoglicolato fosfatasa

Figura 11. Contexto genómico de microorganismos con proteínas CMO-like. (A), Contexto genómico de los microorganismos *G. obscuriglobus* y *Pseudoxanthomonas*. En ambos casos se marca en rojo el gen *cmo-like*, (B). Función predicha de los genes contiguos en base al dominio conservado (cd), pfam o COG (B).

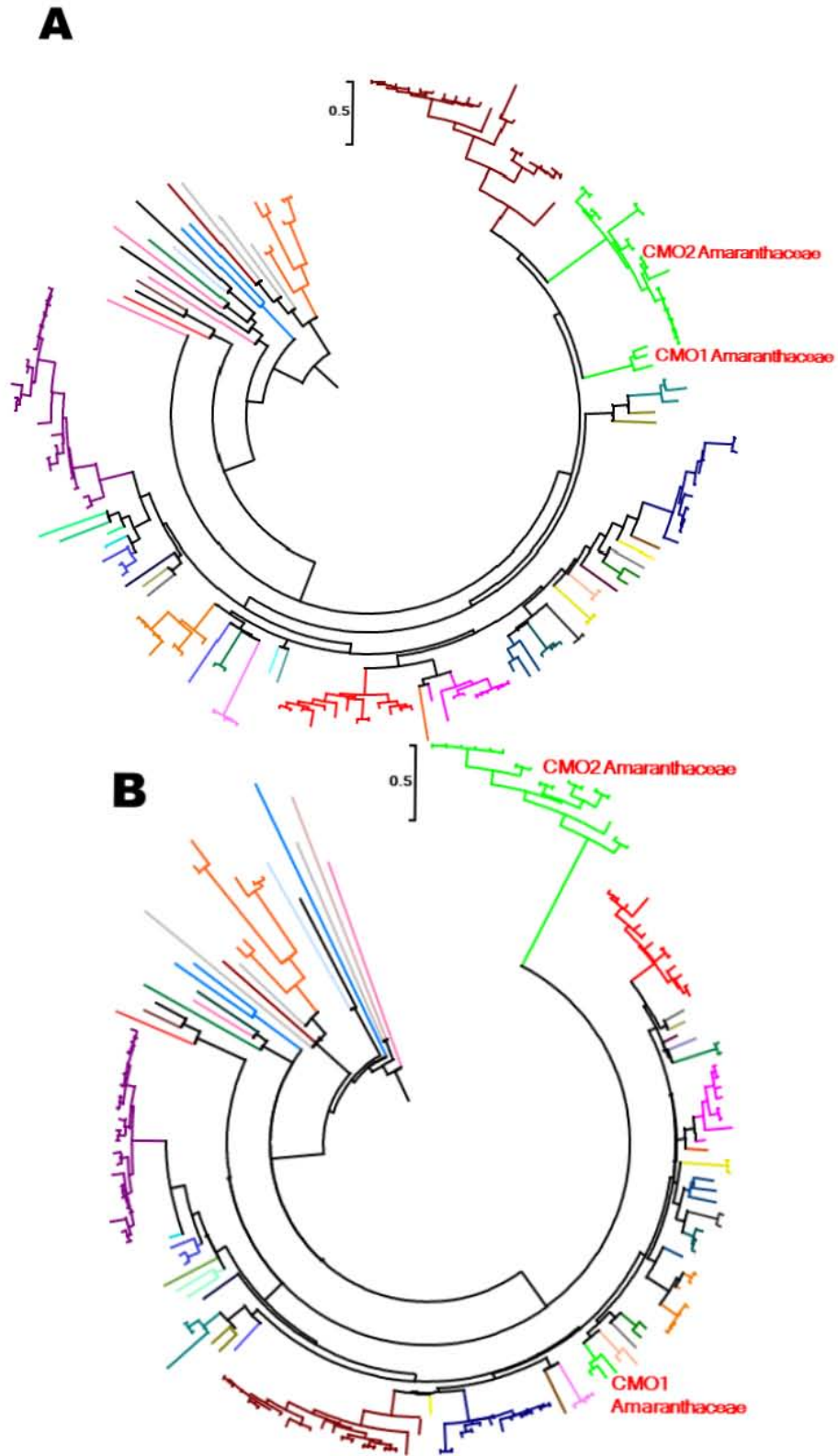


Figura 12. Árboles filogenéticos de los dos dominios de las CMO. Las familias de plantas se encuentran marcadas por color. **A.** Análisis filogenético realizado solo con el dominio Rieske. **B.** Análisis filogenético realizado solo con el dominio del Fe mononuclear no héptico.

VI.1.2. Obtención de la enzima *SoCMO2* recombinante y determinación del estado oligomérico

Después de la purificación mediante cromatografía de intercambio iónico de la proteína recombinante *SoCMO2* madura sin etiqueta de histidinas, se obtuvieron en promedio de 10-15 mg de proteína por litro de cultivo, el proceso de purificación se analizó mediante SDS-PAGE (Fig. 13). Dado que la enzima no tiene una etiqueta de histidinas se observaron algunas impurezas en las fracciones con mayor contenido de proteína, sin embargo, la enzima *SoCMO2* está presente en mayor cantidad que el resto de las proteínas.

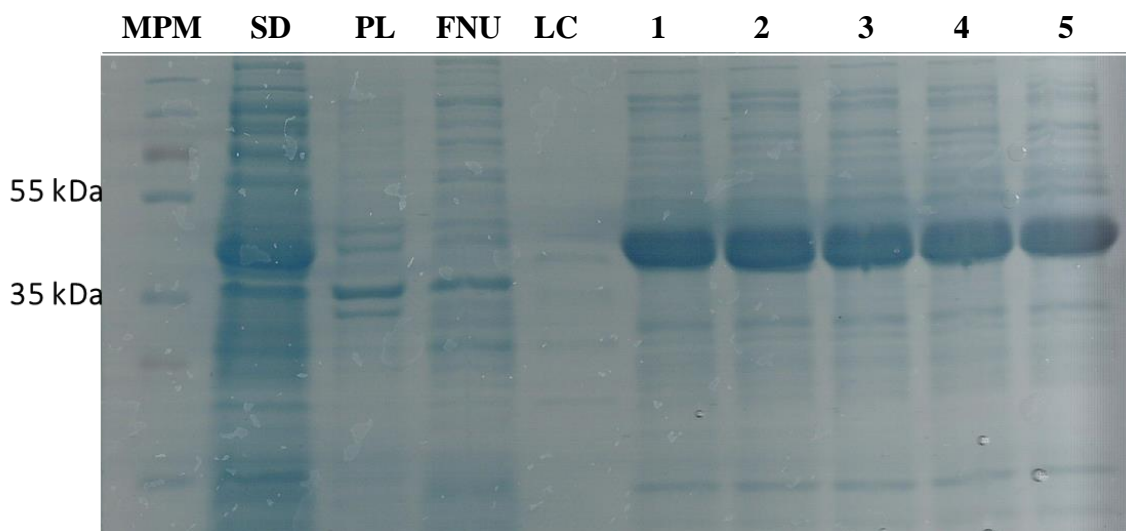


Figura 13. Purificación de *SoCMO2* madura sin etiqueta de histidinas por cromatografía de intercambio iónico. MPM, marcador de peso molecular (Precision Plus Protein™ Dual Color Standards, BioRad, EUA); SD, sobrenadante del lisado celular; PL, pastilla del lisado celular; FNU, fracción no unida a la columna; LC, lavado de columna; 1-5, fracciones con mayor contenido de proteína.

La cromatografía de exclusión molecular de la enzima recombinante *SoCMO2* pura mostró dos picos (Fig. 14), el pico con mayor área tuvo un tiempo de retención de 14.13 min. De acuerdo con la curva de calibración utilizada (Fig. 14, inserto), este pico tiene una masa molecular de 119

kDa, lo que con base a la masa molecular estimada del monómero de la *SoCMO2* (41 kDa) puede corresponder a un trímero como se ha descrito para otras oxigenasas Rieske /mononuclear.

CMO sin His

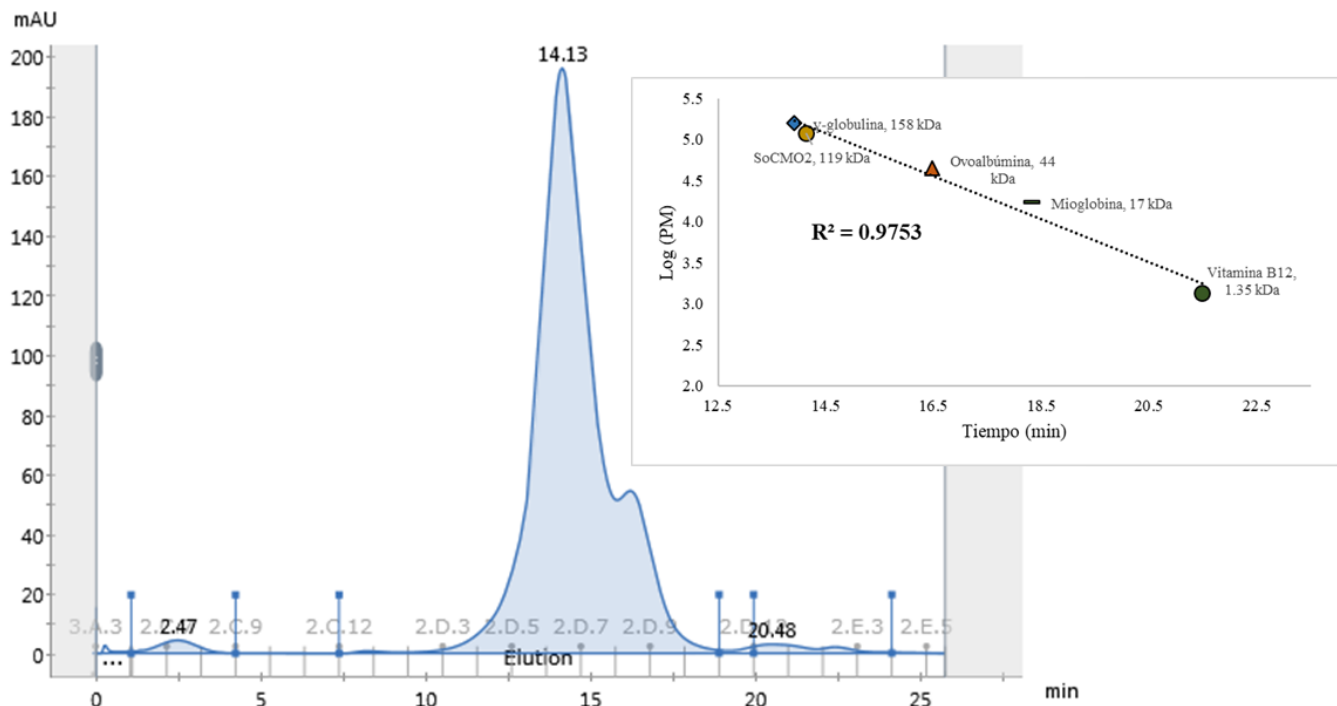


Figura 14. Perfil de elución de la enzima *SoCMO2* de la cromatografía de exclusión molecular. Inserto: curva de calibración de los estándares de masa molecular. La cantidad de proteína total aplicada fue de 3 mg. La absorbancia a 280 nm está dada en miliunidades de absorbancia (mAU)

VI.1.3. Comprobación de integridad de los centros Rieske de la *SoCMO2* madura sin etiqueta de histidinas

Los espectros de la *SoCMO2* oxidada y reducida se muestran en la Figura 15. Se puede apreciar un cambio en la absorbancia en la región de los 400 nm, cambio que fue parcialmente revertido por la adición de colina lo cual es consistente que la enzima es activa con colina como sustrato.

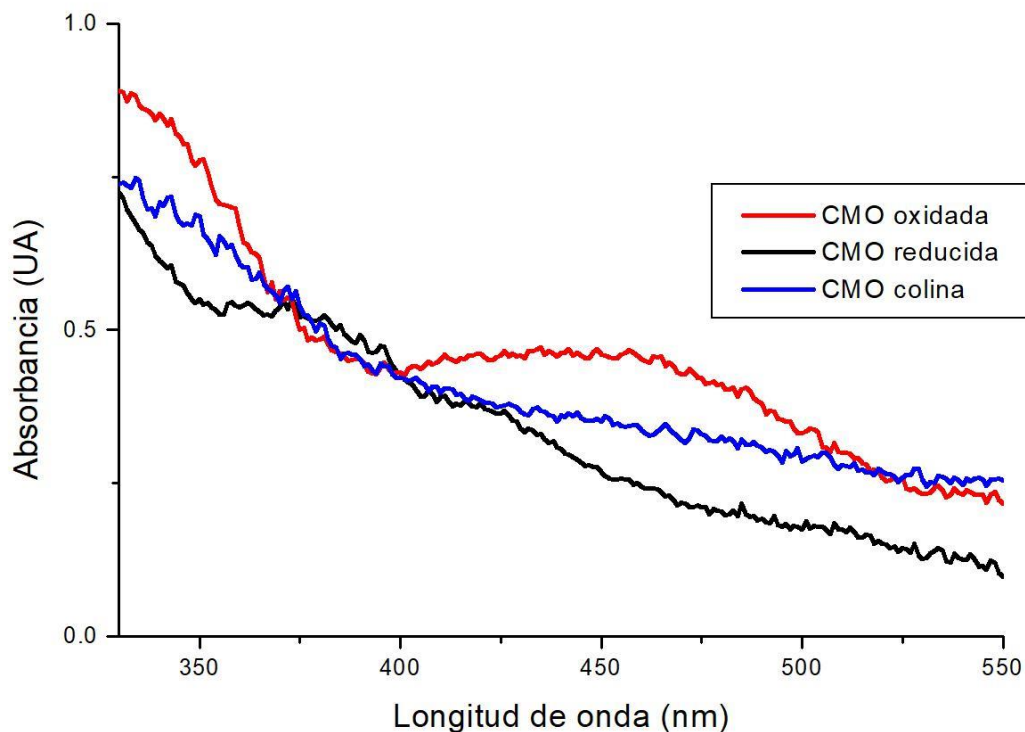


Figura 15. Espectro UV-visible de la SoCMO2. En rojo la enzima oxidada, negro la enzima reducida, azul la enzima reducida tras adicionar colina.

VI.1.4. Cinética de un único ciclo catalítico de la SoCMO2

Con la proteína reducida con ditionita de sodio se realizaron experimentos de un único ciclo catalítico para lo cual, como se describió en la sección de Métodos, la solución de la enzima se combinó en el equipo de cinética rápida con la solución que contenía el posible sustrato y oxígeno ambiental y se siguió el incremento en la absorbancia a 464 nm resultante de la oxidación del centro Rieske (Fig. 16A). Anteriormente, la actividad de colina monooxigenasa de la CMO2 purificada de hojas de espinaca se había comprobado mediante un ensayo enzimático acoplado con ferredoxina, ferredoxina reductasa y NADPH, también de espinaca (Brouquisse *et al.*, 1989; Burnet *et al.*, 1995). En el ensayo se usó como sustrato colina marcada con C14 y se determinó la aparición de GB marcado agregando al ensayo betaína aldehído deshidrogenasa. De

esta forma se pudo comprobar la oxidación de colina. El ensayo de un único ciclo catalítico ya ha sido utilizado con oxigenasas bacterianas Rieske/mononuclear (Wolfe *et al.*, 2001 y 2002) y presenta la ventaja de no depender de la presencia de ferredoxina y ferredoxina reductasa, ya que los electrones los provee la ditionita. Este ensayo permite también conocer si la enzima requiere formar o no complejo con la ferredoxina para ser activa. Simplifica además la búsqueda de otros posibles sustratos ya que lo que se observa es la oxidación de la enzima y no la aparición de algún producto, que habría que identificar posteriormente, o la desaparición del sustrato, lo que metodológicamente también es más complicado porque son sustratos que no poseen color y tampoco fluorescencia.

Una vez comprobada la actividad colina monooxigenasa utilizando el ensayo de un único ciclo catalítico, decidimos probar la capacidad de la SoCMO2 de usar a otros compuestos como sustrato y oxidarlos. Dado que la especificidad por el sustrato de la colina monooxigenasa no había sido estudiada hasta el momento, se tomó en cuenta para la selección de probables sustratos que la mayoría de los sustratos de las oxigenasas Rieske/mononuclear son compuestos aromáticos y que solo se ha reportado otro compuesto con un grupo trimetilamonio en su estructura, la prolina betaína, que es oxidado por este tipo de enzimas. Por ello, probamos la prolina betaína (estaquidrina) compuesto que como ya se mencionó es sustrato de la prolina betaína monooxigenasa, enzima Rieske/mononuclear del microorganismo *S. meliloti* que puede utilizar este compuesto como única fuente de carbono (Burnet *et al.*, 2000; Daughtry *et al.*, 2012). Se probó también el ácido salicílico, una hormona de plantas que es un compuesto aromático estructuralmente parecido a los sustratos descritos de muchas oxigenasas Rieske/mononuclear bacterianas. Este compuesto es metabolizado por algunos microorganismos como *Ralstonia* sp., mediante la enzima tipo Rieske/mononuclear salicilato 5-hidroxilasa que

produce gentisato (Fang & Zhou, 2014). En plantas el ácido salicílico es también transformado en gentisato, pero la enzima que cataliza esta reacción aún no ha sido descrita. La SoCMO2 es capaz de oxidar ambos compuestos (Figs. 16B y C), indicando que por lo menos *in vitro* la enzima puede oxidar compuestos distintos a la colina. Una baja especificidad por el sustrato se ha observado en la enzima Rieske/mononuclear naftaleno dioxigenasa de *Pseudomonas* sp., que puede llevar a cabo reacciones de monooxigenación, desaturación, sulfoxidación y desalquilación utilizando sustratos estructuralmente diferentes al naftaleno (Resnick *et al.*, 1996). Este pudiera ser el caso de la SoCMO2, por lo que pensamos que es necesario el análisis de un mayor número de compuestos. Sin embargo, dado que el ensayo de un único ciclo catalítico requiere concentraciones considerables de enzima y de sustrato, se podría usar otro método para identificar aquellos compuestos que se unen a la enzima, como el corrimiento térmico (“thermal-shift”), y una vez identificados analizar su probable oxidación por ensayos de cinética de un solo ciclo.

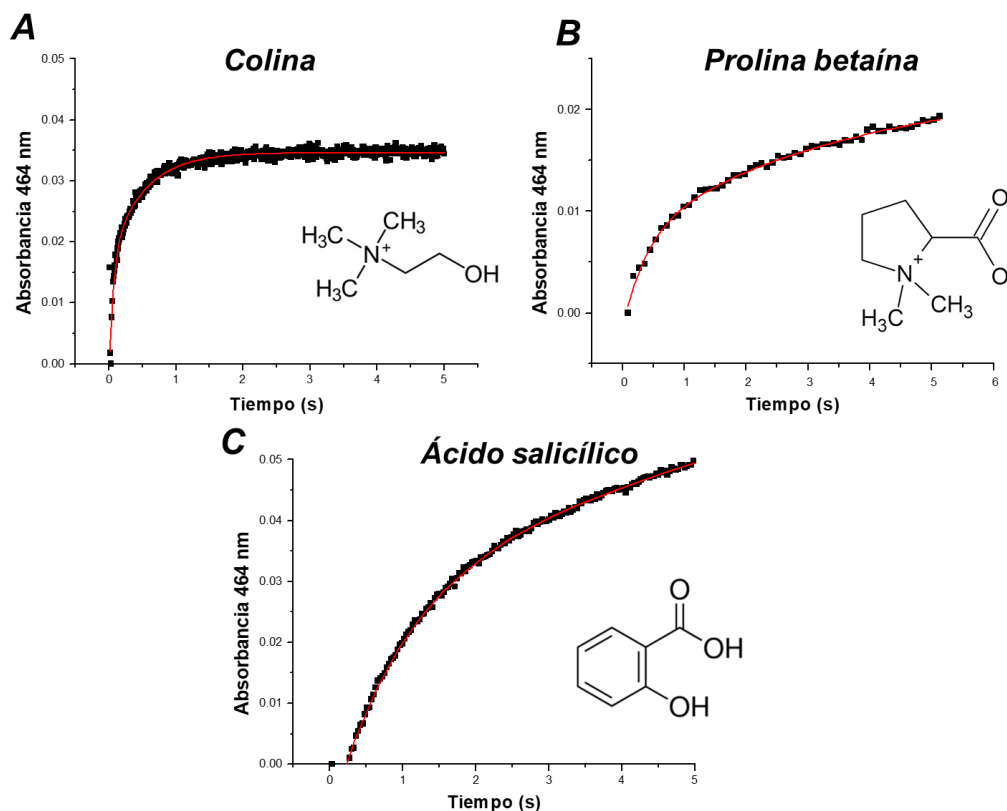


Figura 16. Cinética de un solo ciclo catalítico de SoCMO2. La enzima (30 μ M de sitios activos) se redujo con ditionita de sodio (5 mM) y se mezcló en un equipo de cinética rápida con distintos compuestos, todos a una concentración de 30 mM. Las reacciones se llevaron a cabo a 25 $^{\circ}$ C y pH 8.0. La línea roja sobrepuesta sobre los datos experimentales (en negro) es el resultado del ajuste a la Eq. 1. En las gráficas, los cursos temporales mostrados son representativos de los obtenidos en 3 repeticiones.

VI.1.5. Modelos por homología de las dos isozimas CMO de espinaca (SoCMO1 y SoCMO2)

Uno de los objetivos del presente trabajo de tesis era encontrar características estructurales que permitieran identificar a posibles sustratos y/o justificar la especificidad por los sustratos que hubiéramos encontrado en los ensayos de actividad. Para ello, intentamos obtener la estructura tridimensional por medio de cristalografía y difracción por Rayos X. Pero debido a que después de numerosos intentos no logramos obtener cristales que difractaran rayos X con resolución suficiente, decidimos investigar sus características estructurales, tanto de la SoCMO2 como de la

SoCMO1, de la que no existe hasta el momento ninguna caracterización bioquímica ni estructural, generando modelos por homología, como se describió en la Sección de Métodos.

El mejor molde para obtener el modelo por homología de la *SoCMO2* resultó ser una oxigenasa Rieske/mononuclear bacteriana de *Silicibacter* sp (PDB: 3N0Q) que tiene un 27% de identidad con la secuencia de la CMO XP_021849509 de espinaca, que fue la primera enzima modelada (Fig. 17A). El modelo de la *SoCMO2* (XP_021849509) se obtuvo usando la secuencia de aminoácidos que no incluye al péptido señal para cloroplasto, el cual se había predicho que está formado por los primeros 60 residuos del amino terminal (Rathinasabapathi *et al.*, 1997). Además, eliminamos los 19 siguientes residuos ya que se predecían con un alto desorden usando el programa descrito por Dosztány (2018). El modelo de la *SoCMO1* (XP_021866412) se obtuvo sin los primeros 102 residuos del amino terminal que de igual forma se predecían que eran altamente desordenados (Fig. 17B). Acorde al protocolo Rd.HMM (Martínez-Castilla & Rodríguez-Sotres, 2010), la secuencia de aminoácidos de las reconstrucciones libres generadas por Rosetta design corresponde a la secuencia objetivo y ya no a la secuencia de la proteína que sirvió como molde, tanto en el caso de la *SoCMO2* como en el de la *SoCMO1* (Tabla 2). El Rd.HMM index de los modelos se acerca bastante al 0.6 que es el que se obtiene al analizar estructuras de proteínas determinadas por cristalografía de rayos-X, lo cual indica la confiabilidad del modelo obtenido. Los valores de RMSD (root median standard deviation) incluidos en la tabla se calcularon con respecto al molde original.

Tabla 2. Evaluación de los modelos por homología de SoCMO1 y SoCMO2 con el protocolo Rd.HMM

Modelo/ Subunidad	Puntaje de búsqueda HMM	Primer resultado en función del alineamiento					Rd-HMM index (puntaje/cobertura)	RMSD (Å ²)
		ID	AA de inicio	AA final	Gaps	Marco corrido		
SoCMO2/A	204.7	XP_021849509	87	439	0	no	0.582	2.64
SoCMO2/B	156.4	XP_021849509	95	428	0	no	0.470	2.28
SoCMO2/C	193.7	XP_021849509	92	422	0	no	0.587	2.36
SoCMO1/A	105.1	XP_021866412	106	410	0	no	0.349	1.96
SoCMO1/B	137.5	XP_021866412	111	421	0	no	0.444	1.84
SoCMO1/C	148.1	XP_021866412	105	423	0	no	0.466	1.97

En la Figura 17C se muestra el modelo por homología del trímero de la SoCMO2, así como un acercamiento a los centros metálicos coordinados por los residuos correspondientes y el residuo de aspártico que se propone que participa en la transferencia de electrones. Es interesante que en el sitio activo de la SoCMO2 se encontraron tres residuos aromáticos (Tyr281, Tyr295, Phe301) y un residuo de glutámico (Glu346) (Fig. 18A) que pudieran participar en la unión del grupo trimetilamonio de la colina como se ha encontrado en otras proteínas (receptores, transportadores y enzimas) capaces de unir a compuestos con grupos de amonio cuaternario (Oswald *et al.*, 2008; Haga *et al.*, 2012). En el acoplamiento molecular de la colina en el sitio activo, se inspeccionaron 60 posiciones de la molécula y se seleccionó aquella en la que el hidroxilo de la colina estuviera orientado hacia el Fe mononuclear (Fig. 18A). El resultado obtenido indica que la SoCMO2 sí puede unir la colina en este sitio interaccionando en la forma esperada con los residuos aromáticos y el residuo glutámico. Los primeros forman una especie de caja que rodea al grupo trimetilamonio con el que pudiera hacer interacciones catión-pi mientras que el grupo carboxilo de la cadena lateral del glutámico neutralizaría la carga positiva del trimetilamonio. El grupo hidroxilo de la colina quedó orientado hacia el Fe mononuclear no hémico, lo que es necesario para que ocurra la reacción de hidroxilación sobre el carbono cinco. Un hallazgo interesante fue el que los residuos que con base en estos resultados proponemos que están involucrados en la unión de la colina a la SoCMO2 están totalmente conservados en las CMO2, pero no lo están en

ninguna de las proteínas CMO1, incluyendo a la *So*CMO1 (Fig. 18C). En las proteínas CMO1 los residuos aromáticos Tyr295 y Tyr301 (numeración de *So*CMO2) cambian a leucinas y el Glu346 cambia a una asparagina (Fig. 18B). Dado el cambio en el tamaño de la cadena lateral en el caso de las leucinas es probable que el sitio de unión del sustrato de las proteínas CMO1 sea más grande y que pueda utilizar sustratos más voluminosos. La Tyr281 se conserva y pudiera ser que su papel esté relacionado con la catálisis más que con la especificidad por el sustrato.

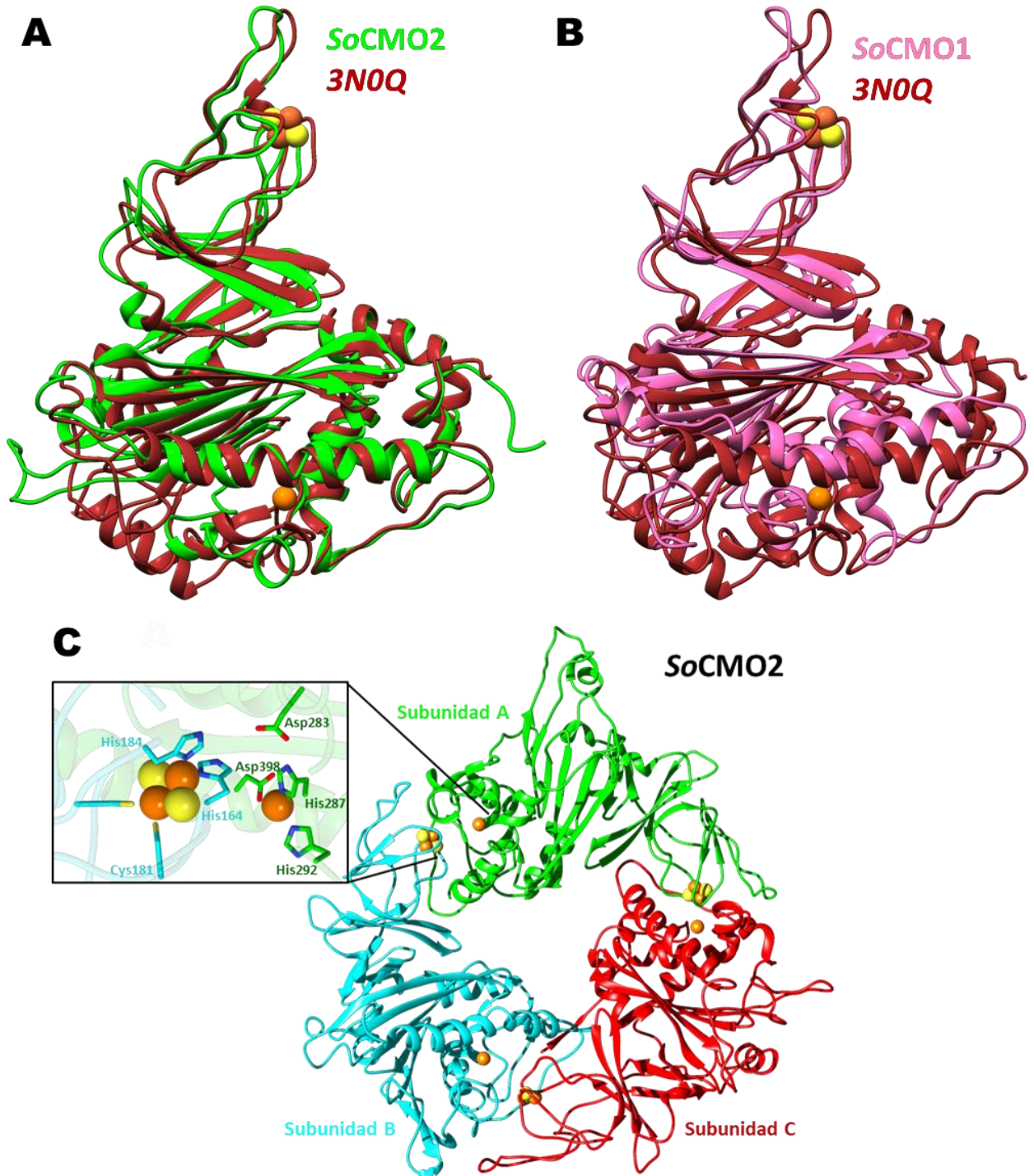


Figura 17. Modelo por homología de las SoCMO. (A) Representación esquemática del alineamiento estructural entre el modelo de la SoCMO2 y el molde PDB: 3NOQ. (B) Representación esquemática del alineamiento estructural entre el modelo de la SoCMO1 y el molde PDB: 3NOQ. (C) Representación en caricatura del modelo tridimensional de la CMO2 de espinaca. En el acercamiento, se muestran los residuos involucrados en la coordinación del Fe mononuclear y los residuos involucrados en la unión de los dos Fe del centro Rieske. Se muestran en representación de esferas los átomos de azufre (amarillo) y hierro (naranja). La numeración de los residuos corresponde a la SoCMO2 madura.

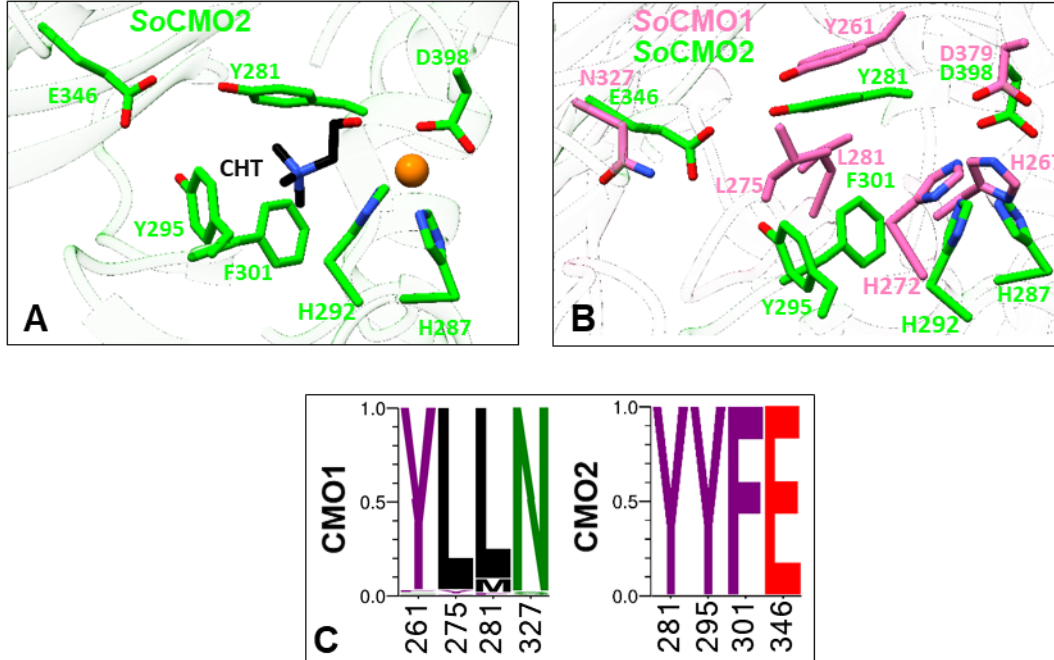


Figura 18. Predicción del sitio activo de *SoCMO* en los modelos por homología. **A.** Representación en caricatura donde se muestran los potenciales residuos de unión al sustrato de la *SoCMO2*. El Fe mononuclear se muestra en color naranja y la molécula de colina con los carbonos en color negro. **B.** Comparación de sitios activos entre *SoCMO2* (verde) y *SoCMO1* (rosa). **C.** Conservación en las enzimas CMO1 y CMO2 de los residuos del sitio activo posiblemente involucrados en la unión al sustrato. La numeración corresponde a la *SoCMO1* y *SoCMO2*, respectivamente.

VI.2. Glicina betaína monooxigenasa (GbcA) y glicina betaína monooxigenasa reductasa (GbcB) de *P. aeruginosa*

Al inicio de esta tesis, no se había caracterizado ninguna monooxigenasa de glicina betaína, pero recientemente apareció un reporte de la caracterización de la glicina betaína monooxigenasa de la gamma-proteobacteria *Chromohalobacter selxigens* (llamada BmoA) (Shao *et al.*, 2018). Esta enzima tiene un 66% de identidad a nivel secuencia de aminoácidos con GbcA. A pesar de ello, consideramos de interés continuar con el estudio de GbcA para aportar nuevos resultados sobre el catabolismo de colina y de GB en *P. aeruginosa*, que es de gran relevancia en la fisiopatología de este importante patógeno oportunista humano (Wargo *et al.*, 2008; Wargo 2013).

VI.2.1. Obtención de las proteínas recombinantes GbcA y GbcB y determinación del estado oligomérico de las enzimas.

De la purificación de GbcA se obtuvieron 20 mg por litro de cultivo de células transformadas y de la de GbcB se obtuvieron 15 mg por litro de cultivo. Los resultados de la purificación de GbcA y GbcB se analizaron mediante SDS-PAGE (Figs. 19 y 20)

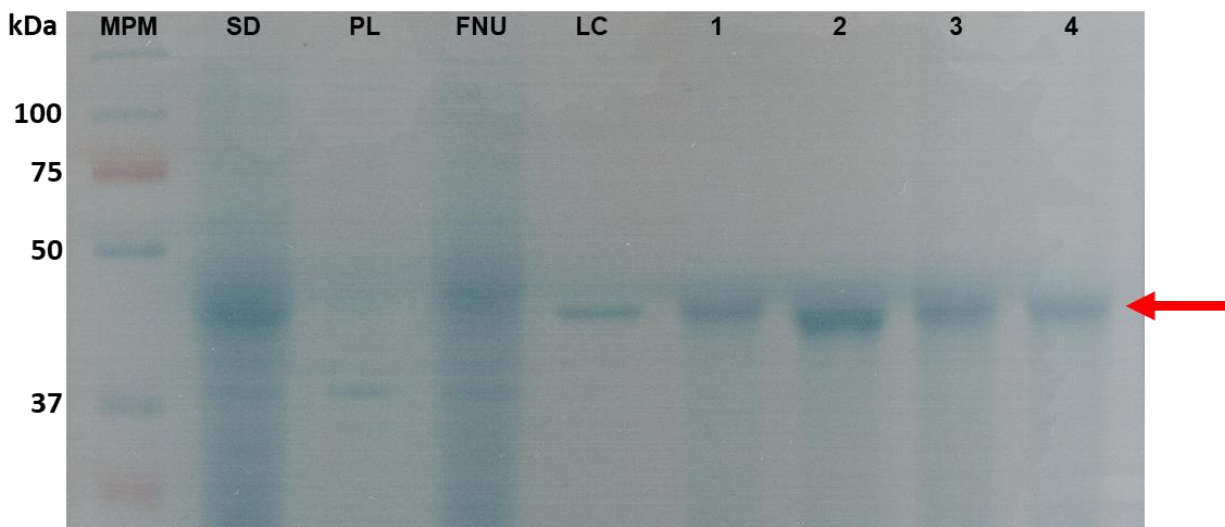


Figura 19. Purificación de GbcA. MPM, marcador de peso molecular (Precision Plus Protein™ Dual Color Standards, BioRad, EUA); SD, sobrenadante del lisado celular; PL, pellet del lisado celular; FNU, fracción no unida a la columna Protino; LC, lavado de columna; 1-4, fracciones con mayor contenido de proteína.

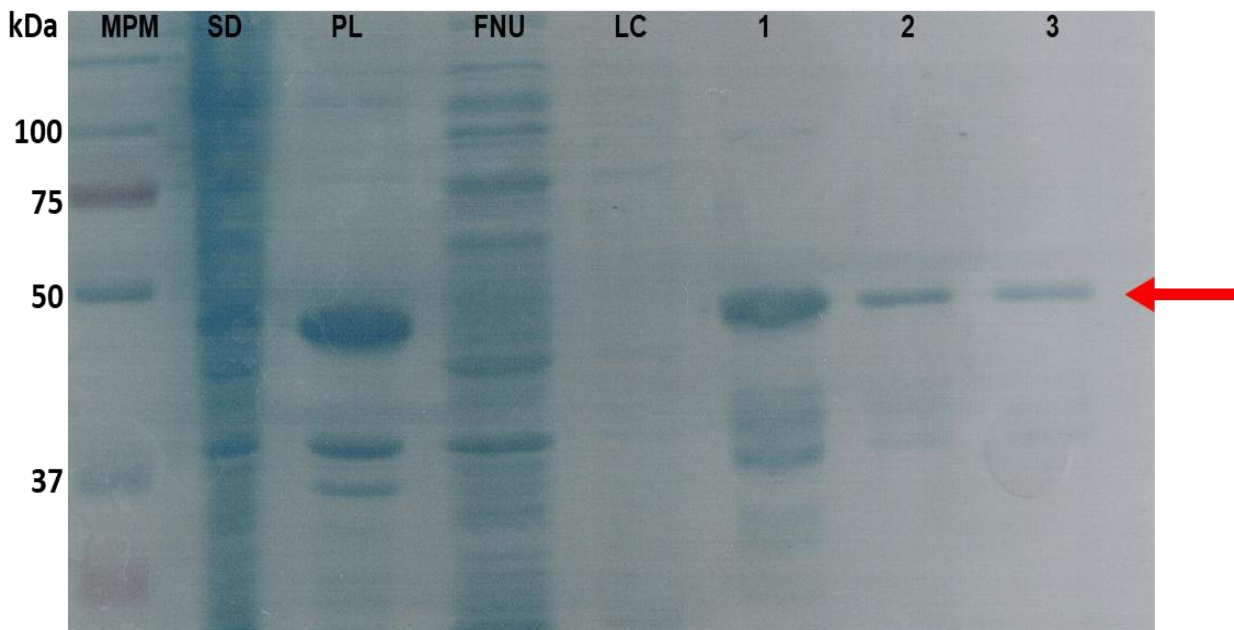


Figura 20. Purificación de GbcB. MPM, marcador de peso molecular (Precision Plus Protein™ Dual Color Standards, BioRad, EUA); SD, sobrenadante del lisado celular; PL, pellet del lisado celular; FNU, fracción no unida a la columna Protino; LC, lavado de columna; 1-4, fracciones con mayor contenido de proteína.

La cromatografía de exclusión molecular de la enzima recombinante GbcA pura mostró dos picos (Fig. 21), el pico con mayor área tuvo un tiempo de retención de 13.42 min. De acuerdo con la curva de calibración utilizada (Fig. 21, inserto), este pico tiene una masa molecular de 114 kDa, lo que con base a la masa molecular estimada del monómero de GbcA (49 kDa) puede corresponder a un trímero, BmoA, la glicina betaína monooxigenasa de *C. salexigens* que es la enzima homóloga de GbcA se describió como un trímero, para la exclusión molecular de BmoA se utilizó un amortiguador (50 mM Tris-HCl pH 7.0, NaCl 150 mM) diferente al de GbcA (HEPES-KOH 50 mM, pH 7.0, KCl 50 mM, glicerol 10% (v/v)).

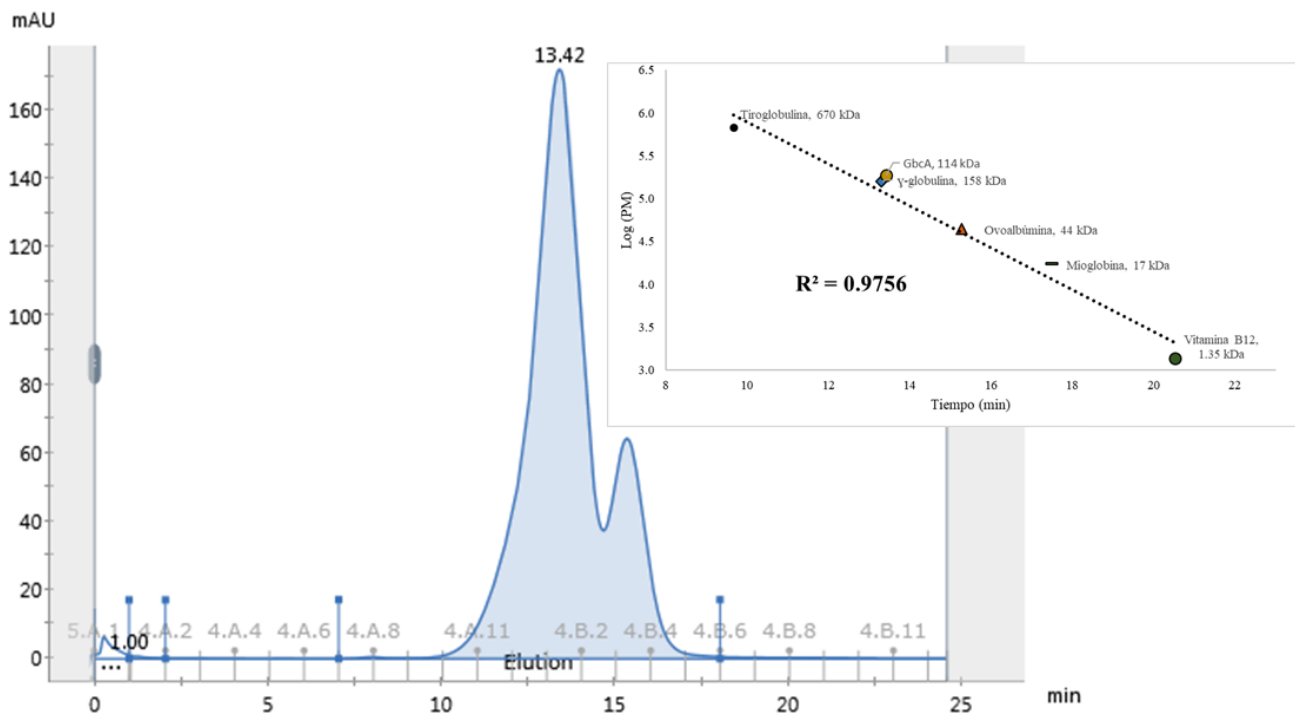


Figura 21. Perfil de elución de la enzima GbcA de la cromatografía de exclusión molecular. Inserto: curva de calibración de los estándares de masa molecular. La cantidad de proteína total aplicada fue de 3 mg. La absorbancia a 280 nm está dada en miliunidades de absorbancia (mAU),

VI.2.2. Caracterización cinética de la enzima recombinante GbcB

La cinética de saturación de GbcB por NADH nos permitió conocer la concentración de NADH saturante necesaria para que la actividad de GbcB no sea limitante en el ensayo de GbcA acoplado con GbcB. En este trabajo se decidió utilizar el método acoplado al DCPIP ya que este permite ensayos en un intervalo amplio de concentraciones de NADH, pudiéndose usar concentraciones de NADH superiores a las que se podrían usar en el método en el que se siguen los cambios de absorbancia consecuencia del consumo de NADH. La cinética de saturación se realizó como se describió en la sección de Métodos y se muestra en la Figura 22. Los parámetros cinéticos estimados a partir de los datos de velocidad inicial se incluyen en la Tabla 3. De acuerdo a la caracterización cinética realizada en este trabajo, la K_m aparente de GbcB por NADH es de 20 μM , un valor inferior al reportado (40 μM) para la enzima BmoB (Shao *et al.*,

2018), que es la enzima homóloga de GbcB del microorganismo *C. salexigens*. Estas diferencias pudieran deberse a que en el ensayo enzimático utilizado para la caracterización cinética de BmoB, a diferencia del que nosotros utilizamos para GbcB, se medía la oxidación del NADH en ausencia de un aceptor de electrones. Por otra parte, el valor de h estimado no corresponde a una verdadera cooperatividad positiva entre sitios activos, dado que GbcB es una proteína monomérica. Probablemente se trata de una cooperatividad cinética que pudiera surgir de una activación de la enzima en el medio de ensayo.

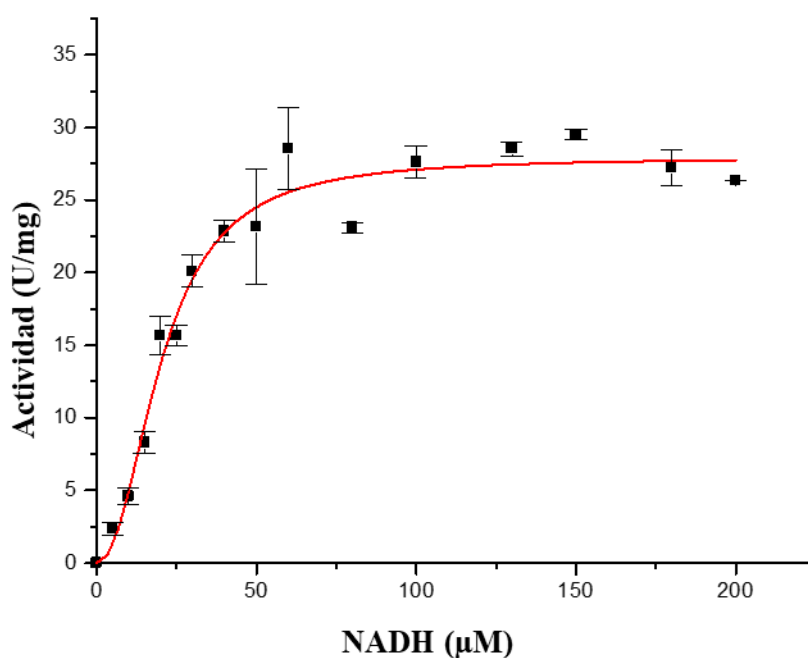


Figura 22. Curva de saturación de GbcB con NADH. Los puntos son los valores experimentales obtenidos; la línea trazada es la calculada a partir del mejor ajuste global de estos datos mediante regresión no lineal a la ecuación de Hill (Ecuación 3).

Tabla 3. Parámetros cinéticos aparentes de la enzima GbcB con NADH. Los valores de los parámetros se obtuvieron por el ajuste de los datos experimentales mostrados en la Figura 22 a la ecuación de Hill (Ecuación 3).

$V_{\text{máx}}$ ($\mu\text{mol min}^{-1} \text{mg proteína}^{-1}$)	27.9 ± 0.8
$S_{0.5}$ (μM)	20.3 ± 1.4
h	2.2 ± 0.3

VI.2.3. Caracterización cinética de la enzima recombinante GbcA

Para realizar la cinética de saturación de la enzima GbcA por GB se utilizó el ensayo acoplado con GbcB descrito en la sección de Métodos. En este ensayo GbcB obtiene electrones del NADH y los transfiere a GbcA, la cual finalmente los usa para hidroxilar a GB. La cinética de saturación se realizó como se describió en la sección de Métodos y los resultados se muestran en la Figura 23.

Los parámetros cinéticos estimados a partir de estos datos se incluyen en la Tabla 4. Son diferentes a los reportados para BmoA, la glicina betaína monooxigenasa de *C. salexigens* (Shao *et al.*, 2018). Si bien el método de ensayo enzimático es el mismo para ambas enzimas, la concentración de enzima utilizada y la relación reductasa/oxigenasa son distintas. Nosotros utilizamos GbcA 2 μM y GbcB 4 μM , es decir, una relación reductasa/oxigenasa de 2:1, mientras que en el ensayo reportado usaron BmoA (la oxigenasa) 17.5 μM y 3.4 μM de BmoB (la reductasa), es decir una relación reductasa/oxigenasa de aproximadamente 1:5. Esto no es congruente con el hecho de que las oxigenasas Rieske/mononuclear son trímeros en solución que unen una molécula reductasa por subunidad. Por ello, la relación reductasa-oxigenasa teórica debería de ser 3:1, en este trabajo decidimos utilizar la relación 2:1 ya que una relación 3:1 mostró un consumo basal elevado del NADH por parte de GbcB. Sin embargo, en el caso del ensayo de la glicina betaína monooxigenasa de *C. salexigens*, se reporta que la relación reductasa-oxigenasa en donde se observa una mayor actividad enzimática es con 1:5 (Shao *et al.*, 2018). En *P. aeruginosa*, se ha demostrado que, a nivel transcrito, la relación reductasa-oxigenasa es de 1:6 (Wargo *et al.*, 2008), *in vivo* esto sirve como mecanismo para evitar que en el caso de *P. aeruginosa*, GbcB oxide NADH y se encuentre cediendo los electrones al medio.

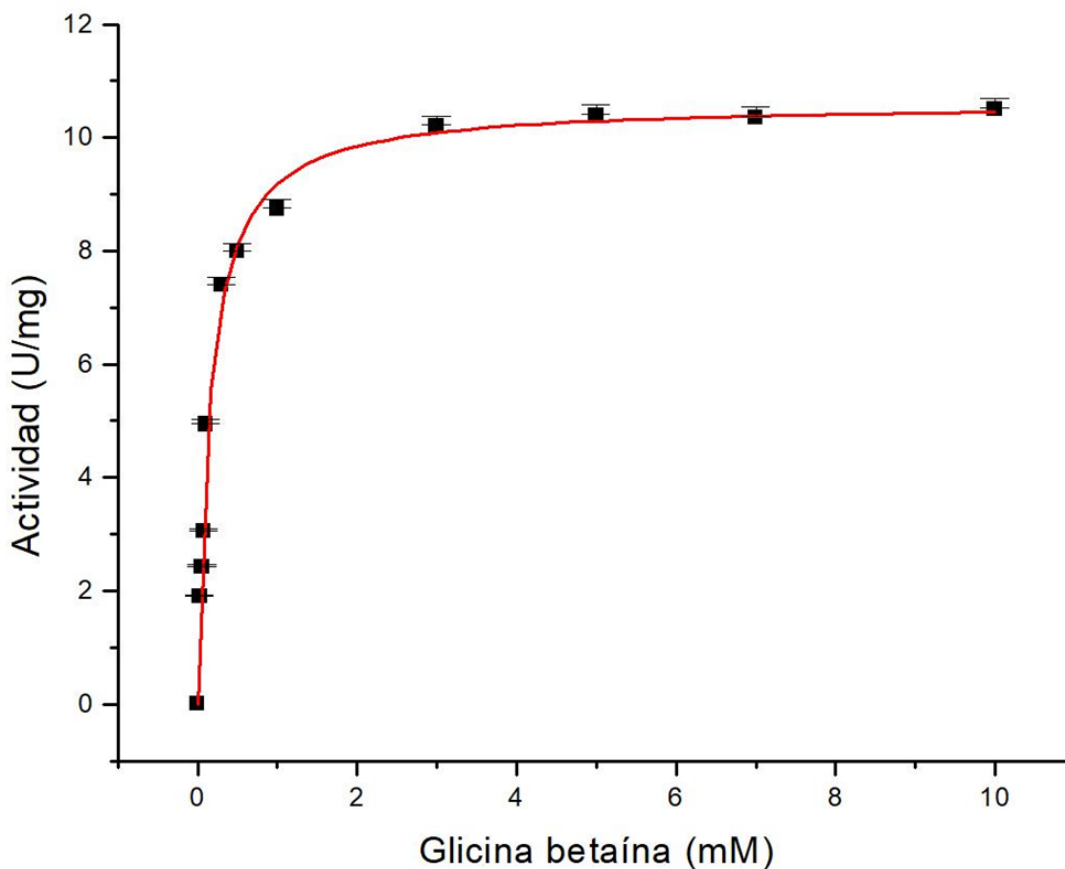


Figura 23. Curva de saturación de GbcA con GB. En la gráfica, los puntos son los valores experimentales obtenidos; la línea trazada es la calculada a partir del mejor ajuste global de estos datos mediante regresión no lineal a la ecuación de Michaelis-Menten (Ecuación 2).

Tabla 4. Parámetros cinéticos aparentes de la enzima GbcA con GB. Los valores de los parámetros se obtuvieron por el ajuste de los datos experimentales mostrados en la Figura 21 a la ecuación de Michaelis-Menten (Ecuación 2).

$V_{m\acute{a}x}$ ($\mu\text{mol. min}^{-1}$, mg proteína $^{-1}$)	10.6 ± 0.2
K_m (mM)	0.15 ± 0.02

Se exploró la especificidad por el sustrato de la enzima GbcA probando posibles sustratos (Fig. 24), se encontró que la enzima es altamente específica para GB y que no oxida a otros sustratos que son parecidos estructuralmente a este último. Esto último también fue reportado para BmoA, la glicina betaína monooxigenasa de *C. selxigens*.

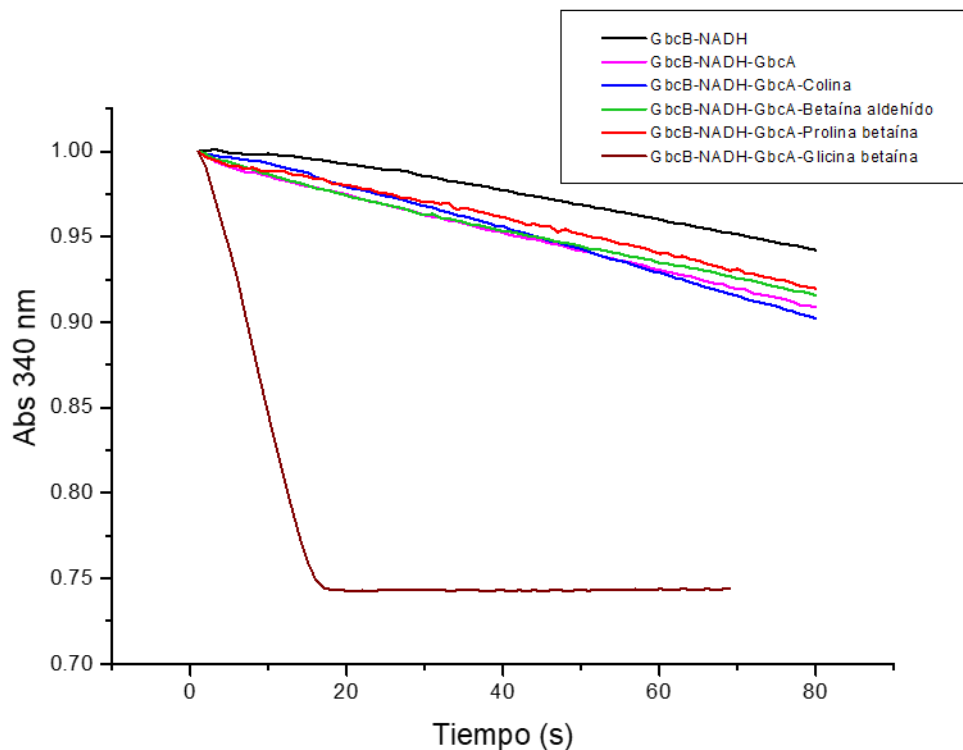


Figura 24. Especificidad por el sustrato de GbcA. Curso temporal de la actividad de la enzima GbcA con diferentes sustratos, cada uno se utilizó a una concentración fija de 10 mM. En la gráfica, los cursos temporales mostrados son representativos de los obtenidos en 3 repeticiones.

VI.2.4. Modelo por homología de la enzima GbcA

El modelo por homología de la enzima GbcA (Fig. 25A) se obtuvo utilizando como molde la estructura cristalográfica de la enzima prolina betaína desmetilasa que se obtuvo en ausencia de un ligando (código PDB: 3VCA; Daughtry *et al.*, 2012). En el modelo se incorporaron los centros metálicos tomando como referencia los del molde. El sitio activo de la prolina betaína desmetilasa está formado por seis residuos, cuando comparamos el probable sitio activo de GbcA con el del molde, encontramos que cuatro (Asn212, Cys215, Cys218, Leu225 e His319, numeración de GbcA) de los seis residuos que forman el sitio activo de la enzima usada como molde están conservados en GbcA. Los dos residuos cambiados son una fenilalanina que en GbcA es una leucina (Leu229), y una alanina que en GbcA es una treonina (Thr236) (Fig. 25B).

El cambio de una fenilalanina a una leucina pudiera alterar las dimensiones de la cavidad del sitio activo y esto estar relacionado con la especificidad de cada una de estas enzimas, ya que la prolina es más voluminosa que la GB. A pesar de que tanto GB como la prolina betaína son compuestos con un grupo trimetilamonio, en su probable sitio de unión tanto en la estructura cristalográfica de la prolina betaína desmetilasa como en el modelo por homología de GbcA no hay residuos aromáticos que pudieran establecer interacciones catión-pi con el grupo trimetilamonio, y tampoco residuos con carga negativa que pudieran contrarrestar la carga positiva del mismo, aunque la presencia de este tipo de residuos en el sitio activo se ha observado en otras enzimas cuyos sustratos poseen un grupo trimetilamonio (Nagy *et al.*, 2014). De hecho, creemos que estos residuos son determinantes para la unión del grupo trimetilamonio de la colina por las CMO2, como ya se discutió. En el caso de la prolina betaína desmetilasa, no se ha estudiado su especificidad por el sustrato por lo que se desconoce si esta enzima es capaz de oxidar otros compuestos y tampoco se conoce el papel que juegan los residuos que conforman el sitio activo en la unión del sustrato o en la catálisis. En el caso de GbcA sería interesante probar si puede oxidar a la dimetil glicina, que es el producto de su reacción, así como la sarcosina que es el producto de la dimetil glicina desmetilasa. BmoA no presenta actividad con estos dos últimos sustratos, pero de manera interesante, la Thr236 de GbcA, es un aspártico en la enzima BmoA, lo que pudiera sugerir alguna diferencia en la especificidad por el sustrato entre estas dos enzimas cuya actividad catalítica es la misma.

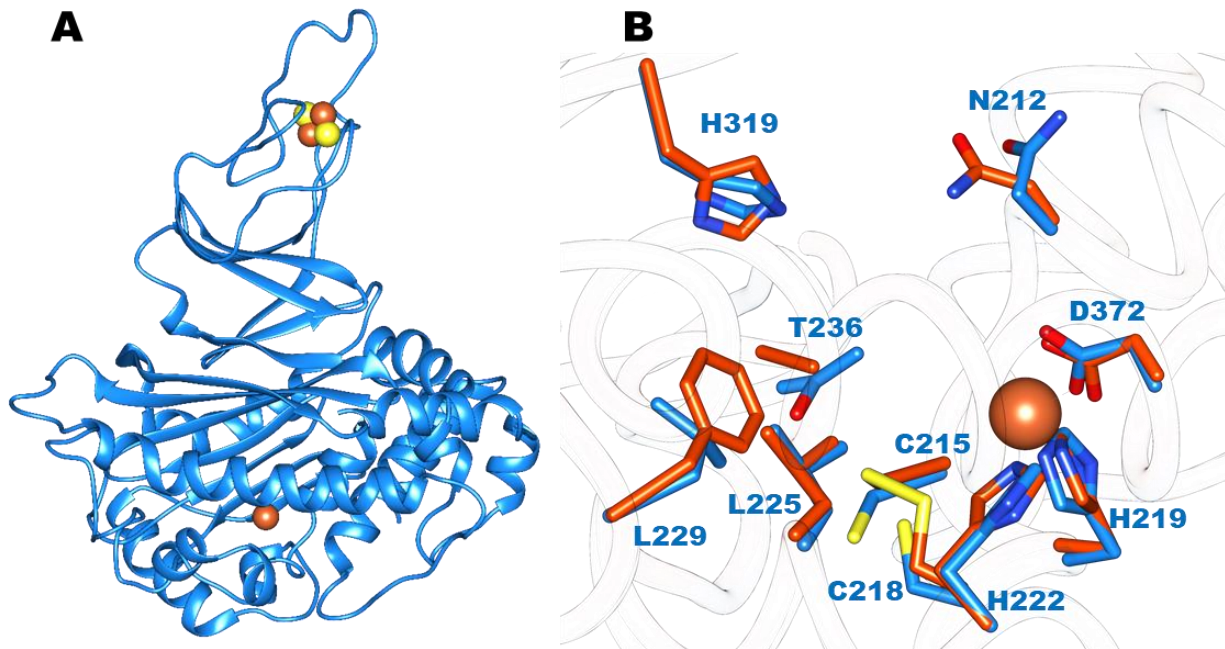


Figura 25. Modelo por homología de GbcA. (A) Representación esquemática de una subunidad del modelo por homología de GbcA. (B) Comparación del sitio de unión del sustrato entre el modelo por homología de GbcA (azul) y la estructura cristalográfica de la prolina betaína desmetilasa (naranja, PDB: 3VCA). Se muestran en representación de esferas los átomos de azufre (amarillo) e hierro (naranja). La numeración de los residuos del sitio activo corresponde a GbcA.

VII. CONCLUSIONES

COLINA MONOOXIGENASA

- El hallazgo de la existencia de dos isozimas CMO. La CMO1 presente en todas las plantas de genomas conocidos y la CMO2 presente solo en la familia Amaranthaceae y los resultados de los estudios filogenéticos nos permiten concluir:

1) Que la actividad colina monooxigenasa de las CMO2 se originó por subfuncionalización, a partir de las CMO1, cuyo papel fisiológico es aún desconocido, gracias a un evento de duplicación génica y a una fuerte presión de selección impuesta por el hábitat de la amarantáceas, de manera que estas plantas adquirieron la capacidad de sintetizar GB y ello les proporcionó una importante ventaja adaptativa.

2) Que la actividad de las CMO1 debe ser importante para las plantas ya que esta isozima parece estar presente en todas ellas.

- Los resultados de la caracterización cinética de la *S*oCMO2 nos permiten concluir:

1) Que las CMO2 no son tan específicas como se pensaba y que su sitio activo es capaz de unir y usar como sustratos compuestos químicamente tan dispares como la colina y el ácido salicílico, aunque su sustrato fisiológico sea la colina, según toda la evidencia presente.

2) Que esta plasticidad funcional posiblemente ha sido importante para la evolución no solo de la actividad colina monooxigenasa de las CMO2 sino también de la actividad fisiológica desconocida de las CMO1 a partir de una enzima bacteriana “CMO-like” ancestral.

3) Que los residuos aromáticos que forman la llamada “caja de aromáticos” que se han relacionado con la unión de grupos cargados positivamente, y sobre todo de grupos trimetilamonio, también es un arreglo estructural que puede unir compuestos aromáticos

GLICINA BETAÍNA MONOOXIGENASA

- Los resultados de la caracterización cinética de la GbcA nos permiten concluir que esta sí es una enzima específica para GB, lo que es fisiológicamente importante para el catabolismo de colina ya que ni colina ni betaína aldehído pueden competir con GB inhibiendo su degradación.
- Los resultados del modelo por homología y su comparación con la enzima BmoA de *C. selxigens* de nuevo indican:
 - 1) La plasticidad de los sitios activos de las monooxigenasas, en los que cambios de unos pocos residuos alteran la especificidad por el sustrato sin modificar el tipo de reacción catalizada.
 - 2) La diversidad de los residuos que pueden ser usados para unir a un mismo tipo de grupo químico, en este caso el grupo trimetilamonio.

VIII. PERSPECTIVAS

COLINA MONOOXIGENASA

- Ampliar el estudio de la especificidad por el sustrato de la *SoCMO2* para conocer qué otros compuestos esta enzima puede oxidar y determinar si la betaína aldehído o la GB pueden ser sustratos o inhibidores de la reacción con colina.
- Determinar la estructura cristalográfica de la *SoCMO2* para validar o rectificar la información obtenida sobre el sitio activo en el modelo por homología.
- Cambiar por alaninas mediante mutagénesis sitio-dirigida los residuos propuestos como importantes para la unión de la colina en la *SoCMO2* y caracterizar cinéticamente las enzimas mutantes para evaluar si en efecto estos residuos están involucrados en la especificidad por el sustrato y/o tienen algún papel en la catálisis.
- Cambiar mediante mutagénesis sitio-dirigida los residuos que se propone son importantes para la unión de la colina en la *SoCMO2* por los residuos en la posición equivalente en la *SoCMO1* para ver si se produce el cambio esperado en la especificidad por el sustrato.
- Determinar la estructura cristalográfica de la *SoCMO1* para validar o rectificar la información obtenida sobre el sitio activo en el modelo por homología.
- Para conocer la posible función fisiológica de la *SoCMO1*, investigar su especificidad por sustrato usando en primer lugar colina y los otros sustratos *in vitro* de la *SoCMO2*, y en segundo lugar otros compuestos aromáticos que sean fisiológicamente relevantes para las plantas.
- Obtener plantas de *Arabidopsis* carentes del gen *CMO1* e investigar su fenotipo bajo diferentes condiciones de crecimiento.

GLICINA BETAÍNA MONOOXIGENASA

- Ampliar el estudio de la especificidad por el sustrato de GbcA para conocer si esta enzima es capaz o no de oxidar otras moléculas.
- Cambiar por medio de mutagénesis sitio-dirigida los residuos involucrados en la unión del sustrato para evaluar su papel en la unión del sustrato y su importancia en la catálisis.

IX. REFERENCIAS

- Axcell BC, Geary PJ (1973) The metabolism of benzene by bacteria. *Biochem J* 136:927–934. doi: 10.1042/bj1360927
- Axcell BC, Geary PJ (1975) Purification and some properties of a soluble benzene-oxidizing system from a strain of *Pseudomonas*. *Biochem J* 146:173–183. doi: DOI 10.1042/bj1460173
- Barry SM, Challis GL (2013) Mechanism and catalytic diversity of rieske non-heme iron-dependent oxygenases. *ACS Catal* 3:2362–2370. doi: 10.1021/cs400087p
- Bashton M, Chothia C (2007) The generation of new protein functions by the combination of domains. *Structure* 15:85–99. doi: 10.1016/j.str.2006.11.009
- Batie CJ, Lahaie E, Ballous P (1987) Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from *Pseudomonas cepacia*. *J Biol Chem* 262:1510–1518
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. *Nucleic Acids Res* 41:36–42. doi: 10.1093/nar/gks1195
- Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, Raytselis Y, Sayers EW, Tao T, Ye J, Zaretskaya I (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* 41:29–33. doi: 10.1093/nar/gkt282
- Brouquisse R, Weigel P, Rhodes D, Yocum CF, Hanson AD (1989) Evidence for a ferredoxin-dependent choline monooxygenase from spinach chloroplast stroma. *Plant Physiol* 90:322–329. doi: 10.1104/pp.90.1.322

- Bruijninx PCA, Van Koten G, Gebbink RJMK (2008) Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: Recent developments in enzymology and modeling studies. *Chem Soc Rev* 37:2716–2744. doi: 10.1039/b707179p
- Burnet M, Lafontaine PJ, Hanson AD (1995) Assay, purification, and partial characterization of choline monooxygenase from spinach. *Plant Physiol* 108:581–588. doi: 10.1104/pp.108.2.581
- Burnet MW, Goldmann A, Message B, Drong R, El Amrani A, Loreau O, Slightom J, Tepfer D (2000) The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. *Gene* 244:151–161. doi: 10.1016/S0378-1119(99)00554-5
- Butler CS, Mason JR (1997) Structure-function Analysis of the Bacterial Aromatic Ring-hydroxylating Dioxygenases. In: Poole RK (ed) *Advances in Microbial Physiology*. Academic Press, Sheffield, UK., pp 47–84
- Chakraborty J, Dutta TK (2011) From lipid transport to oxygenation of aromatic compounds: Evolution within the Bet v1-like superfamily. *J Biomol Struct Dyn* 29:67–78. doi: 10.1080/07391102.2011.10507375
- Chakraborty J, Ghosal D, Dutta A, Dutta TK (2012) An insight into the origin and functional evolution of bacterial aromatic ring-hydroxylating oxygenases. *J Biomol Struct Dyn* 30:419–436. doi: 10.1080/07391102.2012.682208
- Chakraborty J, Suzuki-Minakuchi C, Okada K, Nojiri H (2017) Thermophilic bacteria are potential sources of novel Rieske non-heme iron oxygenases. *AMB Express* 7:2-15. doi: 10.1186/s13568-016-0318-5

- Chakraborty S, Behrens M, Herman PL, Arendsen AF, Hagen WR, Carlson DL, Wang XZ, Weeks DP (2005) A three-component dicamba O-demethylase from *Pseudomonas maltophilia*, strain DI-6: Purification and characterization. *Arch Biochem Biophys* 437:20–28. doi: 10.1016/j.abb.2005.02.024
- Chen THH, Murata N (2008) Glycine betaine: an effective protectant against abiotic stress in plants. *Trends Plant Sci* 13:499–505. doi: 10.1016/j.tplants.2008.06.007
- Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Bürglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Heng Tan CS, Hutchins AP, Weinmeier T, Rattei T, Chu JSC, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick D a, Lorenzo-Morales J, Bateman A, Chiu C-H, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus BJ (2013) Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14:1–14. doi: 10.1186/gb-2013-14-2-r11
- Cline K, Henry R (1996) Import and routing of nucleus-encoded chloroplast proteins. *Annu Rev Cell Dev Biol* 12:1–26. doi: 10.1146/annurev.cellbio.12.1.1
- Colbert CL, Couture MMJ, Eltis LD, Bolin JT (2000) A cluster exposed: Structure of the Rieske ferredoxin from biphenyl dioxygenase and the redox properties of Rieske Fe-S proteins. *Structure* 8:1267–1278. doi: 10.1016/S0969-2126(00)00536-0
- Craciun S, Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycy radical enzyme. *Proc Natl Acad Sci* 109:21307–21312. doi: 10.1073/pnas.1215689109

- Daughtry KD, Xiao Y, Stoner-Ma D, Cho E, Orville AM, Liu P, Allen KN (2012) Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J Am Chem Soc* 134:2823–2834. doi: 10.1021/ja2111898
- Davidson E, Ohnishi T, Atta-Asafo-Adjei E, Daldal F (1992) Potential ligands to the [2Fe-2S] Rieske cluster of the cytochrome bc₁ complex of *Rhodobacter capsulatus* probed by site-directed mutagenesis. *Biochemistry* 31:3342–3351. doi: 10.1021/bi00128a006
- Dong X, Fushinobu S, Fukuda E, Terada T, Nakamura S, Shimizu K, Nojiri H, Omori T, Shoun H, Wakagi T (2005) Crystal structure of the terminal oxygenase component of cumene dioxygenase from *Pseudomonas fluorescens* IP01. *J Bacteriol* 187:2483–2490. doi: 10.1128/JB.187.7.2483-2490.2005
- Dosztányi Z (2018) Prediction of protein disorder based on IUPred. *Protein Sci* 27:331–340. doi: 10.1002/pro.3334
- Ensley BD, Gibson DT (1983) Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. *J Bacteriol* 155:505–511
- Ensley BD, Gibson DT, Laborde AL (1982) Oxidation of naphthalene by a multicomponent enzyme system from *Pseudomonas* sp. strain NCIB 9816. *J Bacteriol* 149:948–954
- Ertekin E, Konstantinidis KT, Tezel U (2017) A Rieske-type oxygenase of *Pseudomonas* sp. BIOMIG1 converts benzalkonium chlorides to benzyldimethyl amine. *Environ Sci Technol* 51:175–181. doi: 10.1021/acs.est.6b03705

- Eswar N, Eramian D, Webb B, Shen M-Y, Sali A (2008) Protein structure modeling with MODELLER. In: Structural Proteomics: High-throughput Methods. pp 145–159
- Flagel L, Wendel J (2009) Gene duplication and evolutionary novelty in plants. *New Phytol* 183:557–564
- Fujiwara T, Hori K, Ozaki K, Yokota Y, Mitsuya S, Ichiyanagi T, Hattori T, Takabe T (2008) Enzymatic characterization of peroxisomal and cytosolic betaine aldehyde dehydrogenases in barley. *Physiol Plant* 134:22–30. doi: 10.1111/j.1399-3054.2008.01122.x
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40:1178–1186. doi: 10.1093/nar/gkr944
- Haddock JD, Horton JR, Gibson DT (1995) Dihydroxylation and dechlorination of chlorinated biphenyls by purified biphenyl 2,3-dioxygenase from *Pseudomonas* sp. strain LB400. *J Bacteriol* 177:20–26
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482:547–551. doi: 10.1038/nature10753
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98
- Hanson AD, May a M, Grumet R, Bode J, Jamieson GC, Rhodes D (1985) Betaine synthesis in chenopods: Localization in chloroplasts. *Proc Natl Acad Sci U S A* 82:3678–82

- Hayaishi O, Katagiri M, Rothberg S (1955) Mechanism of the pyrocatechase reaction. *J Am Chem Soc* 77:5450–5451. doi: 10.1021/ja01625a095
- Hegg EL, Jr LQ (1997) The 2-His-1-Carboxylate facial triad - an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* 250:625–629. doi: 10.1111/j.1432-1033.1997.t01-1-00625.x
- Hibino T, Meng YL, Kawamitsu Y, Uehara N, Matsuda N, Tanaka Y, Ishikawa H, Baba S, Takabe T, Wada K, Ishii T, Takabe T (2001) Molecular cloning and functional characterization of two kinds of betaine-aldehyde dehydrogenase in betaine-accumulating mangrove *Avicennia marina* (Forsk.) Vierh. *Plant Mol Biol* 45:353–363. doi: 10.1023/A:1006497113323
- Huijbers MME, Montersino S, Westphal AH, Tischler D, Van Berkel WJH (2014) Flavin dependent monooxygenases. *Arch Biochem Biophys* 544:2–17. doi: 10.1016/j.abb.2013.12.005
- Hurtubise Y, Barriault D, Sylvestre M (1998) Involvement of the terminal oxygenase β -subunit in the biphenyl dioxygenase reactivity pattern toward chlorobiphenyls. *J Bacteriol* 180:5828–5835
- Iozumi K, Hoganson GE, Pennella R, Everett MA, Fuller BB (1993) Role of tyrosinase as the determinant of pigmentation in cultured human melanocytes. *J Invest Dermatol* 100:806–811. doi: 10.1111/1523-1747.ep12476630
- Jiang H, Parales RE, Gibson DT (1999) The α subunit of toluene dioxygenase from *Pseudomonas putida* F1 can accept electrons from reduced ferredoxin (TOL) but is catalytically inactive in the absence of the β subunit. *Appl Environ Microbiol* 65:315–318

- Jiang H, Parales RE, Lynch NA, Gibson DT (1996) Site-directed mutagenesis of conserved amino acids in the alpha subunit of toluene dioxygenase: Potential mononuclear non-heme iron coordination sites. *J Bacteriol* 178:3133–3139. doi: 10.1128/jb.178.11.3133-3139.1996
- Joseph S, Murphy D, Bhave M (2013) Glycine betaine biosynthesis in saltbushes (*Atriplex* spp.) under salinity stress. *Biol* 68:879–895. doi: 10.2478/s11756-013-0229-8
- Kadereit G, Borsch T, Weising K, Freitag H (2004) Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *Int J Plant Sci* 164:959–986. doi: 10.1086/378649
- Karlsson A, Parales J V., Parales RE, Gibson DT, Eklund H, Ramaswamy S (2000) The reduction of the Rieske iron-sulfur cluster in naphthalene dioxygenase by X-rays. *J Inorg Biochem* 78:83–87. doi: 10.1016/S0162-0134(99)00213-5
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780. doi: 10.1093/molbev/mst010
- Kauppi B, Lee K, Carredano E, Parales RE, Gibson DT, Eklund H, Ramaswamy S (1998) Structure of an aromatic-ring-hydroxylating dioxygenase naphthalene 1,2-dioxygenase. *Structure* 6:571–586. doi: 10.1016/S0969-2126(98)00059-8
- Koehntop KD, Emerson JP, Que L (2005) The 2-His-1-carboxylate facial triad: A versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J Biol Inorg Chem* 10:87–93. doi: 10.1007/s00775-005-0624-x
- Kojima Y, Fujisawa H, Nakazawa A, Nakazawa T, Kanetsuna F, Taniuchi H, Nozaki M, Hayaishi O (1967) Studies on pyrocatechase. *J Biol Chem* 242:3270–3278

- Kortstee GJJ (1970) The aerobic decomposition of choline by microorganisms - I. The ability of aerobic organisms, particularly coryneform bacteria, to utilize choline as the sole carbon and nitrogen source. *Arch Mikrobiol* 71:235–244. doi: 10.1007/BF00410157
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. doi: 10.1093/molbev/msw054
- Kweon O, Kim SJ, Baek S, Chae JC, Adjei MD, Baek DH, Kim YC, Cerniglia CE (2008) A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem* 9:1–20. doi: 10.1186/1471-2091-9-11
- Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. doi: 10.1093/bioinformatics/btm404
- Leahy JG, Batchelor PJ, Morcomb SM (2003) Evolution of the soluble diiron monooxygenases. *FEMS Microbiol Rev* 27:449–479. doi: 10.1016/S0168-6445(03)00023-8
- Levin E, Levenberg B, Kaufman S (1960) The enzymatic conversion to norepinephrine. *J Biol Chem* 235:2080–2086
- Linder T (2014) CMO1 encodes a putative choline monooxygenase and is required for the utilization of choline as the sole nitrogen source in the yeast *Scheffersomyces stipitis* (syn. *Pichia stipitis*). *Microbiol (United Kingdom)* 160:929–940. doi: 10.1099/mic.0.073932-0

- Mallick S, Chakraborty J, Dutta TK (2011) Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular-weight polycyclic aromatic hydrocarbons: A review. *Crit Rev Microbiol* 37:64–90. doi: 10.3109/1040841X.2010.512268
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2017) CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:200–203. doi: 10.1093/nar/gkw1129
- Martínez-Castilla LP, Rodríguez-Sotres R (2010) A score of the ability of a three-dimensional protein model to retrieve its own sequence as a quantitative measure of its quality and appropriateness. *PLoS One* 5:1–19. doi: 10.1371/journal.pone.0012483
- Mason HS, Fowlks WL, Peterson E (1955) Oxygen transfer and electron transport by the phenolase complex. *J Am Chem Soc* 77:2914–2915. doi: 10.1021/ja01615a088
- Mason JR, Butler CS, Cammack R, Shergill JK (1997) Structural studies on the catalytic component of benzene dioxygenase from *Pseudomonas putida*. *Biochem Soc Trans* 25:90–95. doi: 10.1042/bst0250090
- Mason JR, Cammack R (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases. *Annu Rev Microbiol* 46:277–305. doi: 10.1146/annurev.micro.46.1.277
- Mayer AM, Harel E (1979) Polyphenol oxidases in plants. *Phytochemistry* 18:193–215. doi: 10.1016/0031-9422(79)80057-6

- Meng YL, Wang YM, Zhang DB, Nii N (2001) Isolation of a choline monooxygenase cDNA clone from *Amaranthus tricolor* and its expressions under stress conditions. *Cell Res* 11:187–193. doi: 10.1038/sj.cr.7290085
- Mitchell AJ, Weng J-K (2019) Unleashing the synthetic power of plant oxygenases: from mechanism to application. *Plant Physiol* 179:813-829. doi: 10.1104/pp.18.01223
- Mitra D, George SJ, Guo Y, Kamali S, Keable S, Peters JW, Pelmeshnikov V, Case DA, Cramer SP (2013) Characterization of [4Fe-4S] cluster vibrations and structure in nitrogenase Fe protein at three oxidation levels via combined NRVS, EXAFS, and DFT analyses. *J Am Chem Soc* 135:2530–2543. doi: 10.1021/ja307027n
- Muñoz-Clares RA, Riveros-Rosas H, Garza-Ramos G, González-Segura L, Mújica-Jiménez C, Julián-Sánchez A (2014) Exploring the evolutionary route of the acquisition of betaine aldehyde dehydrogenase activity by plant ALDH10 enzymes: Implications for the synthesis of the osmoprotectant glycine betaine. *BMC Plant Biol* 14:1–16. doi: 10.1186/1471-2229-14-149
- Nagy GN, Marton L, Contet A, Ozohanics O, Ardelean LM, Révész Á, Vékey K, Irimie FD, Vial H, Cerdan R, Vértessy BG (2014) Composite aromatic boxes for enzymatic transformations of quaternary ammonium substrates. *Angew Chemie - Int Ed* 53:13471–13476. doi: 10.1002/anie.201408246
- Nordlund P, Eklund H (1995) Di-iron-carboxylate proteins. *Curr Opin Struct Biol* 5:758–766. doi: 10.1016/0959-440X(95)80008-5
- Oka T, Simpson FJ (1971) Quercetinase, a dioxygenase containing copper I. *Biochem Biophys Res Commun* 43:1–5

- Oswald C, Smits SHJ, Höing M, Sohn-Bösser L, Dupont L, Le Rudulier D, Schmitt L, Bremer E (2008) Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J Biol Chem* 283:32848–32859. doi: 10.1074/jbc.M806021200
- Parales RE, Parales J V., Gibson DT (1999) Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity. *J Bacteriol* 181:1831–1837
- Park D (2007) Genomic DNA isolation from different biological materials. *Methods Mol Biol* 353:3–13. doi: 10.1385/1-59745-229-7:3
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi: 10.1002/jcc.20084
- Prabhakaran K, Kirchheimer WF, Harris EB (1968) Oxidation of phenolic compounds by *Mycobacterium leprae* and inhibition of phenolase by substrate analogues and copper chelators. *J Bacteriol* 95:2051–2053
- Pruzinska A, Tanner G, Anders I, Roca M, Hortensteiner S (2003) Chlorophyll breakdown: Pheophorbide a oxygenase is a Rieske-type iron-sulfur protein, encoded by the accelerated cell death 1 gene. *Proc Natl Acad Sci* 100:15259–15264. doi: 10.1073/pnas.2036571100
- Que L (2000) One motif - Many different reactions. *Nat Struct Biol* 7:182–184. doi: 10.1038/73270
- Radauer C, Lackner P, Breiteneder H (2008) The Bet v 1 fold: An ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol* 8:1–19. doi: 10.1186/1471-2148-8-286

- Rathinasabapathi B, Burnet M, Russell BL, Gage DA, Liao P-C, Nye GJ, Scott P, Golbeck JH, Hanson AD (1997) Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: Prosthetic group characterization and cDNA cloning. *Proc Natl Acad Sci* 94:3454–3458. doi: 10.1073/pnas.94.7.3454
- Resnick S, Lee K, Gibson D (1996) Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp strain NCIB 9816. *J Ind Microbiol Biotechnol* 17:438–457. doi: 10.1007/bf01574775
- Rhodes D, Hanson AD (1993) Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Annu Rev Plant Physiol.* 1993; 44: 357–384. doi: 10.1146/annurev.pp.44.060193.002041
- Rieske JS, MacLennan DH, Coleman R (1964) Isolation and properties of an iron-protein from the (reduced coenzyme Q)-cytochrome C reductase complex of the respiratory chain. *Biochem Biophys Res Commun* 15:338–344. doi: 10.1016/0006-291X(64)90171-8
- Rosche B, Tshisuaka B, Fetzner S, Lingens F (1995) 2-Oxo-1,2-dihydroquinoline 8-Monooxygenase, a two-component enzyme system from *Pseudomonas putida* 86. *J Biol Chem* 270:17836–17842. doi: 10.1074/jbc.270.30.17836
- Sagane Y, Zech K, Bouquet JM, Schmid M, Bal U, Thompson EM (2010) Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans. *Development* 137:1483–1492. doi: 10.1242/dev.044503
- Sato S, Nam J, Kasuga K, Nojiri H, Yamane H, Omori T (1997) Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10. *J Biol Chem* 272:4850–4858

- Scannell DR, Wolfe KH (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res* 18:137–147. doi: 10.1101/gr.6341207
- Schmidt CL, Shaw L (2001) A comprehensive phylogenetic analysis of rieske and rieske-type iron-sulfur proteins. *J Bioenerg Biomembr* 33:9–26. doi: 10.1023/A:1005616505962
- Shao Y-H, Guo L-Z, Zhang Y-Q, Yu H, Zhao B-S, Pang H-Q, Lu W-D (2018) Glycine betaine monooxygenase, an unusual rieske-type oxygenase system, catalyzes the oxidative n-demethylation of glycine betaine in *Chromohalobacter salexigens* DSM 3043. *84*:1–18
- Smith LT, Pocard J, Bernard T, Ruduliert DLE (1988) Osmotic control of glycine betaine biosynthesis and degradation in *Rhizobium meliloti*. *J Bacteriol* 170:3142–3149
- Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7:1–12. doi: 10.1186/gb-2006-7-s1-s10
- Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-containing oxygenases. *Chem Rev* 96:2841–2888. doi: 10.1021/cr9500500
- Summers RM, Louie TM, Yu CL, Gakhar L, Louie KC, Subramanian M (2012) Novel, highly specific N-demethylases enable bacteria to live on caffeine and related purine alkaloids. *J Bacteriol* 194:2041–2049. doi: 10.1128/jb.06637-11
- Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K, Okada K (1998) Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proc Natl Acad Sci* 95:12719–12723. doi: 10.1073/pnas.95.21.12719

- Trott O, Olson A (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 31:455–461. doi: 10.1002/jcc.21334.AutoDock
- van Berkel WJH, Kamerbeek NM, Fraaije MW (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* 124:670–689. doi: 10.1016/j.jbiotec.2006.03.044
- Van Doren SR, Gennis RB, Barquera B, Crofts AR (1993) Site-directed mutations of conserved residues of the rieske iron-sulfur subunit of the cytochrome *bc1* complex of *Rhodobacter sphaeroides* blocking or impairing quinol oxidation. *Biochemistry* 32:8083–8091. doi: 10.1021/bi00083a005
- Wargo MJ (2013) Homeostasis and catabolism of choline and glycine betaine: Lessons from *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 79:2112–2120. doi: 10.1128/AEM.03565-12
- Wargo MJ, Szwegold BS, Hogan DA (2008) Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J Bacteriol* 190:2690–2699. doi: 10.1128/JB.01393-07
- Weretilnyk EA, Hanson AD (1989) Betaine aldehyde dehydrogenase from spinach leaves: Purification, in vitro translation of the mRNA, and regulation by salinity. *Arch Biochem Biophys* 271:56–63. doi: 10.1016/0003-9861(89)90255-5
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699. doi: 10.1093/oxfordjournals.molbev.a003851

- White RF, Kaplan L, Birnbaum J (1973) Betaine-homocysteine transmethylase in *Pseudomonas denitrificans*, a vitamin B 12 overproducer. *J Bacteriol* 113:218–223
- Wolfe MD, Altier DJ, Stubna A, Popescu C V., Münck E, Lipscomb JD (2002) Benzoate 1,2-dioxygenase from *Pseudomonas putida*: Single turnover kinetics and regulation of a two-component rieske dioxygenase. *Biochemistry* 41:9611–9626. doi: 10.1021/bi025912n
- Wolfe MD, Parales J V., Gibson DT, Lipscomb JD (2001) Single turnover chemistry and regulation of O₂ activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J Biol Chem* 276:1945–1953. doi: 10.1074/jbc.M007795200
- Yoder MD, Thomas LM, Tremblay JM, Oliver RL, Yarbrough LR, Helmkamp GM (2002) Structure of a multifunctional protein. *J Biol Chem* 276:9246–9252. doi: 10.1074/jbc.m010131200
- Yoshizawa K, Kihara N, Kamachi T, Shiota Y (2006) Catalytic mechanism of dopamine β-monooxygenase mediated by Cu(III)-oxo. *Inorg Chem* 45:3034–3041. doi: 10.1021/ic0521168
- Zhu Y, Jameson E, Crosatti M, Schafer H, Rajakumar K, Bugg TDH, Chen Y (2014) Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *Proc Natl Acad Sci* 111:4268–4273. doi: 10.1073/pnas.1316569111

X.ANEXO

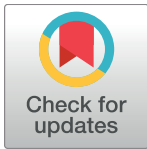
RESEARCH ARTICLE

Bona fide choline monoxygenases evolved in Amaranthaceae plants from oxygenases of unknown function: Evidence from phylogenetics, homology modeling and docking studies

Javier Carrillo-Campos¹, Héctor Riveros-Rosas², Rogelio Rodríguez-Sotres¹, Rosario A. Muñoz-Clares^{1*}

1 Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Ciudad de México, México, **2** Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México

* clares@unam.mx



OPEN ACCESS

Citation: Carrillo-Campos J, Riveros-Rosas H, Rodríguez-Sotres R, Muñoz-Clares RA (2018) Bona fide choline monoxygenases evolved in Amaranthaceae plants from oxygenases of unknown function: Evidence from phylogenetics, homology modeling and docking studies. *PLoS ONE* 13(9): e0204711. <https://doi.org/10.1371/journal.pone.0204711>

Editor: Sara Amancio, Universidade de Lisboa Instituto Superior de Agronomia, PORTUGAL

Received: July 30, 2018

Accepted: September 12, 2018

Published: September 26, 2018

Copyright: © 2018 Carrillo-Campos et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by DGAPA-UNAM PAPIIT grants (IN217814 and IN220317 to RAMC, IN225016 to HRR) and LANCAD-UNAM-DGTIC-215 to RRS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Few land plants can synthesize and accumulate the osmoprotectant glycine betaine (GB) even though this metabolic trait has major adaptive importance given the prevalence of drought, hypersaline soils or cold. GB is synthesized from choline in two reactions catalyzed by choline monoxygenases (CMOs) and enzymes of the family 10 of aldehyde dehydrogenases (ALDH10s) that gained betaine aldehyde dehydrogenase activity (BADH). Homolog genes encoding CMO and ALDH10 enzymes are present in all known land plant genomes, but since GB-non-accumulators plants lack the BADH-type ALDH10 isozyme, they would be expected to also lack the CMO activity to avoid accumulation of the toxic betaine aldehyde. To explore CMOs substrate specificity, we performed amino acid sequence alignments, phylogenetic analysis, homology modeling and docking simulations. We found that plant CMOs form a monophyletic subfamily within the Rieske/mononuclear non-heme oxygenases family with two clades: CMO1 and CMO2, the latter diverging from CMO1 after gene duplication. CMO1 enzymes are present in all plants; CMO2s only in the Amaranthaceae high-GB-accumulators plants. CMO2s, and particularly their mononuclear non-heme iron domain where the active site is located, evolved at a faster rate than CMO1s, which suggests positive selection. The homology model and docking simulations of the spinach CMO2 enzyme showed at the active site three aromatic residues forming a box with which the trimethylammonium group of choline could interact through cation- π interactions, and a glutamate, which also may interact with the trimethylammonium group through a charge-charge interaction. The aromatic box and the carboxylate have been shown to be critical for choline binding in other proteins. Interestingly, these residues are conserved in CMO2 proteins but not in CMO1 proteins, where two of these aromatic residues are leucine and the glutamate is asparagine. These findings reinforce our proposal that the CMO1s physiological substrate is not choline but a still unknown metabolite.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Land plants are sessile organisms that have evolved a great variety of strategies to escape from, or contend with, the many kinds of abiotic and biotic stresses to which they may be exposed during their lives. Osmotic stress—caused by drought, saline soils or low temperatures—constitutes the major limitation of agricultural production worldwide [1]. To cope with osmotic stress plants synthesize neutral, highly soluble, small organic compounds known as compatible solutes or osmoprotectants because they can be accumulated up to high concentrations without any toxic effect, thus preventing water loss and maintaining cell turgor in a hypertonic environment [2], as well as protecting the intracellular proteins from the noxious effects of abnormally high ion concentrations [3]. Glycine betaine (GB) is the most effective osmoprotectant [4] and those plants that synthesize it—known as GB-accumulators—tolerate osmotic stress much better than the non-accumulators [5]. In addition, given the observed health benefits of the intake of GB in humans [6] and animals [7], the ability of an edible plant to accumulate GB is important not only from an agricultural but also from a nutritional point of view. Because of these reasons and because many important crops and forage plants are GB-non-accumulators, the engineering of the ability to synthesize this osmoprotectant has been, and still is, a biotechnological goal [8]. However, although the synthesis of GB was increased in some transgenic plants, the levels of GB attained under osmotic stress conditions were well below those attained by the natural GB-accumulators (reviewed in [9]). This outcome emphasizes the need to get a deeper understanding of the enzymes involved in GB biosynthesis.

In land plants, GB is formed from choline in a short pathway consisting only of two steps (Fig 1A): formation of betaine aldehyde in a reaction catalyzed by choline monoxygenases (E.C.1.14.15.7; CMOs)—enzymes apparently found only in plants [10]—and the oxidation of betaine aldehyde to GB in a reaction catalyzed by betaine aldehyde dehydrogenases (E.C.1.2.1.8; BADHs) [11]. Plant BADHs belong to the family 10 of the aldehyde dehydrogenase (ALDH) superfamily [12,13]. CMOs are monoxygenases that contains Rieske-type [2Fe–2S] and mononuclear non-heme iron centers [14]. They catalyze the irreversible hydroxylation of

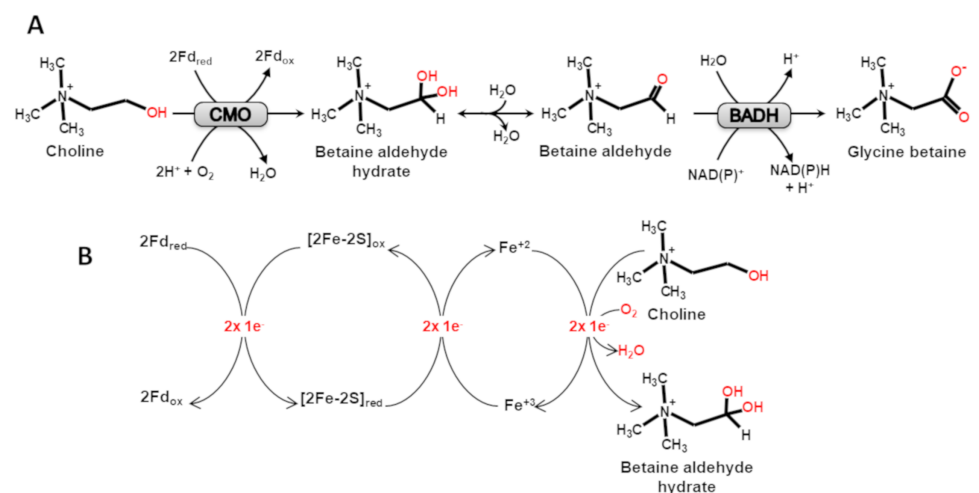


Fig 1. Biosynthesis of glycine betaine in plants. (A) Schematic representation of the two steps from choline. CMO, choline monoxygenase; BADH, betaine aldehyde dehydrogenase; Fd_{red} and Fd_{ox} , reduced and oxidized ferredoxin, respectively (B) Schematic representation of the electron transfer pathway from Fd_{red} to choline in the CMO catalyzed reaction. $[2Fe-2S]_{red}$ and $[2Fe-2S]_{ox}$, reduced and oxidized state of the CMO Rieske center, respectively; Fe^{2+} and Fe^{3+} , reduced and oxidized iron state, respectively, of the CMO mononuclear non-heme iron. The betaine aldehyde hydrate formed in this reaction is in equilibrium with the aldehyde form, which is the BADH substrate.

<https://doi.org/10.1371/journal.pone.0204711.g001>

choline (Fig 1B), using molecular oxygen and the electrons provided by reduced ferredoxin (Fd_{red}) to form betaine aldehyde hydrate, which in aqueous solution is in equilibrium with betaine aldehyde—the BADH substrate [15].

Gene duplication and a single change of a residue (Ala/Cys441 for Ile441, spinach BADH numbering) allowed the acquisition of the BADH activity by some ALDH10 isozymes [13,16]. The absence of the Ala/Cys441-type isozyme has been proposed to be a major limitation for the synthesis of GB in plants [13,16]. Given the high toxicity of betaine aldehyde [17], the CMO activity would be deleterious for the plant if it takes place without an accompanying BADH activity that converts betaine aldehyde to the innocuous GB. It would be then expected that CMO enzymes would be present only in GB-accumulator plants. However, the CMO gene has also been found in GB-non-accumulator plants [18,19], which raises the possibility that in these GB non-accumulators plants either: (i) their CMO genes are not expressed or expressed at a very low level; (ii) their CMO proteins are non-functional; or (iii) their CMO enzymes oxidize a substrate different from choline and, therefore, participate in a metabolic pathway different than the synthesis of GB. Abnormally processed CMO transcripts were found in rice [20], a GB-non-accumulator plant, and the oxidation of choline to betaine aldehyde by CMO proteins has been so far measured only in GB-accumulator plants that also have active BADH enzymes: *Spinacea oleracea* (spinach) [10,14], *Beta vulgaris* (sugar beet) [21], *Amaranthus caudatus* [21], and *Hordeum vulgare* (barley) [19]. Moreover, *Escherichia coli* cells transformed with the CMO gene from *Arabidopsis thaliana*, a GB-non-accumulator plant, expressed the CMO protein but were unable to produce GB when the cells were grown in the presence of choline, whereas the same *E. coli* cells transformed with a spinach CMO gene did produce GB [22]. These results were interpreted as a proof of *A. thaliana* CMO being non-functional, but these data are equally consistent with a different substrate specificity of this enzyme. In other words, the recombinant *A. thaliana* CMO protein may not use choline as substrate and, therefore, not be a functional choline monooxygenase, but still be a functional oxygenase acting on a different substrate.

In the present work we aimed at getting a deeper understanding of how the synthesis of GB from choline evolved in land plants by focusing on the study of the evolutive history of the CMO proteins. We carried out a comprehensive phylogenetic analysis of the available plant CMO sequences and found solid evidence supporting that they form two clades, which we named CMO1 and CMO2. CMO1 proteins are present in every higher plant of known sequence, including Amaranthaceae species, while CMO2 is only present in all species studied from the Amaranthaceae family. Also, to explore whether there are structural differences between these two kinds of plant CMO enzymes that could cause possible differences in their substrate specificity, and given that no CMO three-dimensional structure is known, we used homology modeling, molecular dynamics (MD) and docking simulations to study the two CMO proteins from the high-GB-accumulator Amaranthaceae plant *Spinacia oleracea* (SoCMO1 and SoCMO2). The active site aromatic and negatively charged residues that probably interact with the trimethylammonium group of choline in SoCMO2 are conserved in every CMO2 sequence but absent in the CMO1 sequences, a finding that strongly suggests that the physiological substrate of the latter enzymes is not choline but a still unknown compound.

Methods

Sequence analyses and phylogenetic studies

CMO amino acid sequences were retrieved by BLAST searches [23] from the non-redundant (NR) collection at the NCBI site (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [24] or on Phytozome v12.1.6 database (<https://phytozome.jgi.doe.gov>) [25], using the spinach CMO (NCBI

accession number XP_021849509) as bait. Progressive multiple amino acid sequence alignments were performed with ClustalX 2.0 [26] and were corrected manually using BioEdit [27] (<http://www.mbio.ncsu.edu/bioedit/bioedit>). For each retrieved CMO protein we performed the identification of their conserved domains using the NCBI's Conserved Domain Database (CDD; <http://www.ncbi.nlm.nih.gov/cdd/>) [28]. Only those sequences that have the cd03541 Rieske domain and the cd08883 mononuclear non-heme iron domain were considered as CMOs in our study (see [Results](#) section). When a retrieved amino acid sequence was incomplete or showed atypical insertions or deletions, the genomic or cDNA sequence for this protein was retrieved and the gene prediction software, Softberry FGENESH+ [29] (<http://www.softberry.com/>) was used to re-predict intron/exon gene structure, taking into account protein homology information from the more similar known complete CMO sequence. Likewise, if a CMO gene was not reported as such in a fully sequenced plant genome of the Phytozome database, a TBLASTN search [23] was performed to locate putative genomic sequences containing this CMO gene. Then these sequences were analyzed using FGENESH+ to predict the CMO gene structure. The corrected or newly predicted CMO amino acid sequences were included in the final multiple amino acid sequence alignment.

Phylogenetic analysis was carried out with MEGA7 [30]. Phylogenetic relationships were inferred by using the Maximum Likelihood. Initial phylogenetic trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The amino acid substitution model of Whelan and Goldman [31], using a discrete Gamma distribution with 5 categories, was chosen as the best substitution model, since it gave the lowest Bayesian Information Criterion values and corrected Akaike Information Criterion values in MEGA7 [30]. The gamma shape parameter value (+G parameter) was estimated directly from the data with MEGA7. The rate variation model allowed for some sites to be evolutionarily invariable. Confidence for the internal branches of the phylogenetic trees was determined through bootstrap analysis with 500 replicates. The same strategy was used to separately infer the evolutionary history of the Rieske and the mononuclear non-heme iron domains of the CMO proteins. With this aim, we identified the amino acid residues at the beginning and end of both domains in the spinach CMO (XP_021849509) sequence and then, based in the alignment of the complete sequences, we split all CMO sequences in their two domains, ensuring that both domains were complete. Consensus sequences were obtained from the multiple sequence alignments with a cut-off of 60% conservation. The software ESPript 3 [32] was used to show the results of the sequence consensus analysis. Sequence logos of selected active-site residues were constructed using the WebLogo server [33] (<http://weblogo.threeplusone.com>).

Homology modeling

To select the crystal structure to be used as template in the generation of the CMOs homology models we performed multiple amino acid sequence alignments with the bacterial Rieske/mononuclear non-heme iron oxygenases with a reported crystal structure, using the L-INS-i method in the MAFFT program [34]. Once the template was selected, the starting model for SoCMO2 as a trimer was prepared with MODELLER v9.14 [35]. This model was improved subjecting it to approximately 200 ns of molecular dynamics (MD) simulation, which was performed on GROMACS 4.6.5 software with the AMBER99SB-ILDN force-field and using an explicit TIP3P water periodical box with roughly 0.15 M NaCl adjusted to a neutral system. Bond constraints were imposed using the LINCS algorithm to allow a 2 fs integration interval. Electrostatics were handled with the Particle Mesh Ewald (PME) method. The system was

maintained at 313 K with a velocity-rescaled Berendsen thermostat and at 1 atm with a Berendsen barostat to provide a constant-temperature, constant-pressure ensemble (NPT). The relaxed model was retrieved by clustering the last 5 ns of simulation, and its geometry was taken to equilibrium by energy minimization. The SoCMO2 model was used as template to obtain the homology model of the *Arabidopsis thaliana* CMO1 (AtCMO1), and this model was in turn used as template to obtain the model for the SoCMO1 sequence. The initial models of the two latter proteins were prepared with MODELLER v9.14 and then subjected to approximately 800 ns (AtCMO1) or 60 ns (SoCMO1) of MD simulation using AMBER force-fields and energy minimization, as described above. Disordered regions at the N-terminal of the proteins were predicted using the program IUPred [36].

Topologies for the [2Fe-2S] Rieske and mononuclear non-heme iron centers were built using QM calculations (Hartree-Fock 6-31G**; GAUSSIAN 09) [37], ACPYPE [38] and AmberTools software [39], as reported by Mitra et al. (2013) [40]. To better represent the geometrical and chemical features of the mononuclear non-heme iron, a dioxygen and a water molecule were linked to it. QM calculations were done with the low-spin configuration for an oxidized Rieske [2Fe-2S] center [41], and with an intermediate number of unpaired spin electrons in the case of the Fe³⁺/O₂/H₂O mononuclear non-heme Fe center [42]. The validation of these topologies was performed analysing the geometrical fluctuations in short simulations (1 ns) of an isolated molecule comprising the [2Fe-2S] Rieske and its covalently linked amino acid residues, or the mononuclear non-heme Fe and its covalently linked amino acid residues. The final topologies of the [2Fe-2S] Rieske center and mononuclear non-heme iron were manually merged with the models final subunit topology, and duplicated declarations were deleted. The Rd.HMM protocol [43] was used to test the appropriateness of the models backbones as hosts for the SoCMO2, AtCMO1 and SoCMO1 sequences. Briefly, this protocol retrieves from the UniProtKB/Swiss-Prot protein sequence database all sequences that are structurally compatible, in a higher or lesser degree, with the backbone of the homology model tested without considering the sequence information in the PDB. For a good quality model, the protein sequence with the highest score, i.e. the first hit, should correspond to the protein modelled [43]. The Rosetta energy function was also used for evaluation of the models. Structural alignments of the models were performed using the Structural Alignment of Multiple Proteins (STAMP) tool in the VMD software [44].

Docking simulations

A flexible choline molecule was docked into the active site of the rigid SoCMO2 model with AutoDock Vina [45], using a box of 10 Å³ centered near the mononuclear iron. Sixty poses were visually inspected to select the one that fulfilled the criterion of having a proper orientation of the choline hydroxyl group with respect to the mononuclear non-heme iron, as it would be needed for the reaction to take place. Figures of the homology models and docking simulations were made using the UCSF Chimera software [46].

Results

Plant CMOs are bidomain enzymes that form a monophyletic protein subfamily

All identified CMO proteins possess a Rieske-type domain at their N-terminal half and a mononuclear non-heme iron domain at their C-terminal half. Similar domain architecture of the CMO proteins can be observed in several bacterial oxygenases that catalyze the oxidation of a variety of hydrophobic, mainly aromatic, compounds by the insertion of one or two

hydroxyl groups [47]. According to the Conserved Domain Database (CDD), the N-terminal Rieske domain of CMO proteins belongs to the cd03541 (Rieske_Ro_Alpha_N_CMO) subfamily of the Rieske_Ro_Alpha_N protein family (cd03469), and their C-terminal catalytic domain belongs to the cd08883 (Rho_Alpha_C_CMO-like) subfamily of the Rho_Alpha_C protein family (cd00680). Indeed, all putative CMO sequences retrieved from the NCBI and Phytozome databases, using the amino acid sequence of *Spinacia oleracea* CMO XP_021849509 as bait, have, without exceptions, an N-terminal domain that belong to cd03541 subfamily and a C-terminal domain that belong to the cd08883 subfamily. Therefore, only those protein sequences with these two domains were used for our sequence and phylogenetic analyses. With this criterion, a total of 167 protein sequences (S1 Table) were selected. Although of all them have both domains, some sequences were incomplete or appear to have atypical insertions or deletions; these sequences were revised and most of them were corrected following the procedure described in the Methods section. These corrected or newly predicted CMO amino acid sequences were also included in the final multiple amino acid sequence alignment (S1 Fig). In a few cases, the FGENESH+ program confirmed that the insertion or deletion was correct and therefore the original reported sequences were used in the multiple amino acid alignment, where they introduced gaps (S1 Table). Because of these gaps, there were a total of 743 positions in the final alignment used for the ensuing phylogenetic analysis.

To test whether the CMO subfamily is monophyletic, the N-terminal Rieske domains of the CMO sequences were aligned with the Rieske domains identified at the CDD as members of the Rieske_Ro_Alpha_N protein family (cd03469), including its respective protein subfamilies. In the same way, the C-terminal mononuclear non-heme iron domains of the CMO sequences were aligned with the other mononuclear non-heme iron domains identified at the CDD as members of the Rho_Alpha_C family (cd00680) and its respective protein subfamilies. The results of these analyses showed that the two domains of all retrieved CMO sequences, without exception, belong to subfamily cd03541 and subfamily cd08883 and are from eukaryotes (Fig 2A and 2B). Interestingly, CMO sequences were identified also in chordate animals: *Saccoglossus kowalevskii* (acorn worm; hemichordata; 3 sequences), *Branchiostoma floridae* (Florida lancelet; cephalochordata; 3 sequences) and *B. belcheri* (Belcher's lancelet; cephalochordata; 4 sequences), as well as in two protists: *Nannochloropsis gaditana* (microalgae; heterokontophyta; 1 sequence) and *Acanthamoeba castellanii* (Amoebozoa; 1 sequence). Since the genome of this last species (*A. castellanii*) suffered extensive lateral gene transfer through its evolution [48], it is probable that the presence of a CMO in this organism is also a consequence of a lateral gene transfer event. Likewise, a lateral gene transfer event could also explain the presence of CMO in *S. kowalevskii*, *B. floridae* and *B. belcheri*, because CMOs were not found in any other animal species. Furthermore, CMO proteins were not found either in fungi or archaea, but we retrieved sequences from fungi and bacteria having a C-terminal domain that belongs to the cd08883 subfamily, as do the C-terminal domain of plant CMOs. However, the N-terminal domains (Rieske domains) of these fungal and bacterial proteins do not belong to the cd03541 subfamily, or to any of the other 15 numbered subfamilies of the Rieske_Ro_Alpha_N protein family (cd03469). They do belong to the comprehensive cd03469 family, but a subfamily cd number has not been yet assigned to them (Fig 2A and 2B). We named these proteins as fungal or bacterial CMO-like. Indeed, consistent with this finding, a putative CMO from the yeast *Pichia stipitis* has been reported and named CMO1 because of its homology to plant CMOs and because it appears to be essential for the growth of the yeast on choline as the only nitrogen source, although it was not proved whether choline is a substrate of this enzyme [49].

Other bacterial Rieske/mononuclear non-heme sequences have a variety of different Rieske and mononuclear non-heme domains, as depicted in Fig 2C. As shown in this figure, domain cd03451 is exclusive of plant CMO proteins, as is the particular association of domain cd03541

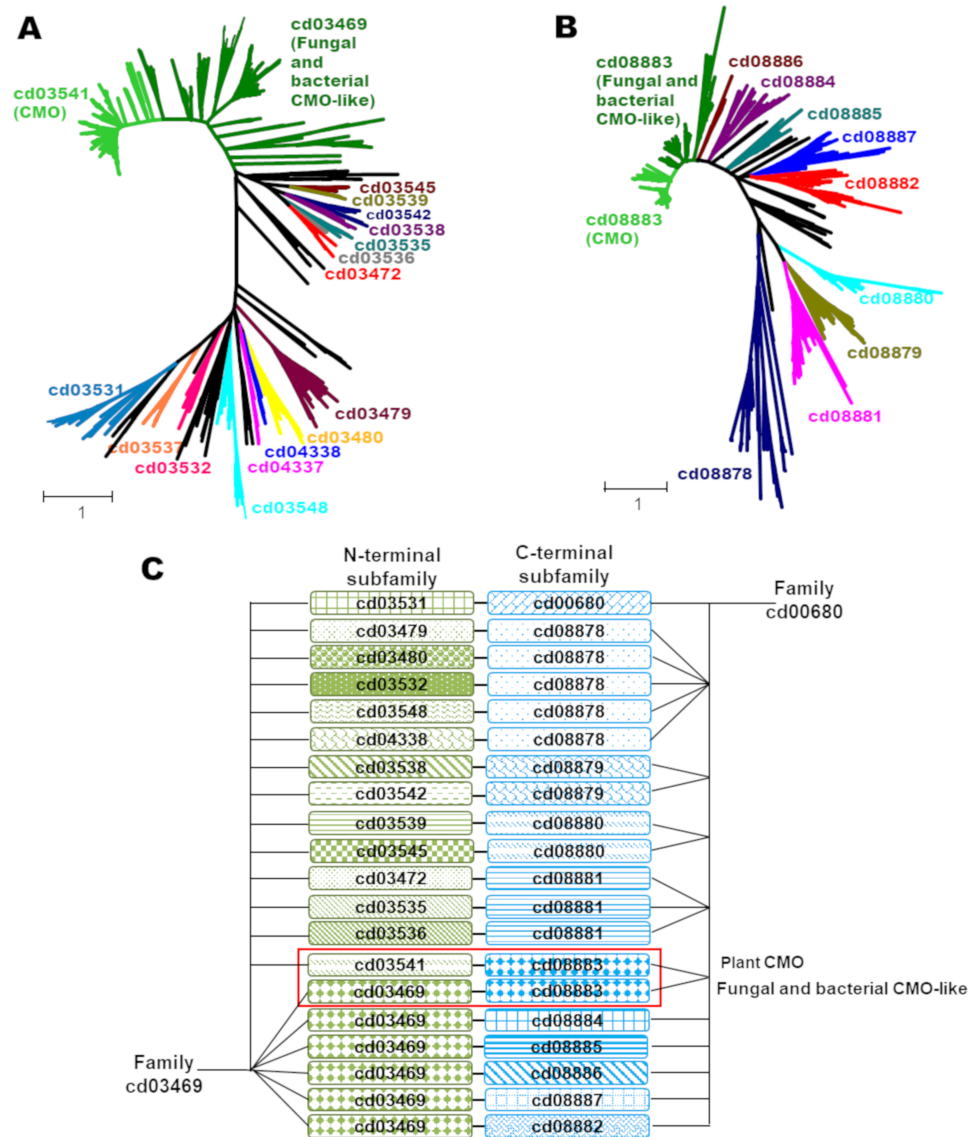


Fig 2. Phylogenetic trees of the bidomain Rieske/mononuclear non-heme iron protein sequences. (A) Tree of the N-terminal Rieske domain of retrieved protein sequences identified at the CDD as members of the oxygenase alpha-subunit N-terminal Rieske domain protein family (RHO_alpha_C, cd03469) or any of its protein subfamilies identified with different cd numbers. The Rieske domain of the bacterial CMO-like (dark green branches) has not a subfamily cd number assigned yet, as well as the Rieske domains of other proteins (black branches). (B) Tree of C-terminal mononuclear non-heme iron domain of retrieved sequences identified at the CDD as members of the oxygenase alpha subunit C-terminal catalytic domain protein family (Rieske_RO_Aplha_N, cd00680), or any of its subfamilies. As in (A) subfamilies to which a cd number has not been assigned yet are depicted as black branches. With the exception of CMOs, all retrieved sequences were bacterial. (C) Schematic representation of the association of the Rieske domains with the mononuclear non-heme iron domains in the retrieved proteins.

<https://doi.org/10.1371/journal.pone.0204711.g002>

with domain cd08883; the domain shuffling events that have occurred in other Rieske/mononuclear non-heme monoxygenases were not detected in CMO proteins. The closer homologs of CMOs are the already mentioned fungal and bacterial CMO-like proteins, which have a mononuclear non-heme iron domain of the same subfamily (cd08883) as the one of plant CMO proteins but different Rieske domain. Thus, bacterial CMO-like, fungal CMO-like and plant CMOs proteins possibly share a common ancestor.

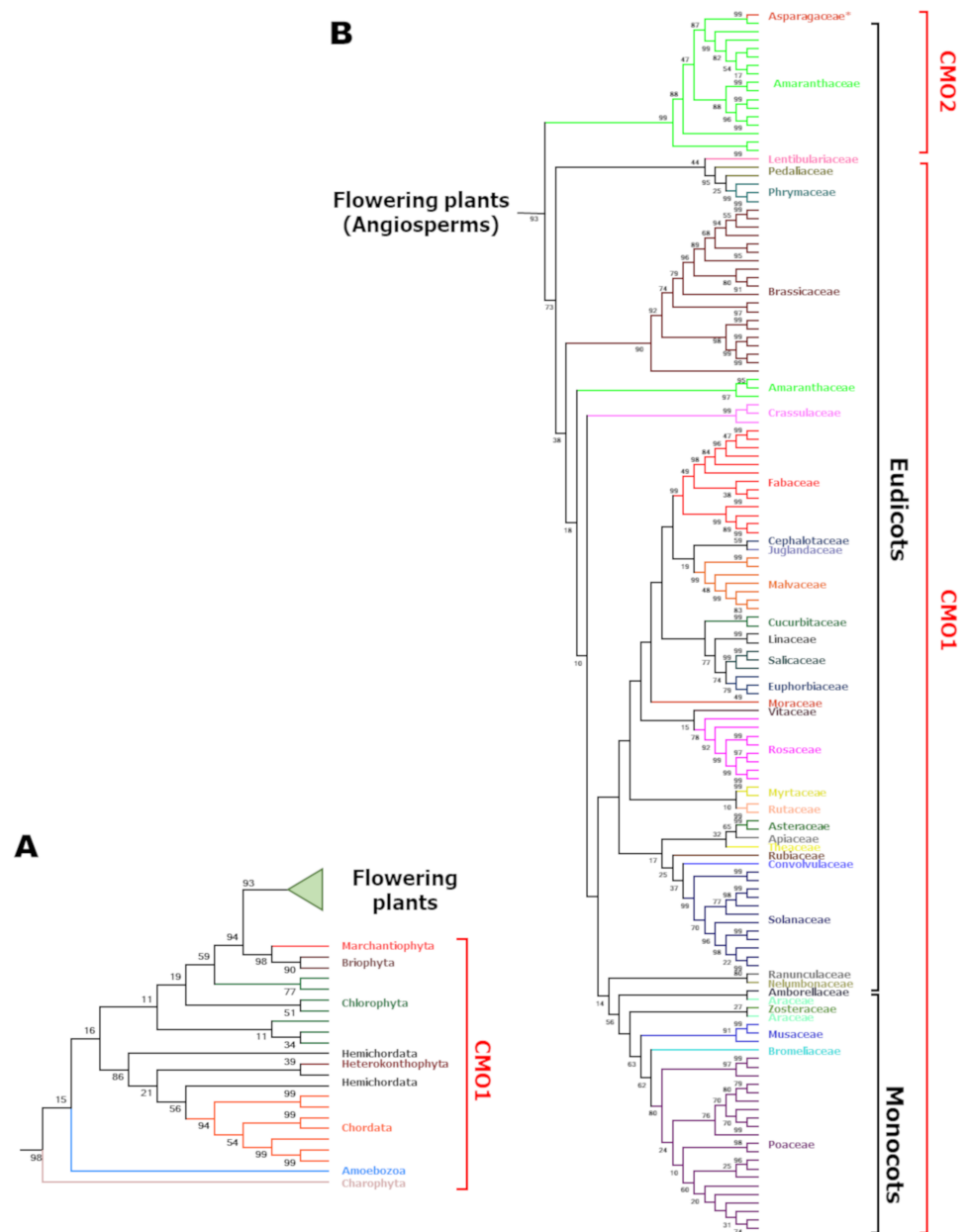


Fig 3. Phylogenetic tree of CMO proteins. (A) Sequences from primitive photosynthetic eukaryotes, chordates and protist species. (B). Flowering Plants. *The mononocot *Ophiopogon japonicus* (Asparagaceae) species groups with the dicot Amaranthaceae species within the CMO2 clade. The tree was inferred using the Maximum Likelihood method based on the Whelan and Goldman model [31] as described in the Methods section. The +G, parameter value was 2.0069 and the [+I] value was 0.94% of sites. Branches are colored according to the taxonomic family. The best tree with the highest log likelihood (-108,453.30) is shown. The percentage of trees in which the associated taxa clustered together in a bootstrap test (500 replicates) is shown next to the branches. The tree is drawn to scale and only the branch topology is shown, with branch lengths measured in the number of substitutions per site. The analysis involved 167 amino acid sequences (156 from plants and 11 from non-plant eukaryotic organisms) given in S1 Table.

<https://doi.org/10.1371/journal.pone.0204711.g003>

We performed multiple sequence alignments considering the identified eukaryotic CMO protein sequences that contain the pair domain cd03541/cd08883, and constructed the phylogenetic tree shown in Figs 3 and 4A and S2 Fig. As expected, flowering plants (angiosperms) form a well-supported monophyletic group, as well as all eukaryotic CMO protein sequences.

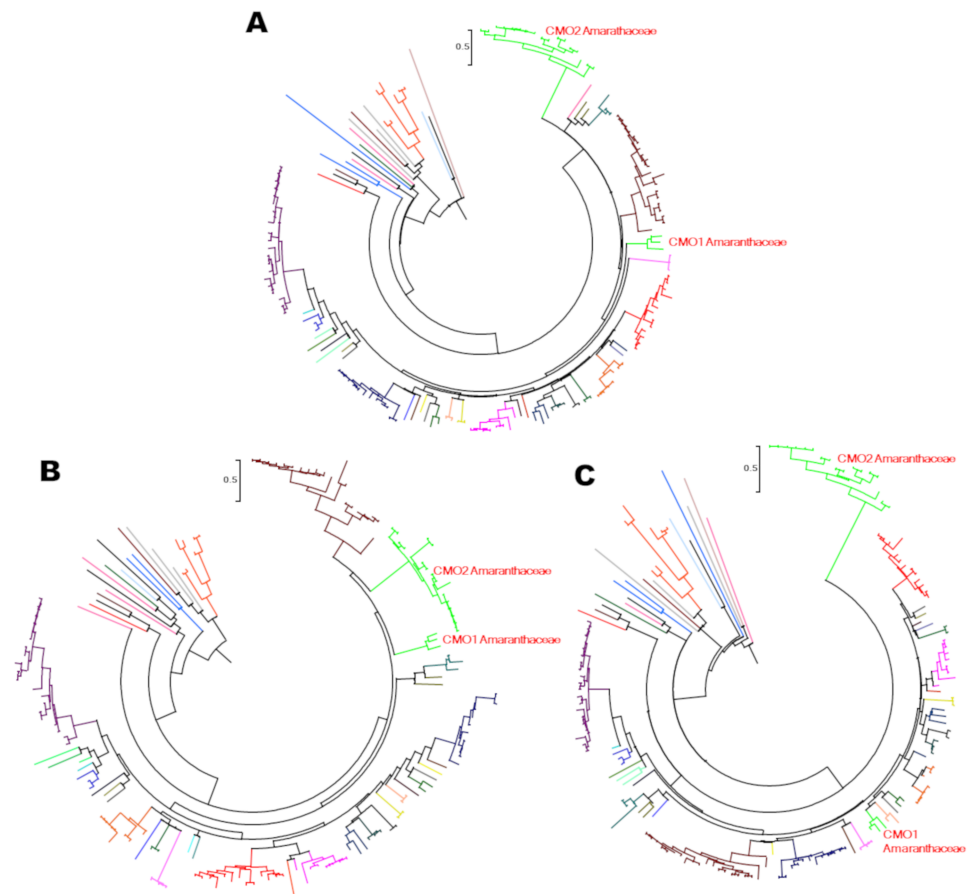


Fig 4. Evolutionary rate of the CMO proteins domains. The evolutionary history of the two domains of CMO proteins was separately inferred by splitting all CMO sequences in their two domains. The phylogenetic trees were obtained using the Maximum Likelihood method, as described in the legend of Fig 3. (A) Circle representation of the tree shown in Fig 3. (B) Circle representation of the phylogenetic tree of the CMO Rieske domain. The tree with the highest log likelihood (-39,113.47) is shown. The +G, parameter value was 0.7879 and the [+I] value was 0.91% of sites. (C) Circle representation of the phylogenetic tree of the CMO non-heme mononuclear iron domain. The tree with the highest log likelihood (-517,254.51) is shown. The +G, parameter value was 0.7024 and the [+I] value was 0.14% of sites. In the three panels, the CMO1 and CMO2 proteins of the Amaranthaceae family are indicated.

<https://doi.org/10.1371/journal.pone.0204711.g004>

Primitive photosynthetic eukaryotes with a known genome like *Nannochloropsis gaditana* (Heterokontophyta), *Micromonas pusilla*, *Micromonas* sp. RCC299, *Coccomyxa subellipsoidea* C-169, *Bathycoccus prasinus* (all of them belonging to Chlorophyta), or *Klebsormidium flaccidum* (Charophyta) also contain a CMO enzyme. However, the genomes of the chlorophytes *Chlamydomonas reinhardtii*, *Volvox carteri*, *Ostrococcus tauri* and *O. lucimarinus* lack CMO sequences; therefore, secondary events of loss should be postulated to explain CMO gene absence in these species. Thus, although genes coding for CMO proteins are present in a few eukaryotic organisms different from photosynthetic eukaryotes ([49] and this work), they more likely resulted from horizontal gene transfer events and therefore it can be safely concluded that CMO proteins containing the pair domains cd03541/cd08883 exclusively evolved in photosynthetic eukaryotes.

The CMO phylogenetic tree shows two clades: one containing CMO proteins from all plant families that have at least one species with a sequenced CMO gene, including Amaranthaceae, and another one including only CMO proteins from Amaranthaceae species, which most likely

resulted from a duplication gene event. One of the two duplicates evolved rapidly, significantly diverging from the other and eventually forming a different clade with a good bootstrap support (94%) (Figs 3 and 4A). We named this clade as CMO2 to differentiate it from the other clade, which includes all other plant CMO proteins, hereafter referred as CMO1. Sequence identity between the proteins grouped in clade CMO1 is higher than 50%, and between proteins of the clade CMO2 is higher than 85%, while sequence identity between CMO1 and CMO2 proteins is 30% at the most. Our results indicate that during plant evolution after monocot-eudicot divergence several other independent CMO gene-duplication events occurred in 9 plant families different from Amaranthaceae. Specifically, we found duplicates in *Camelina sativa*, *Arabis alpina*, *Raphanus sativus*, *Kalanchoe laxiflora*, *Phaseolus vulgaris*, *Gossypium hirsutum*, *Linum usitatissimum*, *Malus domestica*, *Nicotiana tabacum*, *Panicum virgatum*, *Erythranthe guttata*, *Solanum penelli* and *Solanum lycopersicum*. However, these other duplication events are recent, as indicated by the high sequence identity between the proteins encoded by the duplicated genes.

CMO from *Ophiopogon japonicus* (Asparagaceae, a monocot) has higher amino acid sequence identity with CMO2 proteins from the Amaranthaceae family than with CMO1 from monocots (Fig 3 and S2 Fig). This unexpected position of the CMO from *O. japonicus* was also reported by Joseph *et al.* [18]. Interestingly, we have previously found that the *O. japonicus* ALDH10 sequence clusters with the ALDH10 proteins of the Amaranthaceae 2 clade, which includes enzymes with proved or predicted betaine aldehyde dehydrogenase activity [13]. One possible explanation for these findings is that both genes were acquired by *O. japonicus* by horizontal gene transfer from an unidentified Amaranthaceae, a possibility supported by the fact that horizontal gene transfer is a significant force in the evolution of plant genomes [50,51]. Further studies are needed to provide evidence in favor or against this possibility.

The non-heme mononuclear iron domains of CMO1 and CMO2 proteins evolved at different rates

To explore whether the Rieske and the non-heme mononuclear Fe domains of the plant CMO proteins evolved at the same or at a different rate, we constructed phylogenetic trees of each of these two domains separately (Fig 4B and 4C). We found that within the CMO1 sequences the rate of evolution of both domains was the same, and comparable to the rate of evolution of the Rieske domain of the CMO2 proteins, but the non-heme mononuclear domains of the CMO2 proteins appears to have evolved at a much faster rate, as can be inferred by the length of their branches (Fig 4C). These findings suggest that the catalytic domain of the CMO2 proteins was under some kind of evolutionary pressure that probably led to the acquisition of new substrate specificity, as it will be shown below.

The alignments of CMO1 and CMO2 consensus amino acid sequences (Fig 5) show differences between them scattered throughout the entire sequences. The possible chloroplast signal peptide, which was proposed to be located at the N-terminal region in the SoCMO2 protein [14], is well conserved in CMO2 sequences but not in the CMO1 ones. Moreover, alignments of the consensus CMO1 sequences with the consensus CMO1 and CMO2 sequences from Amaranthaceae and CMO1 sequences from Poaceae (Fig 5) showed that the N-terminal region of the CMO1 proteins from these two families is shorter than that of the rest of the CMO1 proteins. The amino acids involved in the Rieske center and in binding the mononuclear non-heme iron, as well as the aspartate residue that participate in the electron transfer between the two, are fully conserved in all CMO1 and CMO2 proteins, indicating that all of them are active oxygenases.

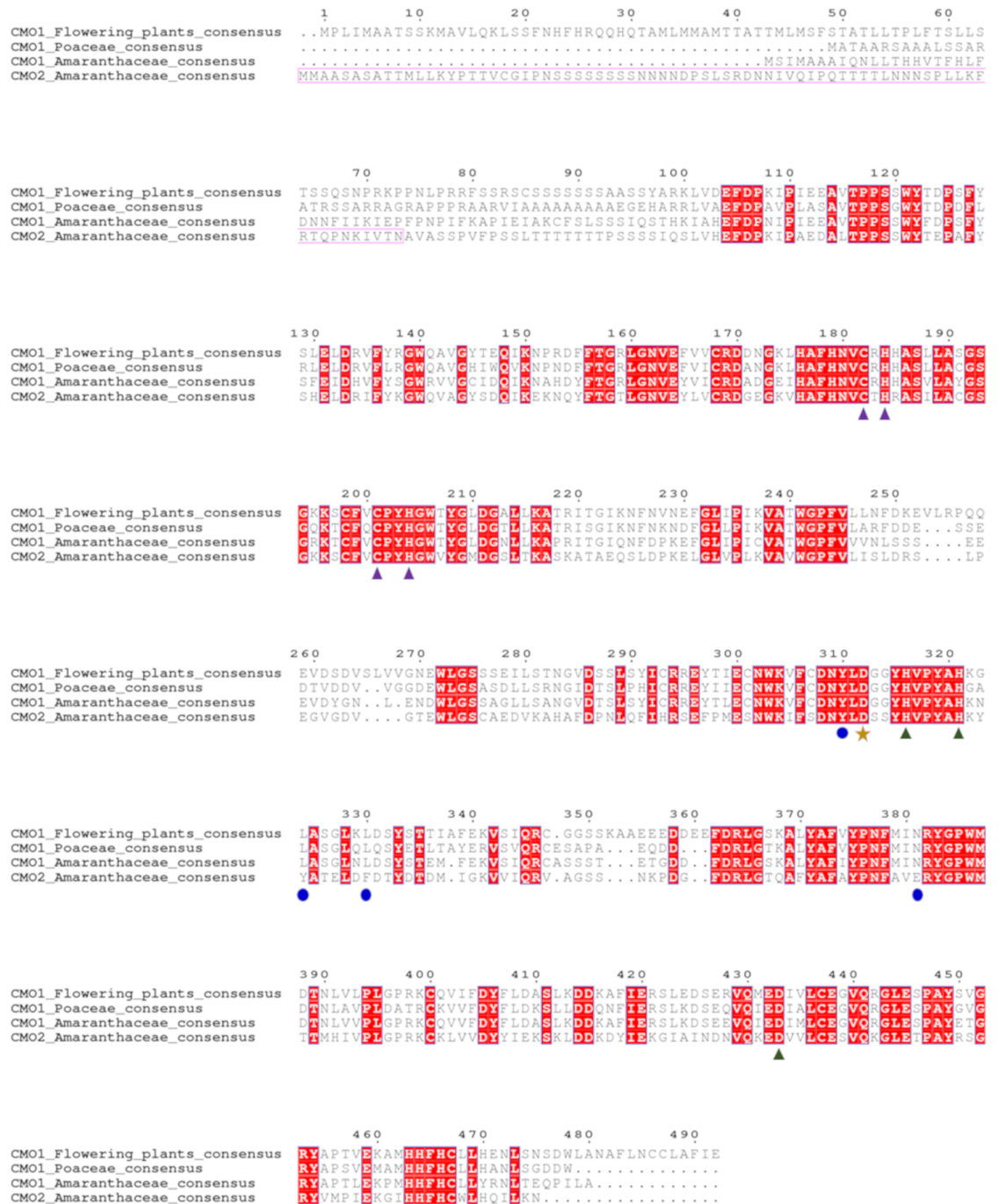


Fig 5. Alignment of CMO consensus sequences of all CMO1 proteins, CMO1 proteins from the Poaceae and CMO1 and CMO2 proteins from the Amaranthaceae. To obtain the consensus sequences we used a cut-off of 60% identity. Residues in red boxes are conserved in at least 60% of all sequences. Residues marked with a purple triangle are those involved in coordinating the iron atoms of the Rieske center; residues marked with green triangles are involved in coordinating the non-heme mononuclear iron; the residue marked with a gold star is the aspartate involved in electron transfer between the Rieske and non-heme mononuclear iron centers; residues corresponding to putative key active site residues involved in choline binding in CMO2 enzymes, as shown below in the docking simulations, are marked with blue circles. The sequence enclosed in a purple frame corresponds to the predicted chloroplast signal peptide in CMO2 enzymes.

<https://doi.org/10.1371/journal.pone.0204711.g005>

The homology models of SoCMO2 and SoCMO1 proteins show important differences between them

Given that to date no three-dimensional structure of any CMO protein has been reported, to analyze structural features that could provide insight into the possible differences in substrate specificity between CMO1 and CMO2 proteins, we constructed and analyzed homology models of the SoCMO2, SoCMO1 and of the CMO1 from *Arabidopsis thaliana* (AtCMO1). We also built topologies for the [2Fe-2S] Rieske and mononuclear non-heme iron centers, which correctly predicted the geometries found in X-ray structures of oxygenase enzymes with the same kinds of iron centers.

We started building the SoCMO2 model because this is the best characterized CMO to date. This model was made without the putative chloroplast peptide signal, which was proposed to consist of 60 residues [14], and without the next segment of 19 residues, which scored as highly mobile in order/disorder predictions [35]. Therefore, the model considers a sequence that is 79 residues shorter than the complete one, i.e., from residues 80 to 439, excluding in this way the highly variable (poorly conserved) region of CMO proteins. From sequence alignment analysis we conclude that the best template to build the SoCMO2 model was the X-ray structure of a bacterial Rieske-monomuclear non-heme dioxygenase from *Rugeria sp.* (PDB: 3N0Q), which has only 24% identity with SoCMO2; other Rieske/mononuclear non-heme oxygenases have even lower identity with SoCMO2. This low degree of identity between the protein model and the protein being modeled made very difficult the construction of a good quality model. Nevertheless, after following the steps described in the Methods section, we obtained a SoCMO2 homology model that complied with the expected functional requirements for a Rieske/mononuclear non-heme oxygenase enzyme. As the first SoCMO2 model obtained with MODELLER was not satisfactory, we improved its quality by repeated MD simulations. A comparison of the final SoCMO2 model with the initial one is shown in Fig 6A and the evolution of the model along the MD simulation, as assessed by the Root Mean Square Deviation (RMSD) per residue between the starting and final models, is shown in Fig 6B.

As an example of CMO1 proteins, and because at the time that we were doing these experiments the SoCMO1 sequence was not yet available, we then followed the same procedure to obtain the *Arabidopsis thaliana* (AtCMO1) homology model using as template the, SoCMO2 model. The AtCMO1 model was constructed from residue 58 to residue 422, eliminating the N-terminal region that contains a putative signal peptide (47 residues) and an additional 11 residues that were predicted to be disordered. When we had a reasonable good AtCMO1 model, we became aware that the SoCMO1 sequence had been reported and so we decided to use this AtCMO1 model as template to make the SoCMO1 3D-structure prediction. For the same reasons as before, only residues 103 to 426 of the complete SoCMO1 sequence (426 residues) were included for the model. Considering this region, AtCMO1 and SoCMO1 share a 63% of sequence identity.

Analyzing the three homology models with the Rd.HMM protocol [42] we found that, out of all sequences in the NCBI-RefSeq database, the SoCMO2 amino acid sequence (XP_021849509), the AtCMO1 amino acid sequence (NP_194718) and the SoCMO1 amino acid sequence (XP_021866412) were the first retrieved when using the SoCMO2, AtCMO1 and SoCMO1 homology models, respectively, indicating that the models indeed can accommodate the respective sequences better than the sequences of any other protein in the NCBI RefSeq database. The results of the validation of the homology models shown in Fig 6C indicate an appropriate backbone for all models, with a Rd.HMM index value just slightly below the value reported for X-ray crystallography determined structures [43]. In addition, the quality of the structural features of the models was tested using the Rosetta-energy scoring

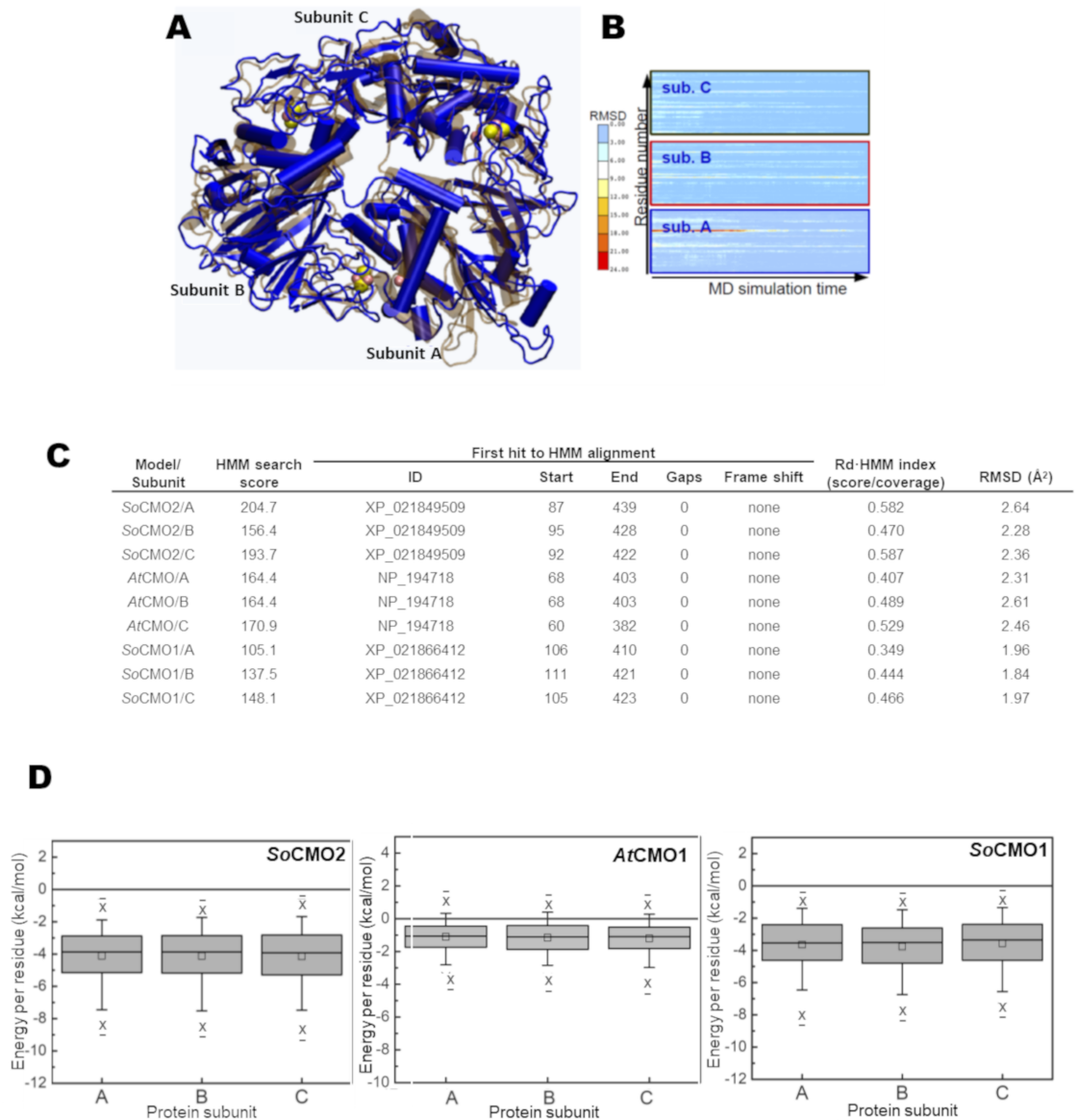


Fig 6. Qualitative evaluation of the final homology models. (A) Three-dimensional predicted structure for SoCMO2 after 200 ns of MD simulations is shown as blue cartoons and the prediction before MD simulations is shown superimposed as other translucent cartoons. Iron and sulfur atoms are shown as orange and yellow spheres, respectively. (B) Heat maps showing the changes in RMSD relative to the final structure per residue of each subunit along MD simulations. The bar plot on the left shows the color range associated with the RMSD values, from red (high) to white (medium) to blue (low). (C) Results of the Rd.HMM analysis of the final SoCMO2, SoCMO1 and AtCMO1 homology models shown per subunit. Note that in this kind of analysis a Rd-HMM index value of 0.6 is obtained for structures solved by X-ray crystallography [42]. The RMSD values given are relative to the bacterial monooxygenase (PDB: 3N0Q) used as template for modeling SoCMO2 and were calculated using STAMP [43] to match equivalent positions in the structures. (D) Box plots of the energy per residue of the SoCMO2, SoCMO1 and AtCMO1 homology models. The evaluation was performed using the Rosetta energy scoring function.

<https://doi.org/10.1371/journal.pone.0204711.g006>

function. The box plots of the energy per residue (Fig 6D) show the absence of abnormally strained, high-energy residues, and that the overall energy of the models was negative.

In the three homology models, the residues involved in iron coordination in the Rieske and mononuclear non-heme centers are properly positioned and the distance between them was similar to the one found in the known crystal structures and the appropriate one for the

transfer of electrons from the Rieske to the mononuclear non-heme iron center, as exemplified in Fig 7A with the SoCMO2 model. Also, the carboxyl group involved in the transfer of electrons between the iron centers in the Rieske/mononuclear non-heme oxygenases [52], which belongs to the side chain of an aspartate in CMO enzymes, was correctly positioned to perform this function. A comparison of the subunit A backbone of the SoCMO2 and SoCMO1 models is shown in Fig 7B.

In the active site of the SoCMO2 model (Fig 7C)—located in the mononuclear non-heme Fe domain as in every Rieske/mononuclear non-heme oxygenase—we found three residues, Tyr281, Tyr295, and Phe301, forming an aromatic box of the kind that has been found in other proteins to be involved in binding a trimethylammonium group, as the one present in choline, through cation- π interactions [16,53–55]. In addition, there was a glutamate residue, Glu346, which could be also involved in an ionic interaction through its side chain carboxyl group with the positively charged trimethylammonium group, as was also found in some choline/acetylcholine binding proteins [55,56]. Indeed, the presence of these aromatic residues and the carboxylate group has been considered a evolutive convergent trait critical for choline binding [56]. Interestingly, except for Tyr281, conserved in most CMO proteins (see below), in the modeled active site of SoCMO1—as well as in that of the AtCMO1 model (not shown)—there are no negatively charged or aromatic residues. In SoCMO1, Glu346 is Asn327, Tyr295 is Leu275, and Phe301 is Leu281 (Fig 7D). These active-site residues changes strongly suggest that both SoCMO1 and AtCMO1 would have a very low affinity for choline, if any. The hydrophobic nature of the SoCMO1 and AtCMO1 active sites, as well as their ample space, suggests that a nonpolar, aromatic compound could be the physiological substrate of these enzymes.

A choline molecule fits well in the active site of SoCMO2

We predicted the possible position and contacts of choline bound into the active site of the SoCMO2 homology model by performing docking simulations, as described in the Methods section. The most populated cluster of choline poses was observed close to the mononuclear non-heme iron center and the three coordinated iron ligands (His287, His292 and Asp398). We selected the best pose (depicted in Fig 7C) by visual inspection and using the orientation of the hydroxyl group of choline relative to the mononuclear non-heme iron as an additional criterion. In the selected pose this hydroxyl group is close to the iron though not near enough to form a coordinated covalent bond, which is not unexpected because the AutoDock Vina forcefield uses molecular mechanics and it cannot predict the formation or breaking of bonds. Interestingly, in this docking pose the quaternary ammonium of choline is inside the aromatic box formed by Tyr281, Tyr295 and Phe301 and near Glu346, as we expected. The three aromatic rings provide an environment rich in electron-density and may participate in cation- π interactions with the trimethylammonium group, whereas the negative charge of the neighbor Glu346 should also compensate the quaternary ammonium electric charge. In addition, this docking pose is consistent not only with the binding of choline but also with the binding of the molecular dioxygen needed for the reaction to take place, since the adduct with choline apparently would not obstruct dioxygen binding. In summary, the docking pose shown in Fig 7C is consistent with what should be expected for a pre-catalytic enzyme-substrate complex for a *bona fide* CMO.

Putative critical active site residues are conserved within the two kinds of CMO enzymes

Interestingly, three of the residues suggested by the SoCMO2 homology model to be in the active site—Tyr295, Phe301 and Glu346 (SoCMO2 numbering)—are conserved in every

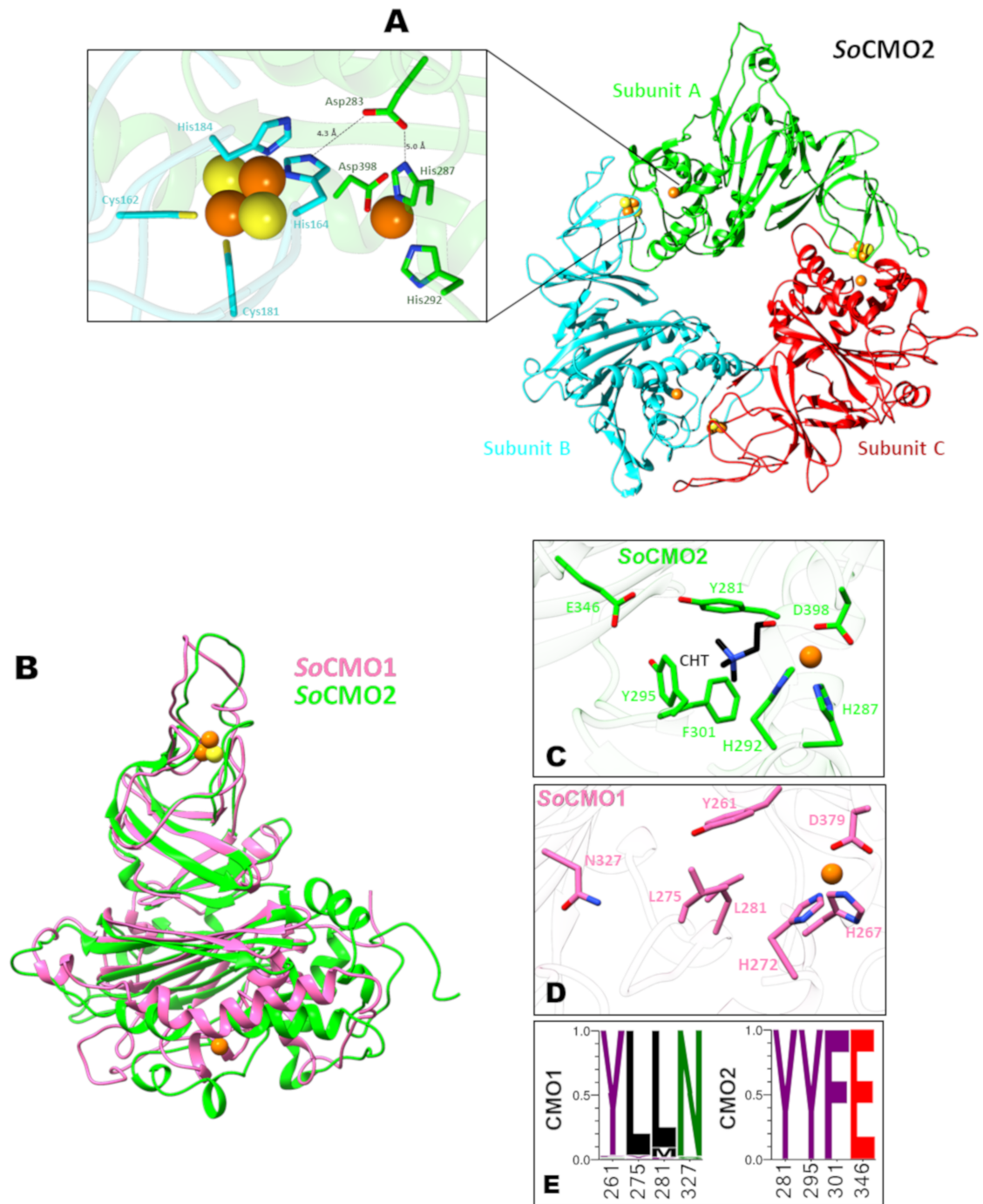


Fig 7. Homology models of SoCMO1 and SoCMO2 proteins. (A) Cartoon representation of the SoCMO2 trimer with each subunit in a different color; the close-up shows the residues involved in iron coordination in the Rieske center (Cys162, His164, Cys181, His184) and in the non-heme mononuclear iron (His287, His292, Asp398), as well as the aspartate residue involved in the transfer of electrons between the two iron centers (Asp283). (B) Cartoon representation of aligned SoCMO1 (pink) and SoCMO2 (green) monomers. In panels A and B, carbon atoms are colored in pink (SoCMO1) or green (SoCMO2), oxygen atoms in red and nitrogen atoms in blue and iron and sulfur atoms are shown as spheres colored orange and yellow, respectively. (C) Cartoon representation of the modeled SoCMO2A active site showing a choline molecule (black carbon atoms) docked into it. The choline trimethylammonium group is in a favorable position to interact with the aromatic box formed by residues Tyr281, Tyr295, Phe301, as well as with the near Asp346 residue. The choline hydroxyl group is oriented towards the iron atom as needed to establish a coordination bond with it. (D) Cartoon representation of the modeled SoCMO1 active site, showing the putative active site residues at position equivalent to the ones in the SoCMO2 model. (E) Sequence logos of selected positions of CMO1 and CMO2 proteins

showing the conservation of residues putatively involved in choline binding in CMO2. Residue conservation is shown as probability. The amino acids color scheme of the logos is according to their chemical properties: polar (S, N), green; aromatic (Y, F), purple; acidic (E), red; and hydrophobic (L, M), black. Residue numbers correspond to the complete SoCMO2 sequence (XP_021849509) or to the complete SoCMO1 sequence (XP_021866412), as appropriate.

<https://doi.org/10.1371/journal.pone.0204711.g007>

known CMO2 sequence, while the residues at equivalent positions in SoCMO1—Leu275, Leu291 and Asn327 (SoCMO1 numbering)—are also highly conserved in all known CMO1 sequences (Fig 7E). Out of the 150 plant sequences included in the CMO1 clade, only in one of the duplicates from *E. guttata*, *S. penelli* and *S. lycopersicum* Leu275 has been changed for a phenylalanine (in *E. guttata*) or tyrosine (in *S. penelli* and *S. lycopersicum*). Also, only in these three enzymes Asn327 has been changed for a serine residue. The conservation of such different active site residues within the two kinds of CMO enzymes strongly suggests differences between them in substrate specificity. In addition to these active-site residues and to those involved in coordinating the mononuclear non-heme iron, there are several other residues totally conserved among the CMO enzymes in both the Rieske and mononuclear non-heme iron domains (Fig 8A). Their location in the SoCMO1 and SoCMO2 homology models is

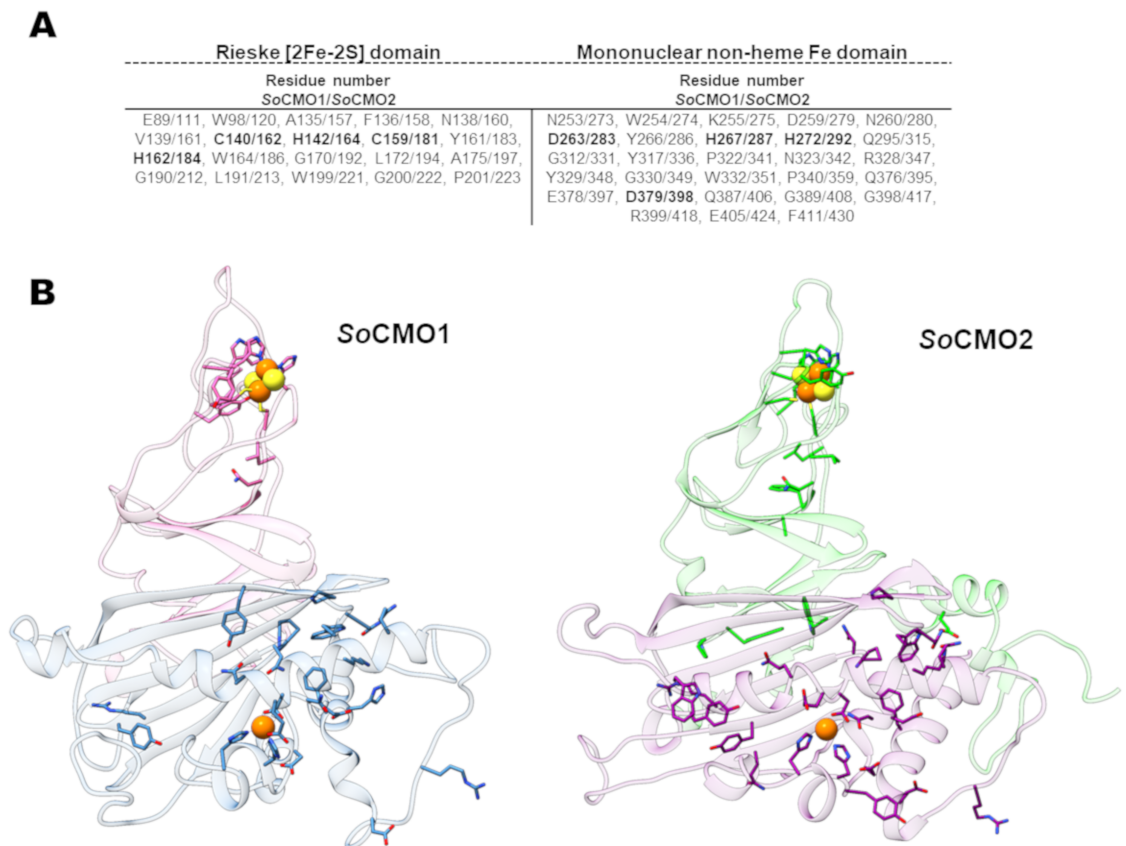


Fig 8. Totally conserved residues in CMO proteins. (A) List of the totally conserved residues in the Rieske [2Fe-2S] and the mononuclear non-heme iron domains of all analyzed CMO proteins. Residues involved in coordination of the Rieske and mononuclear non-heme irons, as well as the aspartate involved in the transfer of electrons between the two iron centers, are in bold. Numbering is according to SoCMO1 and SoCMO2 complete sequences. (B) Cartoon representation of a monomer of the SoCMO1 and SoCMO2 homology models showing the location of the totally conserved residues in the predicted three-dimensional structure. Rieske domain and mononuclear non-heme iron domains are colored pink/blue and green/purple in the SoCMO1 and SoCMO2 monomers, respectively. Residues are shown as sticks, with nitrogen atoms colored in blue, oxygen in red, iron (shown as a sphere) in orange and sulfur (shown as a sphere) in yellow.

<https://doi.org/10.1371/journal.pone.0204711.g008>

depicted in Fig 8B). These conserved residues may be important for activity (particularly, for binding of the Fd protein and/or electron transfer) and/or for the stability of the native trimeric structure. In addition, the active-site tyrosine at position 281 in SoCMO2 is highly conserved in plant CMO proteins; only in one of the CMO1 duplicate of *E. guttata* and the CMO1 of *Capsella rubella* the residue at this position has been changed to phenylalanine. The aromatic nature of this residue at this position is not conserved in other Rieske/mononuclear non-heme oxygenases (results not shown), which suggests that, although it is not needed for the oxygenase activity, it may be important for substrate specificity, not only of the CMO2 but also of the CMO1 enzymes.

Discussion

During evolution of land plants some of them gained the ability to synthesize the osmoprotectant GB, which allowed their growing in very adverse environments where saline soils are prevalent, like in coastal highly saline lands, or in dry or cold climates. This metabolic trait confers a great adaptive advantage not only to halophytes plants, but also to mesophytes, which may experience sporadic episodes of water deficit or freezing temperature. As have been previously shown [16], the change of a single residue in some plant ALDH10 enzymes suffices to confer them BADH activity, which is a *sine qua non* condition for the plant to become a GB-accumulator. But the current experimental evidence obtained with wild-type and transgenic plants suggests that the acquisition of the ability to synthesize GB by evolving a BADH activity should have been accompanied by other adaptations, such as the gain of a significant CMO activity. In the present work we provided phylogenetic and structural evidence of how the CMO activity might have evolved in some of the CMO proteins.

The results of our phylogenetic analysis show that plant CMO proteins are monophyletic and clearly form two clades indicating the existence of two kinds of CMO proteins: CMO1 and CMO2. Limited phylogenetic studies have been previously reported [18,19,57] and one of them suggested the existence of two groups within the plant CMO proteins [18]. By the inclusion in our analysis of a much larger number of sequences (almost ten-fold more) we confirmed and better supported this observation. In addition, we found that the proteins belonging to the clade that we named CMO1 are present in every terrestrial plants having a known genome, regardless of its ability to synthesize and accumulate GB, while proteins of the clade that we named CMO2 were only found in the high GB-accumulator plants of Amaranthaceae species. CMO1 proteins are also present in primitive plants and other unicellular photosynthetic eukaryotes, which suggest that they are the closer descendants from to the ancestral plant CMO. Indeed, CMO2 proteins appear to have evolved from CMO1 as a consequence of a unique event of a *CMO1* gene duplication event that seems to have taken place early in the evolution of the Amaranthaceae family. Although we found other episodes of *CMO* gene duplication in other plants, these events seem to be more recent, as suggested by the few residues that have been changed in the resultant proteins, and probably they did not give rise to new functions. CMO2 proteins clearly evolved at a much higher rate than CMO1 proteins, which led to the formation of a new clade within the CMO subfamily. Of the two CMO protein domains, the mononuclear non-heme domain, where the active site is located, is the one that diverged faster. Moreover, the active site residues that probably are important for determining substrate specificity—as indicated by the homology models and docking simulations—are different in CMO1 and CMO2 enzymes, which strongly suggests functional differences between them. Based on the kind of these active site residues—aromatic and glutamate in CMO2 enzymes *versus* glutamine and non-polar, non-aromatic in CMO1 enzymes—, we think that only CMO2 are *bona fide* choline monooxygenases, *i.e.* their physiological activity is

choline monooxygenase, while most of the CMO1s, including those of the Amaranthaceae species, are not. Indeed, our results hint at the possibility of CMO1s acting on aromatic substrates. It is probable that CMO1s retained the original and so far unknown physiological activity of the ancestor, an activity that seems important for the plant given the conservation of the *CMO* gene in every terrestrial plant of known genome. Interestingly, in the fungal and bacterial CMO-like enzymes the active site residues at the equivalent position of the ones that we propose in this work to be characteristic and possibly determinants of the substrate specificity of the CMO2 enzymes—Tyr295, Phe301 and Glu346 (SoCMO2 numbering)—are changed to non-aromatic, non-charged residues identical or similar to the ones in CMO1 proteins, suggesting either that choline is not the substrate of none of these enzymes or that the reaction that they catalyze is different than the one catalyzed by CMO2 enzymes and requires a different mode of binding of choline at the active site. Unfortunately, none of the plant CMO1 or of the bacterial or fungal CMO-like proteins have been biochemically characterized as yet, so their substrate specificity and the reaction that they catalyze are unknown. The aromatic nature of the tyrosine residue at position 281 (SoCMO2 numbering) is also conserved in the CMO-like proteins, which suggest that this aromatic residue is characteristic of the cd08883 domain and may be important for their general function. Regardless of their specificity, these findings suggest that photosynthetic eukaryotic CMO1 proteins and fungal CMO-like proteins evolved in parallel from an ancestral bacterial CMO-like protein, forming independent protein subfamilies with similar architecture and putative catalytic properties, which will be analyzed in a forthcoming work. On the other hand, it is also possible that CMO1 proteins do have a low CMO activity. Even more, this activity might have been present in the ancestral CMO1 protein. To date, most of the experimentally characterized enzymes exhibiting CMO activity belong to the CMO2 clade: spinach [10,14,58], sugar beet [21], and *A. caudatus* [21]. Only a recombinant CMO1 protein from barley has been assayed for CMO activity in the crude extracts of yeast cells where it was expressed and found that it can oxidize choline [19], but how the CMO activity level of this enzyme compares with that of CMO2 enzymes remains to be evaluated.

Many plants of the Amaranthaceae family are halophytes that tolerate salty and dry soils, and probably the ancestor that gave origin to this plant family also colonized osmotically stressed habitats. The acquisition of a functional CMO together with the acquisition of a functional BADH, which also occurs in the Amaranthaceae family [13], granted them a clear evolutionary advantage through the synthesis of glycine betaine. The evolutionary mechanism that gave rise to a functional CMO activity in these plants likely was the same as the one pointed out by others to explain similar acquisition of new enzymatic activities after a gene duplication event [59]: one of the two resulting *CMO* gene copies was free to undertake multiple mutations, which, under the strong pressure of the osmotically stressed environment, eventually lead to the acquisition and fixation of the CMO activity in these proteins. The low number of known CMO1 sequences from the Amaranthaceae family (just three sequences) contrasts with the higher number of known CMO2 sequences from the same family (17 sequences). This numerical discrepancy can be explained by the few known genomes of Amaranthaceae species to date (only those of the same three species) and by the fact that the *CMO2* gene sequences were obtained from cDNA when the plants were subjected to osmotic stress, indicating that in this family the *CMO2* gene, and not the *CMO1* gene, is the one up-regulated by these adverse environmental conditions. In these studies, the increase in the levels of the *CMO2* gene transcripts and/or CMO2 protein correlated with significant increases in the levels of GB, indicating that this is the protein that exhibits CMO activity and participates in the synthesis of this osmoprotector, at least in the high GB-accumulator species of this plant family. Indeed, osmotic stress response elements for salt,

dehydration and cold were found in the *CMO2* gene promoter of the *Amaranthaceae* halophyte, GB-accumulator plant *Suaeda liaotungensis* [60].

Together, our results prompted us to propose as a hypothesis that only *CMO2* proteins have CMO activity *in vivo* and that the *CMO1* proteins do not. But then the question is, which enzyme produces betaine aldehyde from choline in those GB-accumulators plants that do not have the *CMO2* protein and only have *CMO1*? Antibodies raised against CMO from spinach did not detect any protein in mangrove leaves (*Avicennia marina*) [61], barley (both monocot medium GB-accumulator plants) [19], sunflower and *Arabidopsis* (dicot GB-non-accumulator plants) [22], a finding that was taken as indicative of either significant structural differences between the spinach CMO and the CMO proteins of these plants, or of the absence of a CMO protein in these plants and the presence instead of a choline dehydrogenase (CDH) or choline oxidase protein [22]. However, the CMO protein used to raise the antibody was *SoCMO2*; therefore, their results may be explained by the fact that all the other CMO proteins tested against this antibody were *CMO1* proteins, which, as shown here, have low sequence identity with the *CMO2* proteins. It seems therefore that the likely explanation of these results is the lack of cross-reactivity of the polyclonal anti-*CMO2* antibodies against *CMO1* proteins, due to the important differences between the two kinds of CMO proteins. Regarding the proposal of the possible existence of homologs of CDH or choline oxidase in land plants, we did not find any in pBLAST searches. It may be possible to hypothesize that another, non-yet described enzyme able to convert choline to betaine aldehyde was recruited for GB synthesis. In this respect, it results interesting that a recent report described a plant oxidase named GB1 that greatly increased the levels of GB when overexpressed in maize and soybean plants [62]. We found that GB1 enzymes are present in all plants of known genome (data not shown) and that they belong to the hydroxylase superfamily cl27195 (COG3000). Therefore, it could be speculated that they may participate in a biosynthetic route of GB forming the hydrate species of betaine aldehyde from choline, particularly when overexpressed, although this activity may not be their primary one. Unfortunately, their physiological substrates and the reaction that they catalyze are still unknown. Alternatively, *CMO1*s may have a vestigial CMO activity that could be exploited if they have access to choline in their cellular environment and their protein level and/or activity are up-regulated by osmotic stress. Indeed, this may be the case of the barley *CMO1*, which exhibited CMO activity as mentioned above [19], since the plant possesses an ALDH10 enzyme with proved BADH activity. But since barley *CMO1* has been reported as peroxisomal and barley BADH as cytosolic, the different intracellular location of the two enzymes may impede their sharing of substrates or products.

Supporting information

S1 Fig. Complete alignment of CMO protein sequences.

(PDF)

S2 Fig. Unrooted phylogenetic tree of the complete CMO protein sequences given in S1 Table.

(TIFF)

S1 Table. Proteins identified as members of the CMO subfamily.

(PDF)

Acknowledgments

This work was supported by DGAPA-UNAM PAPIIT grants (IN217814 and IN220317 to RAMC, IN225016 to HRR) and LANCAD-UNAM-DGTIC-215 to RRS. JCC is a student of

the Biochemical Sciences Graduate Program at UNAM and a recipient of a CONACYT doctoral scholarship. HRR was supported by PASPA-DGAPA, UNAM program.

Author Contributions

Conceptualization: Rosario A. Muñoz-Clares.

Formal analysis: Javier Carrillo-Campos, Héctor Riveros-Rosas, Rogelio Rodríguez-Sotres.

Funding acquisition: Rosario A. Muñoz-Clares.

Investigation: Javier Carrillo-Campos, Héctor Riveros-Rosas, Rogelio Rodríguez-Sotres.

Methodology: Héctor Riveros-Rosas, Rogelio Rodríguez-Sotres.

Project administration: Rosario A. Muñoz-Clares.

Supervision: Rosario A. Muñoz-Clares.

Visualization: Javier Carrillo-Campos, Héctor Riveros-Rosas, Rogelio Rodríguez-Sotres, Rosario A. Muñoz-Clares.

Writing – original draft: Rosario A. Muñoz-Clares.

Writing – review & editing: Javier Carrillo-Campos, Héctor Riveros-Rosas, Rogelio Rodríguez-Sotres, Rosario A. Muñoz-Clares.

References

1. Boyer JS. Plant productivity and environment. *Science* 1982; 218: 443–448. <https://doi.org/10.1126/science.218.4571.443> PMID: 17808529
2. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water stress: evolution of osmolyte systems. *Science*. 1982; 217: 1214–1222. <https://doi.org/10.1126/science.7112124> PMID: 7112124
3. Papageorgiou GC, Murata N. The unusual strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving Photosystem II complex. *Photosynth Res*. 1995; 44: 243–252. <https://doi.org/10.1007/BF00048597> PMID: 24307094
4. Courtenay ES, Capp MW, Anderson CF, Record MT. Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants *in vivo* and for the interpretation of “osmotic stress” experiments *in vitro*. *Biochemistry*. 2000; 39: 4455–4471. <https://doi.org/10.1021/bi992887i> PMID: 10757995
5. Rhodes D, Hanson AD. Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Annu Rev Plant Physiol*. 1993; 44: 357–384. <https://doi.org/10.1146/annurev.pp.44.060193.002041>
6. Craig SA. Betaine in human nutrition. *Am J Clin Nutr*. 2004; 80: 539–549. <https://doi.org/10.1093/ajcn/80.3.539> PMID: 15321791
7. Ratriyanto A, Mosenthin R, Bauer E, Eklund M. Metabolic, osmoregulatory and nutritional functions of betaine in monogastric animals. *Asian-Austral J Anim Sci*. 2009; 22: 1461–1476. <https://doi.org/10.5713/ajas.2009.80659>
8. McNeil SD, Nuccio ML, Hanson AD. Betaines and related osmoprotectants: targets for metabolic engineering of stress resistance. *Plant Physiol*. 1999; 120: 945–949. <https://doi.org/10.1104/pp.120.4.945> PMID: 10444077
9. Chen THH, Murata N. Glycinebetaine protects plants against abiotic stress: mechanisms and biotechnological applications. *Plant Cell Environ*. 2011; 34: 1–20. <https://doi.org/10.1111/j.1365-3040.2010.02232.x> PMID: 20946588
10. Brouquisse R, Weigel P, Rhodes D, Yocum CF, Hanson AD. Evidence for a ferredoxin-dependent choline monoxygenase from spinach chloroplast stroma. *Plant Physiol*. 1989; 90: 322–9. <https://doi.org/10.1104/pp.90.1.322> PMID: 16666757
11. Weretilnyk EA, Bednarek S, McCue KF, Rhodes D, Hanson AD. Comparative biochemical and immunological studies of the glycine betaine synthesis pathway in diverse families of dicotyledons. *Planta*. 1989; 178: 342–352. <https://doi.org/10.1007/BF00391862> PMID: 24212901

12. Vasilio V, Bairoch A, Tipton KF, N D. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics*. 1999; 9: 421–434. PMID: [10780262](https://pubmed.ncbi.nlm.nih.gov/10780262/)
13. Muñoz-Clares RA, Riveros-Rosas H, Garza-Ramos G, González-Segura L, Mújica-Jiménez C, Julián-Sánchez A. Exploring the evolutionary route of the acquisition of betaine aldehyde dehydrogenase activity by plant ALDH10 enzymes: implications for the synthesis of the osmoprotectant glycine betaine. *BMC Plant Biol*. 2014; 14: 1–16. <https://doi.org/10.1186/1471-2229-14-1> PMID: [24387633](https://pubmed.ncbi.nlm.nih.gov/24387633/)
14. Rathinasabapathi B, Burnet M, Russell BL, Gage DA, Liao P-C, Nye GJ, et al. Choline monoxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: prosthetic group characterization and cDNA cloning. *Proc Natl Acad Sci*. 1997; 94: 3454–3458. <https://doi.org/10.1073/pnas.94.7.3454> PMID: [9096415](https://pubmed.ncbi.nlm.nih.gov/9096415/)
15. Naylor JF, Fridovich I. The form of the aldehyde susceptible to enzymic oxidation. *J Biol Chem*. 1967; 243: 341–345.
16. Díaz-Sánchez AG, González-Segura L, Mújica-Jiménez C, Rudiño-Piñera E, Montiel C, Martínez-Castilla LP, et al. Amino acid residues critical for the specificity for betaine aldehyde of the plant ALDH10 isoenzyme involved in the synthesis of glycine betaine. *Plant Physiol*. 2012; 158: 1570–1582. <https://doi.org/10.1104/pp.112.194514> PMID: [22345508](https://pubmed.ncbi.nlm.nih.gov/22345508/)
17. Rathinasabapathi B, McCue KF, Gage DA, Hanson AD. Metabolic engineering of glycine betaine synthesis: plant betaine aldehyde dehydrogenases lacking typical transit peptides are targeted to tobacco chloroplasts where they confer betaine aldehyde resistance. *Planta*. 1994; 193: 155–162. <https://doi.org/10.1007/BF00192524> PMID: [7764986](https://pubmed.ncbi.nlm.nih.gov/7764986/)
18. Joseph S, Murphy D, Bhavé M. Glycine betaine biosynthesis in saltbushes (*Atriplex* spp.) under salinity stress. *Biologia*. 2013; 68: 879–895. <https://doi.org/10.2478/s11756-013-0229-8>
19. Mitsuya S, Kuwahara J, Ozaki K, Saeki E, Fujiwara T, Takabe T. Isolation and characterization of a novel peroxisomal choline monoxygenase in barley. *Planta*. 2011; 234: 1215–1226. <https://doi.org/10.1007/s00425-011-1478-9> PMID: [21769646](https://pubmed.ncbi.nlm.nih.gov/21769646/)
20. Luo D, Niu X, Wang Y, Zheng W, Chang L, Wang Q, et al. Functional defect at the rice choline monoxygenase locus from an unusual post-transcriptional processing is associated with the sequence elements of short-direct repeats. *New Phytol*. 2007; 175: 439–447. <https://doi.org/10.1111/j.1469-8137.2007.02124.x> PMID: [17635219](https://pubmed.ncbi.nlm.nih.gov/17635219/)
21. Russell BL, Rathinasabapathi B, Hanson AD. Osmotic stress induces expression of choline monoxygenase in sugar beet and amaranth. *Plant Physiol*. 1998; 116: 859–865. <https://doi.org/10.1104/pp.116.2.859> PMID: [9489025](https://pubmed.ncbi.nlm.nih.gov/9489025/)
22. Hibino T, Waditee R, Araki E, Ishikawa H, Aoki K, Tanaka Y, et al. Functional characterization of choline monoxygenase, an enzyme for betaine synthesis in plants. *J Biol Chem*. 2002; 277: 41352–41360. <https://doi.org/10.1074/jbc.M205965200> PMID: [12192001](https://pubmed.ncbi.nlm.nih.gov/12192001/)
23. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res*. 2013; 41: 29–33. <https://doi.org/10.1093/nar/gkt282> PMID: [23609542](https://pubmed.ncbi.nlm.nih.gov/23609542/)
24. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res*. 2013; 41: 36–42. <https://doi.org/10.1093/nar/gks1195> PMID: [23193287](https://pubmed.ncbi.nlm.nih.gov/23193287/)
25. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012; 40: 1178–1186. <https://doi.org/10.1093/nar/gkr944> PMID: [22110026](https://pubmed.ncbi.nlm.nih.gov/22110026/)
26. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23: 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: [17846036](https://pubmed.ncbi.nlm.nih.gov/17846036/)
27. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999. pp. 95–98. PMID: [10780396](https://pubmed.ncbi.nlm.nih.gov/10780396/)
28. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017; 45: D200–D203. <https://doi.org/10.1093/nar/gkw1129> PMID: [27899674](https://pubmed.ncbi.nlm.nih.gov/27899674/)
29. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. 2006; 7: 1–12. <https://doi.org/10.1186/gb-2006-7-s1-s10> PMID: [16925832](https://pubmed.ncbi.nlm.nih.gov/16925832/)
30. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol*. 2016; 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: [27004904](https://pubmed.ncbi.nlm.nih.gov/27004904/)

31. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 2001; 18: 691–699. <https://doi.org/10.1093/oxfordjournals.molbev.a003851> PMID: 11319253
32. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014; 42: 320–324. <https://doi.org/10.1093/nar/gku316> PMID: 24753421
33. Crooks G, Hon G, Chandonia J, Brenner S. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14: 1188–1190. <https://doi.org/10.1101/gr.849004> PMID: 15173120
34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
35. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Protein structure modeling with MODELLER, In: Kobe B, Guss M, Huber T, editors. *Structural Proteomics: High-throughput Methods.* Humana Press; 2008. pp. 145–159. https://doi.org/10.1007/978-1-60327-058-8_8 PMID: 18542861
36. Dosztányi Z. Prediction of protein disorder based on IUPred. *Protein Sci.* 2018; 27: 331–340. <https://doi.org/10.1002/pro.3334> PMID: 29076577
37. Weinhold F, Carpenter JE. The natural bond orbital Lewis structure concept for molecules, radicals, and radical ions, In: Naaman R, Vager Z, editor. *The Structure of Small Molecules and Ions.* Boston: Springer, 1988. pp. 227–236. https://doi.org/10.1007/978-1-4684-7424-4_24
38. Sousa Da Silva AW, Vranken WF. ACPYPE—antechamber python parser interface. *BMC Res Notes.* 2012; 5: 1–8. <https://doi.org/10.1186/1756-0500-5-1> PMID: 22214347
39. Case D.A., Betz R.M., Cerutti D.S., Cheatham T.E. III, Darden T.A., Duke R.E., et al. Amber 2016. University of California, San Francisco; 2016.
40. Mitra D, George SJ, Guo Y, Kamali S, Keable S, Peters JW, et al. Characterization of [4Fe-4S] cluster vibrations and structure in nitrogenase Fe protein at three oxidation levels via combined NRVS, EXAFS, and DFT analyses. *J Am Chem Soc.* 2013; 135: 2530–2543.
41. Carvalho ATP, Swart M. Electronic structure investigation and parametrization of biologically relevant iron-sulfur clusters. *J Chem Inf Model.* 2014; 54: 613–620.
42. Edler E, Stein M. Spin-state-dependent properties of an iron(III) hydrogenase mimic. *Eur J Inorg Chem.* 2014; 2014: 3587–3599. <https://doi.org/10.1002/ejic.201402295>
43. Martínez-Castilla LP, Rodríguez-Sotres R. A score of the ability of a three-dimensional protein model to retrieve its own sequence as a quantitative measure of its quality and appropriateness. *PLoS One.* 2010; 5: e12483 <https://doi.org/10.1371/journal.pone.0012483> PMID: 20830209.
44. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996. 14: 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
45. Trott O, Olson A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.* 2010; 31: 455–461. <https://doi.org/10.1002/jcc.21334> PMID: 19499576
46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25: 1605–1612. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
47. Schmidt CL, Shaw L. A comprehensive phylogenetic analysis of rieske and rieske-type iron-sulfur proteins. *J Bioenerg Biomembr.* 2001; 33: 9–26. <https://doi.org/10.1023/A:1005616505962> PMID: 11460929
48. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, et al. Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* 2013; 14: 1–14. <https://doi.org/10.1186/gb-2013-14-2-r11> PMID: 23375108
49. Linder T. CMO1 encodes a putative choline monoxygenase and is required for the utilization of choline as the sole nitrogen source in the yeast *Scheffersomyces stipitis* (syn. *Pichia stipitis*). *Microbiology (United Kingdom).* 2014; 160: 929–940. <https://doi.org/10.1099/mic.0.073932-0> PMID: 24608175
50. Richardson AO, Palmer JD. Horizontal gene transfer in plants. *J Exp Bot.* 2007; 58: 1–9. <https://doi.org/10.1093/jxb/erl148> PMID: 17030541
51. Bock R. The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* 2010; 15: 11–22. <https://doi.org/10.1016/j.tplants.2009.10.001> PMID: 19910236
52. Parales RE, Parales J V., Gibson DT. Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity. *J Bacteriol.* 1999; 181: 1831–1837. PMID: 10074076
53. Schiefner A, Breed J, Bösser L, Kneip S, Gade J, Holtmann G, et al. Cation- π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic

- ligand-binding protein ProX from *Escherichia coli*. J Biol Chem. 2004; 279: 5588–5596. <https://doi.org/10.1074/jbc.M309771200> PMID: 14612446
54. Horn C, Sohn-Bösser L, Breed J, Welte W, Schmitt L, Bremer E. Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from *Bacillus subtilis* for the compatible solutes glycine betaine and proline betaine. J Mol Biol. 2006; 357: 592–606. <https://doi.org/10.1016/j.jmb.2005.12.085> PMID: 16445940
 55. Oswald C, Smits SHJ, Höing M, Sohn-Bösser L, Dupont L, Le Rudulier D, et al. Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. J Biol Chem. 2008; 283: 32848–32859. <https://doi.org/10.1074/jbc.M806021200> PMID: 18779321
 56. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, et al. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature 2012; 482: 547–551. <https://doi.org/10.1038/nature10753> PMID: 22278061
 57. Park J, Okita TW, Edwards GE. Salt tolerant mechanisms in single-cell C₄ species *Bienertia sinuspersici* and *Suaeda aralocaspica* (Chenopodiaceae). Plant Sci. 2009; 176: 616–626. <https://doi.org/10.1016/j.plantsci.2009.01.014>
 58. Burnet M, Lafontaine PJ, Hanson AD. Assay, purification, and partial characterization of choline monoxygenase from spinach. Plant Physiol. 1995; 108: 581–588. <https://doi.org/10.1104/pp.108.2.581> PMID: 12228495
 59. James LC, Tawfik DS. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. Trends Biochem Sci. 2003; 28: 361–368. [https://doi.org/10.1016/S0968-0004\(03\)00135-X](https://doi.org/10.1016/S0968-0004(03)00135-X) PMID: 12878003
 60. Li Q, Yin H, Li D, Zhu H, Zhang Y, Zhu W. Isolation and characterization of CMO gene promoter from halophyte *Suaeda liaotungensis* K. J Genet Genomics. 2007; 34: 355–361. [https://doi.org/10.1016/S1673-8527\(07\)60038-1](https://doi.org/10.1016/S1673-8527(07)60038-1) PMID: 17498634
 61. Hibino T, Meng YL, Kawamitsu Y, Uehara N, Matsuda N, Tanaka Y, et al. Molecular cloning and functional characterization of two kinds of betaine-aldehyde dehydrogenase in betaine-accumulating mangrove *Avicennia marina* (Forsk.) Vierh. Plant Mol Biol. 2001; 45: 353–363. <https://doi.org/10.1023/A:1006497113323> PMID: 11292080
 62. Castiglioni P, Bell E, Lund A, Rosenberg AF, Galligan M, Hinchey BS, et al. Identification of GB1, a gene whose constitutive overexpression increases glycine betaine content in maize and soybean. Plant Direct. 2018; 2018: 1–7. <https://doi.org/10.1002/pld3.40>