



Tesis de Maestría en Bioinformática



Caracterización *in silico* de dioxigenasas responsables del clivaje de carotenoides de especies de citrus

Ing. Jorge R. Cantero Piñanez

Directores: Prof. Margot Paulino Zunini Prof. Fabio Polticelli

Montevideo - Uruguay - 2021

TRIBUNAL

Dr. Antonio Jesús Meléndez-Martínez Laboratorio de Color y Calidad de Alimentos. Facultad de Farmacia – Universidad de Sevilla, España

Dr. Sergio Pantano

Laboratorio de simulaciones biomoleculares Institut Pasteur de Montevideo, Uruguay

Dr. Federico Iribarne

Departamento de Experimentación y Teoría de la Estructura de la Materia Facultad de Química – Universidad de la República, Uruguay

TABLA DE CONTENIDOS

Capítulo I. INTRODUCCIÓN

Los carotenoides	1
Estructura y clasificación	2
Biosíntesis de carotenoides	4
Biosíntesis de apocarotenoides	6
Propiedades de los carotenoides	6
Dioxigenasas de clivaje de carotenoides (CCDs)	7
Subfamilias CCD1 y CCD2	9
Subfamilias CCD7 y CCD8	9
Subfamilias CCD4	10
Caracterización funcional de las CCD4	11
Características de las CCD4 de especies de citrus	13
Importancia de los cítricos en el Uruguay	15
Referencias	15

Capítulo II. ANÁLISIS DE SECUENCIAS DE LAS ENZIMAS CCD4 DE *Citrus clementina*

Introducción	22
Objetivos	24
Materiales y Métodos	24
Identificación y caracterización de CCDs de Citrus clementina	24
Alineamiento de las secuencias	25
Predicción de la estructura secundaria	25
Base de datos de secuencias CCDs	25
Predicción de motivos	26
Análisis de variabilidad de las secuencias	26
Reconstrucción filogenética	28
Resultados y discusión	28
Identificación y caracterización de CCDs de Citrus clementina	28
Predicción de motivos conservados	31
Predicción del desorden de las secuencias	32
Predicción de la estructura secundaria	33
Alineamiento y análisis de la conservación de las secuencias CCD4	34
Análisis de la variabilidad de las secuencias	36
Análisis filogenético	37
Conclusiones	40
Referencias	41

TABLA DE CONTENIDOS

Capítulo III. IMODELADO MOLECULAR Y SIMULACIÓN DE LA CCD4a Y CCD4b DE Citrus clementina

Introducción	56
Objetivos	50 60
Materiales y Métodos	60
Software, hardware, campo de fuerza y parametrización	60
Modelado por homología de las proteínas	60
Refinamiento y dinámica molecular	61
Procedimiento de <i>docking</i> : Base de datos de ligandos	62
Validación del docking	62
Dinámica molecular de los complejos	64
Medidas de la energía de interacción	64
Entorno de interacciones del ligando	64
Resultados y discusión	66
Alineamiento de secuencias y selección del template	66
Modelado molecular	67
Resultados observados en el docking	70
Esfera de coordinación del Fe ⁺² y la influencia en la reactividad de las	70
enzimas CCD4	76
Conclusiones	77
Referencias	79

Capítulo IV. CCD4 DE Citrus clementina Y SU INTERACCIÓN CON MEMBRANAS

Introducción	82
Objetivos	85
Materiales y Métodos	85
Software, hardware, campo de fuerza y parametrización	85
Construcción de modelos 3D de las proteínas	86
Refinamiento y dinámica molecular	87
Construcción de los complejos CCD4-ligandos	88
Dinámica molecular de los complejos	89
Construcción de los complejos en membrana	89
Energía de interacción del ligando (Uab)	90
Interacciones proteína-membrana	91
Influencia de la membrana en la movilidad de residuos	91
Penetración en la membrana	93
Estimación de la energía libre de unión (MMPBSA)	94

TABLA DE CONTENIDOS

Capítulo IV. CCD4 DE Citrus clementina Y SU INTERACCIÓN CON MEMBRANAS

Resultados y Discusión	96
Inserción del dominio α-hélice en las membranas	98
Análisis energético y equilibrio termodinámico	102
Análisis estructural (RMSD y RMSF)	104
Distancias al Fe ⁺² y energías libres de unión	106
Análisis de la interacción proteína – membrana	109
Conclusiones	113
Referencias	114

Capítulo V. CONCLUSIONES Y PERSPECTIVAS

117
120
121

ANEXO A1 – Artículo completo.

"An In Silico Study of the Citrus Dioxygenases CCD4 family Substrates". *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–28. (2018). Vega-Teijido, M., **Cantero, J.,** Rodrigo, M. J., López, C., & Paulino Zunini, M. https://doi.org/10.1080/07391102.2018.1477619

A2 – Manuscrito en revisión

"Citrus carotenoids cleavage dioxygenases type CCD4 family and their interaction with membranes". *International Journal of Molecular Sciences* (IJMS). Autores: **Jorge Cantero**, Fabio Polticelli y Margot Paulino. ID del manuscrito: ijms-1185032. Estado: en revisión.

A3 – Estructura de carotenoides utilizadas en esta tesis

CAPÍTULO I INTRODUCCIÓN

CAROTENOIDES

Los carotenoides son pigmentos isoprenoides o terpenos que abundan en las membranas de todos los organismos fotótrofos y la mayoría de los heterótrofos. La fórmula empírica del β -caroteno C₄₀H₅₆ se estableció por primera vez en 1907 por Willstatter y Mieg (Johnson, 2002) y su estructura química fue dilucidada por Paul Karrer en 1931. Fue la primera provitamina establecida y por este trabajo recibió el premio Nobel de Química en 1937 (Karrer, 1966). Estos compuestos adquieren su nombre del mayor miembro de este grupo, el β -caroteno (Ahmad et al., 2019), una molécula abundante en las zanahorias cuyo nombre científico es *Daucus carota*.

La abundancia y diversidad de los carotenoides en los vegetales obedecen a las múltiples funciones y propiedades que derivan de su estructura química. La mayoría de los animales no pueden sintetizarlos por lo que deben incluirlos en su dieta (Nisar et al., 2015). Como excepción algunos áfidos son capaces de sintetizarlos. Estudios han demostrado que esta capacidad deriva de un evento de transferencia horizontal de los genes implicados en la biosíntesis de carotenoides provenientes de bacterias y hongos (Zhao y Nabity, 2017).

ESTRUCTURA Y CLASIFICACIÓN

La estructura elemental de los carotenos se constituye a partir de una cadena de 8 unidades de isoprenos con un total de 40 carbonos (C_{40}) (Meléndez-Martínez et al., 2007). La modificación de esta estructura básica genera toda la variedad estructural que conocemos de los carotenoides (Yabuzaki, 2017). Los carotenos son sustancias hidrocarbonadas puras, cuentan solamente con átomos de carbono e hidrógeno (Fig. 1) mientras que aquellas que contienen átomos de oxígeno se conocen como xantofilas (Rodriguez-Amaya, 2015) (Fig. 2a-b).





Figura 1. Sistema de numeración y estructura de los carotenos clasificados según su estructura. Los números representan al sistema estándar de numeración del esqueleto de carbonos en carotenoides. a) Caroteno lineal correspondiente al licopeno. b) Caroteno monocíclico correspondiente al γ-caroteno. c) Caroteno bicíclico correspondiente al βcaroteno.

Los grupos oxigenados pueden ser muy variados. Entre los más comunes se encuentran las cetonas, alcoholes, aldehídos, metoxi, carbometoxi (Fraser y Bramley, 2004). Según su estructura los podemos clasificar en lineales (Fig. 1a), monocíclicos (Fig. 1b) y bicíclicos (Fig. 1c). A los productos derivados de los carotenoides, ya sean carotenos o xantofilas, poseen menos de 40 carbonos se los denomina apocarotenoides (Beltran y Stange, 2016) (Fig. 2c)



Figura 2. Estructura de las xantófilas más comunes. a) Grupo correspondiente a las xantófilas que incluye a las especies de oxígeno en forma de OH representados por zeaxantina y luteína. b) Grupo que incluye a las especies de oxígeno en forma de epoxi, representados por anteraxantina y violaxantina. c) Estructura de apocarotenoides, carotenoides con menos de 40 átomos de carbonos, representados por β-ionona, retinal y β-citraurina.

BIOSÍNTESIS DE CAROTENOIDES

En los organismos fotosintetizadores los carotenoides se sintetizan en los cloroplastos, que se localizan en las membranas tilacoidales (Sun et al., 2018). La biosíntesis en las plantas superiores se inicia con la unión de dos moléculas de geranil geranil difosfato (GGPP) mediadas por la enzima fitoeno sintasa (PSY – EC 2.5.1.32) para dar lugar al fitoeno, posteriormente es desaturado por la fitoeno desaturasa (PDS – EC 1.3.99.31) para generar ζ -caroteno. Seguidamente el ζ -caroteno da lugar al licopeno mediante la ζ -caroteno desaturasa (ZDS – EC 1.3.5.6). El licopeno sufre una acción de ciclación y es aquí en donde el camino se bifurca según sea la enzima que actúe: la ε -licopeno ciclasa (ε -LYC – EC 5.5.1.18) o la β -licopeno ciclasa (β -LYC – EC 5.5.1.19). La ε -LYC genera α -caroteno mientras que la β -LYC genera β -caroteno. Ambas moléculas pueden hidroxilarse por enzimas específicas y generar moléculas de xantofilas: El α -caroteno puede sufrir la acción de la ε -caroteno es hidroxilado por la β -caroteno hidroxilasa (β -CHX – EC 1.14.15.24) para generar β -criptoxantina. La α -criptoxantina a su vez puede volver a hidroxilarse por acción de la β -CHX para dar paso a la luteína mientras que la β -criptoxantina genera zeaxantina (Ikoma et al., 2016). (Fig. 3)

La zeaxantina puede seguir sufriendo modificaciones de carácter reversible para generar epoxicarotenoides como la anteraxantina y violaxantina. Ese conjunto de reacciones se conoce como ciclo de las xantófilas y constituyen un importante factor de protección frente al estrés oxidativo en el aparato fotosintético de las plantas (Latowski et al., 2011).

Los carotenoides se acumulan en los cloroplastos en los frutos inmaduros, y a medida que los frutos maduran los cloroplastos pierden gran parte de su aparato fotosintético y se convierten en cromoplastos, la coloración que adquieren los mismos se debe a la acumulación y tipo de pigmentos carotenoides preexistentes (Li et al., 2016).

Todos los genes carotenogénicos en plantas vasculares son nucleares (Baranski y Cazzonelli, 2016) y poseen, al igual que las enzimas de clivaje de carotenoides (CCDs), en el extremo N-terminal un péptido de destinación al cloroplasto (cpTP) donde tienen lugar las reacciones de biosíntesis y clivaje (Majer et al., 2017) (Shumskaya y Wurtzel, 2013). A pesar de

no ser proteínas de transmembrana, interactúan con ella a través de una región hidrófóbica que consigue penetrar en la membrana, desde donde toma directamente sus sustratos (Wurtzel, 2019).



Biosíntesis de carotenoides

Figura 3. Ruta de biosíntesis de los carotenoides en plantas superiores. GGPP (geranil geranil difosfato). En recuadro se indican las enzimas que catalizan las reacciones: PSY (fitoeno sintasa), PDS (fitoeno desaturasa), ZDS (ζ-caroteno desaturasa), ε-LYC (ε-licopeno ciclasa), β-LYC (β-licopeno ciclasa), β-CHX (β-caroteno hidroxilasa), (ε-CHX) ε- caroteno hidroxilasa. Se indican los números EC (*Enzyme Commission numbers*) de acuerdo a la reacción química que catalizan.

BIOSÍNTESIS DE APOCAROTENOIDES

La biosíntesis de apocarotenoides se da por la actividad de las enzimas CCDs (Hou et al., 2016), estas producen un corte oxidativo de la molécula de carotenoide (C₄₀) en las posiciones de doble enlace generando fragmentos más pequeños (Harrison y Bugg, 2014) (Fig. 4). Debido al gran número de posibles sitios de cortes se generan moléculas con distintos tamaños y propiedades. Las moléculas más pequeñas suelen ser moléculas volátiles con propiedades fragantes como las iononas y damascenonas, mientras que algunas de ellas son precursoras de fitohormonas como el ácido abscísico (ABA) (Auldridge et al., 2006) y la estrigolactona (Waters et al., 2017). Las moléculas de mayor tamaño suelen volver a clivarse en fragmentos más pequeños por las enzimas de clivaje de apocarotenoides (ACOs) (Sui et al., 2013) o acumularse en las membranas tilacoidales y ser partes constitutivas de los cromoplastos (Li et al., 2016).

PROPIEDADES

Desde el punto de vista fisicoquímico la presencia de dobles enlaces conjugados convierte a los carotenoides en cromatóforos que absorbe luz en el rango de espectro UV-visible(Meléndez-Martínez et al., 2007). Debido a su capacidad de absorber energía luminosa y fotoprotectora son indispensables en la fotosíntesis (Hashimoto et al., 2016). La característica poliénica habilita a estos compuestos a aceptar electrones de especies reactivas y neutralizar radicales libres, lo que los convierte en antioxidantes (Johnson, 2002).

En los mamíferos son importantes precursores de la vitamina A (retinal) y el ácido retinoico que juegan un papel muy importante en la nutrición (Johnson, 2002), la visión (von Lintig, 2012) y la diferenciación celular (Wang et al., 2015), mientras que en las plantas constituyen un importante rol en la síntesis de las moléculas hormonales tales como el ABA (Parry et al., 1990), que interviene en las respuestas de estrés fisiológico y las estrigolactonas responsables de la ramificación lateral en plantas (Jia et al., 2018).

Algunos carotenoides intervienen como protectores de los complejos antena de los fotosistemas y frente a distintos tipos de estrés (Hashimoto et al., 2016). En frutas y flores la presencia de los carotenoides sirve para dar color y aroma que permite atraer a los polinizadores y organismos que ayudan a la dispersión de semillas (Cazzonelli, 2011).

DIOXIGENASAS DE CLIVAJE DE CAROTENOIDES (CCDs)

Las CCDs constituyen una gran familia de enzimas dependientes de hierro no hémico que participan en la oxidación de carotenoides o apocarotenoides, y se encuentran ampliamente difundidas en todos los organismos vivos(Akemi Ohmiya, 2009). Estas enzimas son capaces de realizar un corte oxidativo en los sitios de doble enlace C=C de los carotenoides (Harrison y Bugg, 2014) (Fig. 4).



Figura 4. Posible mecanismo de reacción de clivaje oxidativo catalizada por las CCDs. El centro catalítico está coordinado por cuatro anillos de imidazol correspondientes a residuos de histidinas conservadas en las CCDs. El Fe⁺² activa al O₂ para iniciar el ataque del doble enlace del carotenoide. En la reacción se forma dioxietano, un intermediario inestable, que rápidamente deviene en los productos finales.

Las características de estas reacciones no están del todo claras para cada caso particular, ya que se desconocen las características estructurales de muchos miembros de esta familia de enzimas, por lo que constituye un campo activo en la investigación actual (Ahrazem et al., 2016).

Las CCDs suelen ser muy específicas en cuanto al sitio de corte y selectivas en cuanto al tipo de carotenoides que aceptan como sustrato. Todas poseen un sitio activo fuertemente conservado formado por cuatro histidinas en coordinación con el hierro y un alto grado de similitud estructural a lo largo de toda la familia (Kloer y Schulz, 2006). La gran mayoría se encuentran asociadas a las membranas tilacoidales en el interior de los cloroplastos. Las funciones de los productos del clivaje son aún poco conocidos (Ahrazem et al., 2016).

La primera CCD cuya estructura fue dilucidada fue la oxigenasa de clivaje de apocarotenoides (ACO – <u>EC 1.13.11.75</u>) de *Synechocystis sp.* expresada en *Escherichia coli* en el año 2015 (PDB id: <u>2BIW</u>), la misma realiza un corte en la posición 15=15' (Kloer et al., 2005) (Fig. 5).



Figura 5. Representación esquemática de la ACO. a) Estructura secundaria: Se observan 7 hojas que forman el β -propeller. Los puntos indican la posición de las histidinas conservadas que coordinan con el Fe⁺² en el centro catalítico. Los cuatro *loops* y el dominio N-terminal resaltados en gris oscuro conforman el domo hidrofóbico ubicado en la parte superior del sitio activo que interactúa con la membrana tilacoidal. b) Las ACOs solamente aceptan apocarotenoides. Se representa el túnel hidrofóbico con el centro catalítico formado por el Fe⁺². También se observa el domo con el área hidrofóbica de interacción con la membrana y el canal de salida del producto de clivaje.

En las plantas se encuentran los grupos más diversos de CCDs (Ahrazem et al., 2016). Las CCDs están clasificadas en dos grandes familias en función a su capacidad o no de producir ABA, una hormona implicada en la respuesta a condiciones de estrés y en el letargo de las semillas (Rodrigo et al., 2013). Las involucradas en la síntesis del ABA se denominan dioxigenasas de 9cis-epoxicarotenoides (NCEDs) (Messing et al., 2010), y fue la primera familia caracterizada (Tan et al., 1997), que clivan selectivamente 9-cis-violaxantina y 9-cis-neoxantina en la posición C11=C12 produciendo xantoxina (Schwartz et al., 1997), precursor del ABA. Las demás subfamilias están compuestas por cinco miembros CCD1, 2, 4, 7 y 8 (Ohmiya, 2009). Muchas de las cuales aún no cuentan con funciones totalmente conocidas (Ahrazem et al., 2016).

SUBFAMILIAS CCD1 Y CCD2

La CCD1 de la *Arabidopsis thaliana* fue la segunda CCD en ser aislada en las plantas (Schwartz et al., 2001). Su actividad enzimática ocurre de forma simétrica en la posición C9=C10(C9'=C10'), sobre una larga lista de substratos: ζ -caroteno, licopeno, fitoflueno, β -caroteno, δ -caroteno, zeaxantina, luteína, violaxantina y algunos apocarotenoides (Ilg et al., 2009)[.] (Tian et al., 2015). Todos sus sustratos requieren la configuración *all-trans* (Schwartz et al., 2001). Del producto del clivaje surgen compuestos volátiles como la geranilacetona, pseudoionona y β -ionona (Kloer y Schulz, 2006). La CCD1 está involucrada en la ruta de producción de las estrigolactonas (Floss y Walter, 2009), la última hormona vegetal descubierta e involucrada en la ramificación y crecimiento de las yemas axilares y la simbiosis entre las raíces de las plantas y los microorganismos del suelo (Zwanenburg et al., 2016).

La CCD2 es una subfamilia cercanamente emparentada con la CCD1, y está asociada a la formación de crocetina que se acumula en altas concentraciones en los estigmas de las flores, que se produce por el clivaje en la posición C7=C8 y C7'=C8' de la zeaxantina. La CCD2 también es capaz de clivar luteína y 3-OH- β -apocarotenal en la posición C7=C8 (Ahrazem et al., 2016).

SUBFAMILIAS CCD7 Y CCD8

Las enzimas CCD7 y CCD8 se encuentran asociados a procesos del desarrollo vegetal, específicamente en las ramificaciones laterales (Alder et al., 2012). La CCD7 es capaz de catalizar las reacciones en C9'=C10' del β -caroteno para producir 10'-apo- β -carotenal (C₂₇) y β -ionona (C₁₃) (Bruno et al., 2014). La CCD8 es capaz de catalizar el corte secundario del apocarotenoide

10'-apo- β -carotenal en la posición C13=C14 para producir 13-apo- β -carotenona (C₁₈) (Liang et al., 2018). A diferencia de la CCD1 que se encuentra alojada en el citoplasma (Floss y Walter, 2009), la CCD7 y la CCD8 se encuentran localizados en los plastidios (Auldridge et al., 2006), al igual que la CCD4 y la CCD2 y que son los lugares preferidos para los carotenoides (Ahrazem et al., 2016).

Las CCD7 y CCD8 están también implicadas en la ruta de producción de la estrigolactonas (Zwanenburg et al., 2016), que gobiernan el crecimiento de las ramas laterales, la asociación de microorganismos con la raíz de las plantas superiores y la germinación de plantas parásitas (Mishra et al., 2017).

SUBFAMILIA CCD4

La subfamilia CCD4 se encuentra en todas las plantas con flores (Ahrazem et al., 2010). Sus funciones están directamente relacionadas a la coloración de frutos y flores y a la producción de sustancias con propiedades fragantes (Huang et al., 2009). Están presentes en hojas, flores, raíces, tallos y frutos (Rodrigo et al., 2013).

El número de genes de la subfamilia CCD4 presente en los individuos varía a lo largo de distintas especies (Ahrazem et al., 2010). La mayoría cuenta con al menos dos genes del tipo CCD4 (Vallabhaneni et al., 2010). En algunos casos algunos miembros se expresan específicamente en algunos órganos (Rodrigo et al., 2013).

En *Chrysantemum morifolium* la CCD4a está asociada a la ausencia de acumulación de carotenoides en los pétalos de sus flores produciendo el color blanco típico en ellas (Ohmiya, 2013). En algunas especies la expresión se da por situaciones de estrés abiótico.

En *Crocus sativus*, se ha observado un aumento en la expresión de la CCD4a y b en las hojas ante las condiciones de estrés hídrico y calor (Rubio et al., 2008); mientras que la CCD4c se sobre expresa en el estigma de las flores ante la presencia de heridas y condiciones de estrés osmótico, frío y calor (Rubio-Moraga et al., 2014).

Los miembros de esta familia poseen al inicio un péptido de destinación al cloroplasto (cpTP) en la región N-terminal (Rodrigo et al., 2013). Se encuentran alojados en el interior de los cloroplastos junto con otras enzimas de la biosíntesis de carotenoides que permite obtener sus sustratos de forma directa (Messing et al., 2010). En el interior de los cloroplastos se constituyen en unas de las proteínas más abundantes, lo que permite sugerir su importancia en las rutas de síntesis de carotenoides en las plantas (Ytterberg et al., 2006).

CARACTERIZACIÓN FUNCIONAL DE LAS CCD4

Los miembros de la subfamilia CCD4 no están aún lo suficientemente caracterizados. El papel biológico de la mayoría de las CCD4 de plantas sigue siendo aún desconocido. Un creciente número de estudios han señalado la relación entre los genes CCD4 y el color de flores y frutas (Rodrigo et al., 2013; Zhang et al., 2015; Ohmiya et al., 2006).

A diferencia de otras subfamilias, las CCD4 son bastante variables en cuanto a la preferencia por substratos (Tabla 1). Su caracterización se realiza a menudo mediante tres estrategias:

- 1) Ensayos *in vivo* en plantas mutantes con deficiencia/sobreexpresión en CCD4.
- Ensayos *in vivo* en bacterias que sobre acumulan distintos tipos de carotenoides y que coexpresan CCD4
- 3) Ensayos *in vitro* de enzimas CCD4 recombinantes

A modo de resumen, en la Tabla 1 se observa que en general las CCD4 de origen vegetal prefieren como sustrato al β -caroteno y la posición de clivaje en el doble enlace C9=C10 formando como productos β -ionona y 10'-apo- β -carotenal (Huang et al., 2009).

La CCD4 proveniente de *Arabidopsis thaliana* es una las pocas en las que se observó la capacidad de clivar epoxy y apocarotenoides (Latari et al., 2015)

En los cítricos se ha encontrado evidencias en ensayos *in vitro* que la CCD4b1 posee la capacidad de clivar en la posición C7=C8 o C7'=C8', recordando la actividad de las CCD2 (Rodrigo et al., 2013). Aún se desconoce si esta característica se encuentra extendida a los demás miembros de la subfamilia CCD4 presentes en especies de citrus.

Tabla 1.Caracterización funcional de las enzimas CCD4. Los ensayos *in planta* se realizan en
mutantes con deficiencia/sobreexpresión en CCD4; *in vivo* en bacterias que sobre
acumulan distintos tipos de carotenoides y que coexpresan CCD4; *in vitro* utilizando
enzimas CCD4 recombinantes.

Organismo	Gen	Ensayo	Sustrato	Posición de clivado	Productos	Ref.	
Arabidopsis	0004	in planta	epoxy-β-xantofilas	C9=C10 o C9'=C10'	C13	(Latari et al., 2015)	
thaliana	CCD4	in vitro	8'-apo-β-carotenal	C9=C10	β-ionona	(Huang et al., 2009)	
Brassica napus	CCD4 C3	in planta	α-caroteno	C9=C10	α-ionona	(Zhang et al.,	
		I	δ-caroteno			2015b)	
Chrysantemum	CCD4a	in vitro	β-caroteno	C9=C10 v C9'=C10'	ß-ionona	(Huang et al.,	
morifolium		in vivo			pionona	2009)	
Malus domestica	CCD4	in vivo	β-caroteno	C9=C10 y C9'=C10'	β-ionona	(Huang et al., 2009)	
Osmanthus in vivo β-caroteno fragans		β-caroteno	C9=C10 y C9'=C10'	β-ionona	(Huang et al., 2009)		
Rosa	CCD4	in vivo	β-caroteno	$C_{0}-C_{10}=C_{0}^{2}-C_{10}^{2}$	Q ionono	(Huang et al., 2009)	
damascena	CCD4	in vitro	8-apo-β-carotenal	C9-C10 y C9 -C10	p-ionona		
				β-caroteno			
		in vitro	α-caroteno		β-ionona		
Solanum tuberosum	CCD4	in vivo	luteína	C9=C10 o C9'=C10'	10'-аро-β-	(Bruno et al., 2015)	
			III VIVO	zeaxantina		carotenal	·
				β-criptoxantina			
Citrus unshiu	CCD4	in vitro	β-caroteno,	$C7=C8 \circ C7'=C8'$	C ₂₀	(Ma et al 2013b)	
	CCD4	in vivo	zeaxantina	0,-0000, 00	030	(Ivia et al., 20150)	
			β-caroteno				
Citerra			β-criptoxantina		C 30		
clementina	CCD4b1	CCD4b1 in vitro	zeaxantina	C7=C8 o C7'=C8'	C10	(Koungo et al., 2013)	
			luteína		-10		
			α-caroteno				

CARACTERÍSTICAS DE LAS CCD4 DE ESPECIES DE CITRUS

En naranjas y mandarinas la CCD4b1 fue caracterizada como la responsable del corte asimétrico de la β -criptoxantina y la zeaxantina para generar β -citraurina (C30) cuya acumulación produce el color naranja-rojizo intenso en la pigmentación de híbridos de mandarina de la variedad fortuna (Rodrigo et al., 2013) (Fig. 6)



Nulo

Alto

Figura 6. Distintos niveles de coloración de tres frutas de mandarina con una composición similar de carotenoides C40 en la cáscara pero con una marcada diferencia en concentración del apocarotenoide β-citraurina (C30).Contenido nulo corresponde a una variedad mutante de *Citrus clementina* (39E7), contenido medio corresponde a *Citrus clementina* convencional y contenido alto a un híbrido (variedad Fortuna) (Rodrigo et al., 2013)

La CCD4a se expresa en todos los tejidos, pero predomina en hojas y flores, mientras que la CCD4b1 se expresa predominantemente en la cáscara y en menor nivel en las flores (Rodrigo et al., 2013), la CCD4c se expresa únicamente en las flores y ha sido propuesta como marcador molecular ideal para el estudio de las relaciones entre especies de citrus basados en su polimorfismo alélico (Zheng et al., 2015). En las CCD4b2 y CCD4d no se han observado patrones de expresión, se sugiere que estas enzimas han perdido residuos importantes y podría tratarse de pseudogenes (Rodrigo et al., 2013).

Medio



Figura 7. Niveles de expresión relativa de los genes CCD4a, CCD4b1 y CCD4c en las fases vegetativas y reproductivas de *Citrus sinensis*, para cada gen los niveles de expresión relativa se normalizaron en función a la muestra con el máximo nivel de expresión al que se le asignó el valor 1.Los valores representan los promedios observados ± la desviación estándar (Rodrigo et al., 2013)

IMPORTANCIA DE LOS CÍTRICOS EN EL URUGUAY

El cultivo de los cítricos es el rubro más importante en la fruticultura del Uruguay, centrada principalmente en la exportación de naranjas, mandarinas y limones. La superficie de producción de cítricos asciende a más de 14.000 hectáreas (Tabla 2).

Las dos principales zonas productoras se encuentran al norte en los departamentos de Salto, Paysandú y Artigas y al sur en los departamentos de San José, Colonia, Canelones y Montevideo. La zona con mayor producción es la zona norte y comprende más del 90% de la superficie cultivada de cítricos.

En el 2019 se exportaron más de 100.000 Tn de frutas cítricas y se calcula unos 2.000 empleos involucrados en la producción. Las mandarinas representan el 32% de la producción total de cítricos a nivel nacional involucrando una superficie de 5.000 ha.

Entre las características más importantes a los que apuntan los procesos de mejora genética son el tamaño de la fruta, presencia o no de semillas, la firmeza, el contenido del jugo, azúcares, nivel de acidez y el color de las frutas.

 Tabla 2. Producción (Tn), número de plantas y superficie cultivada de especies cítricas comerciales en Uruguay correspondientes al año 2019. Fuente: Encuesta Citrícola "Primavera 2020". Ministerio de Ganadería, Agricultura y Pesca - Oficina de Estadísticas Agropecuarias (MGAP-DIEA) Año: 2020.

Año 2019	Producción (Tn)	N° de plantas	Superficie (ha)
Naranja	115.688	3.384.000	6.107
Mandarina	78.744	3.034.000	5.346
Limón	47.422	744	2.582
Pomelo	790	39	72
Total	242.645	7.201.000	14.107

El presente trabajo de tesis se centra en el estudio *in silico* de las CCD4s provenientes de especies de *Citrus clementina* (CCD4a, b1, b2, c y d) combinando distintas técnicas bioinformáticas de análisis de secuencias para inferir sus características, su estructura tridimensional, su selectividad en cuanto a sustratos y los residuos que participan en las interacciones receptor-ligando, así como su interacción con las membranas biológicas mediante técnicas de simulación molecular.

REFERENCIAS

- Ahmad, T., Cawood, M., Iqbal, Q., Arino, A., Batool, A., Tariq, R. M. S., Azam, M., & Akhtar, S. (2019). Phytochemicals in Daucus carota and Their Health Benefits-Review Article. Foods (Basel, Switzerland), 8(9). <u>https://doi.org/10.3390/foods8090424</u>
- Ahrazem, O., Gómez-Gómez, L., Rodrigo, M. J., Avalos, J., & Limón, M. C. (2016). Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: Features and functions. International Journal of Molecular Sciences, 17(11). <u>https://doi.org/10.3390/ijms17111781</u>
- Ahrazem, O., Rubio-Moraga, A., Berman, J., Capell, T., Christou, P., Zhu, C., & Gomez-Gomez, L. (2016). The carotenoid cleavage dioxygenase CCD2 catalysing the synthesis of crocetin in spring crocuses and saffron is a plastidial enzyme. The New Phytologist, 209(2), 650–663. <u>https://doi.org/10.1111/nph.13609</u>
- Ahrazem, O., Trapero, A., Gómez, M. D., Rubio-Moraga, A., & Gómez-Gómez, L. (2010). Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: A deeper

study in Crocus sativus and its allies. Genomics, 96(4), 239–250. https://doi.org/https://doi.org/10.1016/j.ygeno.2010.07.003

- Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., Ghisla, S., Bouwmeester, H., Beyer, P., & Al-Babili, S. (2012). The path from beta-carotene to carlactone, a strigolactone-like plant hormone. Science (New York, N.Y.), 335(6074), 1348– 1351. https://doi.org/10.1126/science.1218094
- Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D. R., & Klee, H. J. (2006). Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. The Plant Journal, 45(6), 982–993. <u>https://doi.org/10.1111/j.1365-313X.2006.02666.x</u>
- Auldridge, M. E., McCarty, D. R., & Klee, H. J. (2006). Plant carotenoid cleavage oxygenases and their apocarotenoid products. Current Opinion in Plant Biology, 9(3), 315–321. <u>https://doi.org/10.1016/j.pbi.2006.03.005</u>
- Baranski, R., & Cazzonelli, C. I. (2016). Carotenoid Biosynthesis and Regulation in Plants. In Carotenoids (pp. 159–189). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118622223.ch10
- Beltran, J. C. M., & Stange, C. (2016). Apocarotenoids: A New Carotenoid-Derived Pathway. Sub-Cellular Biochemistry, 79, 239–272. https://doi.org/10.1007/978-3-319-39126-7_9
- Bruno, M., Beyer, P., & Al-Babili, S. (2015). The potato carotenoid cleavage dioxygenase 4 catalyzes a single cleavage of beta-ionone ring-containing carotenes and non-epoxidated xanthophylls. Archives of Biochemistry and Biophysics, 572, 126–133. https://doi.org/10.1016/j.abb.2015.02.011
- Bruno, M., Hofmann, M., Vermathen, M., Alder, A., Beyer, P., & Al-Babili, S. (2014). On the substrate- and stereospecificity of the plant carotenoid cleavage dioxygenase 7. FEBS Letters, 588(9), 1802–1807. https://doi.org/https://doi.org/10.1016/j.febslet.2014.03.041
- Cazzonelli, C. I. (2011). Carotenoids in nature: Insights from plants and beyond. Functional Plant Biology, 38(11), 833–847. <u>https://doi.org/10.1071/FP11192</u>
- Floss, D. S., & Walter, M. H. (2009). Role of carotenoid cleavage dioxygenase 1 (CCD1) in apocarotenoid biogenesis revisited. Plant Signaling & Behavior, 4(3), 172–175. <u>https://doi.org/10.4161/psb.4.3.7840</u>
- Fraser, P. D., & Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. Progress in Lipid Research, 43(3), 228–265. <u>https://doi.org/10.1016/j.plipres.2003.10.002</u>

- Harrison, P. J., & Bugg, T. D. H. (2014). Enzymology of the carotenoid cleavage dioxygenases: Reaction mechanisms, inhibition and biochemical roles. Archives of Biochemistry and Biophysics, 544, 105–111. https://doi.org/10.1016/j.abb.2013.10.005
- Hashimoto, H., Uragami, C., & Cogdell, R. J. (2016). Carotenoids and Photosynthesis. Sub-Cellular Biochemistry, 79, 111–139. <u>https://doi.org/10.1007/978-3-319-39126-7_4</u>
- Hou, X., Rivers, J., Leon, P., McQuinn, R. P., & Pogson, B. J. (2016). Synthesis and Function of Apocarotenoid Signals in Plants. Trends in Plant Science, 21(9), 792–803. <u>https://doi.org/10.1016/j.tplants.2016.06.001</u>
- Huang, F.-C., Molnár, P., & Schwab, W. (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. Journal of Experimental Botany, 60(11), 3011– 3022. <u>https://doi.org/10.1093/jxb/erp137</u>
- Ikoma, Y., Matsumoto, H., & Kato, M. (2016). Diversity in the carotenoid profiles and the expression of genes related to carotenoid accumulation among citrus genotypes. Breeding Science, 66(1), 139–147. <u>https://doi.org/10.1270/jsbbs.66.139</u>
- Ilg, A., Beyer, P., & Al-Babili, S. (2009). Characterization of the rice carotenoid cleavage dioxygenase 1 reveals a novel route for geranial biosynthesis. The FEBS Journal, 276(3), 736– 747. <u>https://doi.org/10.1111/j.1742-4658.2008.06820.x</u>
- Jia, K.-P., Baz, L., & Al-Babili, S. (2018). From carotenoids to strigolactones. Journal of Experimental Botany, 69(9), 2189–2204. <u>https://doi.org/10.1093/jxb/erx476</u>
- Johnson, E. J. (2002). The role of carotenoids in human health. Nutrition in Clinical Care : An Official Publication of Tufts University, 5(2), 56–65. <u>https://doi.org/10.1046/j.1523-5408.2002.00004.x</u>
- Karrer, P. (1966). Carotenoids, flavins and vitamin A and B2: Nobel lecture, December 11, 1937. In Nobel Lecture (pp. 433–448). Elsevier.
- Kloer, D. P., & Schulz, G. E. (2006). Structural and biological aspects of carotenoid cleavage. Cellular and Molecular Life Sciences, 63(19–20), 2291–2303. https://doi.org/10.1007/s00018-006-6176-6
- Kloer, Daniel P., Ruch, S., Al-Babili, S., Beyer, P., & Schulz, G. E. (2005). The structure of a retinal-forming carotenoid oxygenase. Science, 308(5719), 267–269. https://doi.org/10.1126/science.1108965
- Latari, K., Wust, F., Hubner, M., Schaub, P., Beisel, K. G., Matsubara, S., Beyer, P., & Welsch, R. (2015). Tissue-Specific Apocarotenoid Glycosylation Contributes to Carotenoid Homeostasis in Arabidopsis Leaves. Plant Physiology, 168(4), 1550–1562. <u>https://doi.org/10.1104/pp.15.00243</u>

- Latowski, D., Kuczyńska, P., & Strzałka, K. (2011). Xanthophyll cycle a mechanism protecting plants against oxidative stress. Redox Report, 16(2), 78–90. https://doi.org/10.1179/174329211X13020951739938
- Li, L., Yuan, H., Zeng, Y., & Xu, Q. (2016). Plastids and Carotenoid Accumulation. Sub-Cellular Biochemistry, 79, 273–293. <u>https://doi.org/10.1007/978-3-319-39126-7_10</u>
- Liang, M.-H., Zhu, J., & Jiang, J.-G. (2018). Carotenoids biosynthesis and cleavage related genes from bacteria to plants. Critical Reviews in Food Science and Nutrition, 58(14), 2314–2333. <u>https://doi.org/10.1080/10408398.2017.1322552</u>
- Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., Wahyudi, A., Motohashi, R., & Kato, M. (2013). Enzymatic formation of beta-citraurin from betacryptoxanthin and Zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. Plant Physiology, 163(2), 682–695. https://doi.org/10.1104/pp.113.223297
- Majer, E., Llorente, B., Rodríguez-Concepción, M., & Daròs, J. A. (2017). Rewiring carotenoid biosynthesis in plants using a viral vector. Scientific Reports, 7(January), 1–10. <u>https://doi.org/10.1038/srep41645</u>
- Meléndez-Martínez, A. J., Vicario, I. M., & Heredia, F. J. (2007). Pigmentos carotenoides: Consideraciones estructurales y fisicoquímicas. Archivos Latinoamericanos de Nutricion, 57(2), 109–117.
- Messing, S. A. J., Mario Amzel, L., Gabelli, S. B., Echeverria, I., Vogel, J. T., Guan, J. C., Tan, B. C., Klee, H. J., & McCarty, D. R. (2010). Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. Plant Cell, 22(9), 2970– 2980. <u>https://doi.org/10.1105/tpc.110.074815</u>
- Mishra, S., Upadhyay, S., & Shukla, R. K. (2017). The Role of Strigolactones and Their Potential Cross-talk under Hostile Ecological Conditions in Plants. Frontiers in Physiology, 7, 691. <u>https://doi.org/10.3389/fphys.2016.00691</u>
- Nisar, N., Li, L., Lu, S., Khin, N. C., & Pogson, B. J. (2015). Carotenoid metabolism in plants. Molecular Plant, 8(1), 68–82. <u>https://doi.org/10.1016/j.molp.2014.12.007</u>
- Ohmiya, A. (2013). Involvement of CCD4 in Determining Petal Color. In Carotenoid Cleavage Products (Vol. 1134, pp. 3–21). American Chemical Society. <u>https://doi.org/doi:10.1021/bk-2013-1134.ch003</u>
- Ohmiya, Akemi. (2009). Carotenoid cleavage dioxygenases and their apocarotenoid products in plants. Plant Biotechnology, 26(4), 351–358. https://doi.org/10.5511/plantbiotechnology.26.351

- Ohmiya, Akemi, Kishimoto, S., Aida, R., Yoshioka, S., & Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiology, 142(3), 1193–1201. <u>https://doi.org/10.1104/pp.106.087130</u>
- Parry, A. D., Babiano, M. J., & Horgan, R. (1990). The role of cis-carotenoids in abscisic acid biosynthesis. Planta, 182(1), 118–128. <u>https://doi.org/10.1007/BF00239993</u>
- Rodrigo, M. J., Alquézar, B., Alís, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461– 4478. <u>https://doi.org/10.1093/jxb/ert260</u>
- Rodriguez-Amaya, D. B. (2015). Nomenclature, structures, and physical and chemical properties. In Food Carotenoids (pp. 1–23). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118864364.ch1
- Rubio-Moraga, A., Rambla, J. L., Fernandez-de-Carmen, A., Trapero-Mozos, A., Ahrazem, O., Orzaez, D., Granell, A., & Gomez-Gomez, L. (2014). New target carotenoids for CCD4 enzymes are revealed with the characterization of a novel stress-induced carotenoid cleavage dioxygenase gene from Crocus sativus. Plant Molecular Biology, 86(4–5), 555–569. https://doi.org/10.1007/s11103-014-0250-5
- Rubio, A., Rambla, J. L., Santaella, M., Gomez, M. D., Orzaez, D., Granell, A., & Gomez-Gomez, L. (2008). Cytosolic and plastoglobule-targeted carotenoid dioxygenases from Crocus sativus are both involved in beta-ionone release. The Journal of Biological Chemistry, 283(36), 24816–24825. <u>https://doi.org/10.1074/jbc.M804000200</u>
- Schwartz, S. H., Qin, X., & Zeevaart, J. A. (2001). Characterization of a novel carotenoid cleavage dioxygenase from plants. The Journal of Biological Chemistry, 276(27), 25208–25211. <u>https://doi.org/10.1074/jbc.M102146200</u>
- Schwartz, S. H., Tan, B. C., Gage, D. A., Zeevaart, J. A., & McCarty, D. R. (1997). Specific oxidative cleavage of carotenoids by VP14 of maize. Science (New York, N.Y.), 276(5320), 1872–1874. <u>https://doi.org/10.1126/science.276.5320.1872</u>
- Shumskaya, M., & Wurtzel, E. T. (2013). The carotenoid biosynthetic pathway: Thinking in all dimensions. Plant Science, 208, 58–63. <u>https://doi.org/10.1016/j.plantsci.2013.03.012</u>
- Sui, X., Kiser, P. D., Lintig, J. Von, & Palczewski, K. (2013). Structural basis of carotenoid cleavage: From bacteria to mammals. Archives of Biochemistry and Biophysics, 539(2), 203– 213. <u>https://doi.org/10.1016/j.abb.2013.06.012</u>

- Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., & Li, L. (2018). Carotenoid Metabolism in Plants: The Role of Plastids. Molecular Plant, 11(1), 58–74. https://doi.org/10.1016/j.molp.2017.09.010
- Tan, B. C., Schwartz, S. H., Zeevaart, J. A., & McCarty, D. R. (1997). Genetic control of abscisic acid biosynthesis in maize. Proceedings of the National Academy of Sciences of the United States of America, 94(22), 12235–12240. <u>https://doi.org/10.1073/pnas.94.22.12235</u>
- Tian, X., Ji, J., Wang, G., Jin, C., Guan, C., & Wu, G. (2015). Molecular cloning and characterization of a novel carotenoid cleavage dioxygenase 1 from Lycium chinense. Biotechnology and Applied Biochemistry, 62(6), 772–779. <u>https://doi.org/10.1002/bab.1327</u>
- Vallabhaneni, R., Bradbury, L. M. T., & Wurtzel, E. T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. Archives of Biochemistry and Biophysics, 504(1), 104– 111. https://doi.org/10.1016/j.abb.2010.07.019
- von Lintig, J. (2012). Metabolism of carotenoids and retinoids related to vision. The Journal of Biological Chemistry, 287(3), 1627–1634. <u>https://doi.org/10.1074/jbc.R111.303990</u>
- Wang, C. X., Jiang, H., Yuen, J. J., Lee, S.-A., Narayanasamy, S., Curley Jr, R. W., Harrison, E. H., & Blaner, W. S. (2015). Actions of β-apo-carotenoids in differentiating cells: differential effects in P19 cells and 3T3-L1 adipocytes. Archives of Biochemistry and Biophysics, 572, 2–10. <u>https://doi.org/10.1016/j.abb.2015.01.009</u>
- Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone Signaling and Evolution. Annual Review of Plant Biology, 68, 291–322. <u>https://doi.org/10.1146/annurevarplant-042916-040925</u>
- Wurtzel, E. T. (2019). Changing Form and Function through Carotenoids and Synthetic Biology. Plant Physiology, 179(3), 830–843. <u>https://doi.org/10.1104/pp.18.01122</u>
- Yabuzaki, J. (2017). Carotenoids Database: structures, chemical fingerprints and distribution among organisms. Database: The Journal of Biological Databases and Curation, 2017(1), bax004. <u>https://doi.org/10.1093/database/bax004</u>
- Ytterberg, A. J., Peltier, J.-B., & van Wijk, K. J. (2006). Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. Plant Physiology, 140(3), 984–997. <u>https://doi.org/10.1104/pp.105.076083</u>
- Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., King, G. J., & Liu, K. (2015a). Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in Brassica species. New Phytologist, 206(4), 1513–1526. <u>https://doi.org/10.1111/nph.13335</u>

- Zhao, C., & Nabity, P. D. (2017). Phylloxerids share ancestral carotenoid biosynthesis genes of fungal origin with aphids and adelgids. PLoS ONE, 12(10), 1–14. https://doi.org/10.1371/journal.pone.0185484
- Zheng, X., Xie, Z., Zhu, K., Xu, Q., Deng, X., & Pan, Z. (2015). Isolation and characterization of carotenoid cleavage dioxygenase 4 genes from different citrus species. Molecular Genetics and Genomics, 290(4), 1589–1603. <u>https://doi.org/10.1007/s00438-015-1016-8</u>
- Zwanenburg, B., Pospíšil, T., & Ćavar Zeljković, S. (2016). Strigolactones: new plant hormones in action. Planta, 243(6), 1311–1326. <u>https://doi.org/10.1007/s00425-015-2455-5</u>

CAPÍTULO II

ANÁLISIS DE SECUENCIAS DE LAS ENZIMAS CCD4 DE *Citrus clementina*

INTRODUCCIÓN

Las CCDs constituyen una familia antigua con miembros omnipresentes en los distintos reinos del árbol de la vida: bacterias, hongos, plantas y animales (Kloer y Schulz, 2006; Ahrazem et al., 2016). Las CCDs presentes en especies vegetales constituyen un grupo de enzimas importantes en el metabolismo de los carotenoides que catalizan el corte oxidativo en los dobles enlaces carbono-carbono para formar apocarotenoides (Kato, 2012).

En *Arabidopsis thaliana* se han identificado 9 subfamilias CCDs (CCD1, NCED2, NCED3, CCD4, NCED5, NCED6, CCD7, CCD8 y NCED9). En otras especies vegetales, genes ortólogos

son típicamente nombrados de acuerdo a su homología con las secuencias provenientes de *Arabidopsis thaliana* (Huang et al., 2009).

Las secuencias CCDs en especies de citrus presentan un largo variable en un rango de 1000 a 2000 pares de base (bp) y con un largo de secuencias de aminoácidos predichas entre 450 a 600 aminoácidos (Kato et al., 2006).

Todas ellas comparten una arquitectura típica en β -propeller y un centro catalítico altamente conservado por una esfera de coordinación de cuatro histidinas que coordinan a un hierro no hémico en una disposición de pirámide trigonal, lo que sugiere un mecanismo catalítico común (Kloer y Schulz, 2006).

La conservación es variable en distintas especies. En la mayoría de las CCDs de origen vegetal se reportan la secuencia de péptido de destinación al cloroplasto (cpTP) en la región N-terminal de un largo variable (Liang et al., 2018). La región más variable de la secuencia está constituida por una región de α -hélices y *loops* que establecen un dominio hidrofóbico predichos como un sitio de interacción con membranas tilacoidales (Kloer y Schulz, 2006).

Los miembros de la subfamilia CCD4 no están aún lo suficientemente caracterizados. El papel biológico de la mayoría de las CCD4 de plantas sigue siendo aún desconocido. Un creciente número de estudios han señalado la relación entre los genes CCD4 y el color de flores y frutas (Rodrigo, et al., 2013; Zhang et al., 2015; Ohmiya et al., 2006). En especies cítricas, las CCD4s se han relacionado con la formación de color del flavedo de las frutas mediante la relación en la formación del apocarotenoide β -citraurina, sintetizada por el clivaje de las xantofilas zeaxantina o β -criptoxantina en las posiciones 7,8 o 7',8'(Rodrigo, et al., 2013).

La mayoría de las plantas poseen al menos dos genes CCD4 (Ahrazem et al., 2010; Vallabhaneni et al., 2010), mientras que en *Chrysanthemum morifolium* y *Populus trichocarpa* se han identificado 5 miembros de esta subfamilia (Ahrazem et al., 2010), en *Vitis vinífera* y *Brassica napus* se han identificado al menos cuatro miembros (Lashbrooke et al., 2013; Rubio-Moraga et al., 2014). Por lo general en la mayoría de los genes CCD4 no se detectan intrones o si los poseen suele ser típicamente uno (Ahrazem et al., 2010; Vallabhaneni et al., 2010), la posición y el tamaño de este intrón suele ser variable a lo largo de las distintas especies (Ahrazem et al., 2010). La similitud de las secuencias miembros de la subfamilia CCD4 es variable entre especie y por lo general mayor al 30% al nivel de secuencia proteica (Ahrazem et al., 2016).

La caracterización de patrones de expresión en órganos y tejidos de genes identificados como miembros de la subfamilia CCD4 en citrus han revelado que no todos se expresan y algunos han sido predichos como pseudogenes originados a partir una duplicación en tándem de otros miembros de esta subfamilia (Rodrigo et al., 2013).

En base a lo antes dicho, es que entendimos la necesidad de plantear en este capítulo el estudio de las características de las secuencias CCDs provenientes de citrus y en especial a la subfamilia CCD4, analizando su composición, estructura secundaria, la presencia de motivos conservados y las relaciones filogenéticas entre los distintos miembros de otras especies vegetales.

OBJETIVOS

- Identificar y caracterizar a las secuencias CCDs presentes en *Citrus clementina*
- Analizar la variabilidad de las secuencias CCDs de Citrus clementina
- Identificar los motivos conservados en las secuencias CCDs de Citrus clementina
- Realizar una reconstrucción filogenética que permita establecer las relaciones evolutivas entre las CCD4 de *Citrus clementina* y otras CCDs de origen vegetal

MATERIALES Y MÉTODOS

IDENTIFICACIÓN Y CARACTERIZACIÓN DE CCDs DE Citrus clementina

Para la identificación de las enzimas responsables de la síntesis de apocarotenoides se recurrió a la base de datos <u>Phytozome v13</u> (Goodstein et al., 2012), una base de datos especializada en datos genómicos de plantas. Para la búsqueda en la base de datos se utilizó la secuencia de la CCD4 proveniente de *Arabidopsis taliana* (Uniprot id: <u>O49675</u>) mediante el algoritmo *tBlastn* (Gerts et al., 2006) restringiendo la búsqueda al genoma de *Citrus Clementina*. Las secuencias

resultantes fueron identificadas y caracterizadas en función al número de nucleótidos, presencia o ausencia de intrones, número de aminoácidos.

La presencia de secuencia de péptidos de destinación al cloroplasto (cpTP) se determinó mediante tres programas distintos de predicción en base a la secuencia: ChloroP 1.1(Emanuelsson et al., 1999), TargetP-2.0 (Emanuelsson et al., 2000) y LOCALIZER (Sperschneider et al., 2017).

ALINEAMIENTO DE LAS SECUENCIAS

Con el objetivo de evaluar la similitud entre los distintos subtipos de CCD4 presentes en *Citrus clementina* se procedió al alineamiento múltiple de las secuencias utilizando el programa MUSCLE v3.8.31 (Edgar, 2004) y su posterior corrección manual con el programa AliView v 1.27 (Larsson, 2014).

PREDICCIÓN DE LA ESTRUCTURA SECUNDARIA

Para predecir la estructura secundaría de las secuencias de la CCD4 se procedió primeramente a estudiar las regiones probablemente desestructuradas de la secuencia mediante el software DISOPRED3 v3.1 (Jones y Cozzetto, 2015) y posteriormente se procedió a la predicción de la estructura secundaria en base a la información de la secuencia de aminoácidos mediante el software PSIPRED v4.0 (Jones, 1999), adicionalmente se realizó un análisis de predicción de interacción de las α -hélices con membranas utilizando las herramientas MEMSAT-SVM (*Membrane Helix Prediction*) (Nugent y Jones, 2012) y MEMPACK (*TM Topology and Helix Packing*) (Nugent y Jones, 2009).

BASES DE DATOS DE SECUENCIAS CCDs

Para analizar las relaciones evolutivas de las secuencias CCD4s presentes en *Citrus clementina* con sus homólogos de origen vegetal se construyó una base de datos de secuencias de aminoácidos mediante el algoritmo de búsqueda *Blastp* (Altschul et al., 1990) utilizando como secuencia *Query* cada miembro de la subfamilia CCD4 de *Citrus clementina* y estableciendo como valor de corte un E-value de e^{-20} y un límite de 100 entradas por búsqueda (*Query*) restringiendo el set de datos a secuencias provenientes de plantas. Como la base de datos mediante *Blastp* genera secuencias altamente redundantes se aplicó el método de factorización QR (Sethi et al., 2005) para crear una base de datos de secuencias no redundantes eliminando aquellas secuencias que compartan más de 75% de identidad. Con estas secuencias se procedió al alineamiento utilizando el programa *MUSCLE*.

PREDICCIÓN DE MOTIVOS

La predicción de motivos conservados de las secuencias CCD4 se realizó mediante el software *MEME Suite* v5.3.3 (*Multiple Em for Motif Elicitation*) (Bailey et al., 2009) utilizando la base no redundante de secuencias CCD4 de origen vegetal. Los motivos conservados encontrados se representaron mediante el esquema LOGO (Schneider y Stephens, 1990) e identificadas en las secuencias de las CCD4 de *Citrus clementina*. Para dar información de la localización estructural de dichos motivos se mapearon las secuencias motivo sobre la estructura cristalográfica de la VP14 (PDB: <u>3NPE</u>) una NCED de *Zea mays* que posee los motivos conservados de las secuencias (Messing et al., 2010)

ANÁLISIS DE VARIABILIDAD DE LAS SECUENCIAS

Para el análisis de ocupación del alineamiento y entropía de la secuencia (entropía de Shannon) se tomó el alineamiento del conjunto no redundante de secuencias CCD4 de origen vegetal y se filtraron las posiciones en base a la secuencia más larga de CCD4 de *Citrus clementina* (CCD4a de 603 aminoácidos) eliminando aquellas posiciones del alineamiento en los que se observen *gaps* en esta secuencia conservando las demás posiciones. La ocupación se midió como la frecuencia en la que una posición (columna) del alineamiento está ocupada por algún aminoácido, mientras que la entropía de la secuencia mediante la ecuación de la entropía de la información de Shannon (Strait y Dewey, 1996) descrita en la Ecuación 1:

Ecuación 1.
$$S(i) = \sum_{x_i=1}^{20} P(x_i) \log \frac{1}{P(x_i)}$$

donde,

 $P(x_i)$ es la probabilidad de observar un determinado aminoácido x (de 20 posibles) en una posición (columna) i del alineamiento.

La entropía de Shannon mide la variabilidad en una determinada posición del alineamiento. Típicamente valores de S(i) > 2 se consideran regiones variables, S(i) < 2 regiones conservadas y S(i) < 1 regiones altamente conservadas.

La información mutua para el par de posiciones i,j del alineamiento se calculó mediante el método descrito por Martin et al., (2005) usando la Ecuación 2:

Ecuación 2.
$$I(i, j) = \sum_{x_i=1}^{20} \sum_{y_j=1}^{20} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}$$

donde,

 $P(x_i)$ es la probabilidad de observar un determinado aminoácido x (de 20 posibles) en una posición (columna) i del alineamiento.

 $P(y_j)$ es la probabilidad de observar un determinado aminoácido y (de 20 posibles) en una posición (columna) j del alineamiento.

 $P(x_i, y_j)$ es la probabilidad de observar de forma conjunta al aminoácido x y al aminoácido j en las posiciones i y j del alineamiento, respectivamente

Como la información mutua está afectada por las relaciones filogenéticas entre las secuencias, para eliminar dicha influencia se utilizó la corrección del producto promedio (*Average Product Correction* – APC) siguiendo la Ecuación 3 y 4 (Dunn et al., 2008):

Ecuación 3.
$$MIp(i, j) = I(i, j) - APC$$

en la que,

Ecuación 4. APC =
$$\frac{[I(i, x)I(j, x)]}{\langle I(i, j) \rangle}$$

donde,

I(i, x) es la media de la información mutua de la columna i $(I(i, x) = \sum_j I(i, j))$ I(j, x) es la media de la información mutua de la columna j $(I(j, x) = \sum_i I(j, i))$ $\langle I(i, j) \rangle$ es la media de la información mutua de todos los pares de columnas i,j

Para facilitar la visualización se representaron los valores mediante un *heatmap* normalizando los valores en función al máximo observado (Ecuación 5):

Ecuación 5.
$$MIp(i,j)_{norm} = \frac{MIp(i,j)}{max(MIp(i,j))}$$

La información mutua nos permite identificar las posiciones variables del alineamiento múltiple que coevolucionan (cambios de aa de forma conjunta) a lo largo de las distintas especies.

RECONSTRUCCIÓN FILOGENÉTICA

Finalmente se construyó un árbol filogenético no enraizado en base al alineamiento de las secuencias de aminoácidos de CCD4 de origen vegetal homólogos de *Citrus clementina* para estudiar las relaciones entre ellos. El alineamiento fue previamente filtrado removiendo secuencias espurias y regiones pobremente alineadas mediante el software *trimAl* v1.3 (Capella-Gutiérrez et al., 2009). Posteriormente la reconstrucción del árbol filogenético se realizó mediante el método de Máxima Verosimilitud (ML) utilizando el software *PhyML* v3.0 (Guindon et al., 2010). El modelo y los parámetros utilizados para la reconstrucción filogenética se obtuvo mediante el programa *SMS* v1.8.4 (*Smart Model Selection in PhyML*) (Lefort et al., 2017). La robustez del árbol obtenido se evaluó mediante el método de bootstrap no paramátrico con 1000 réplicas.

Por último, para analizar con una perspectiva más general a las CCOs de origen vegetal homólogas a todas las presentes en *Citrus clementina* se construyó un árbol filogenético incluyendo todas las secuencias CCOs siguiendo el procedimiento descrito en la sección BASES DE DATOS DE SECUENCIAS CCDs y la reconstrucción del árbol mediante el procedimiento descrito en el párrafo anterior.

RESULTADOS Y DISCUSIÓN

IDENTIFICACIÓN Y CARACTERIZACIÓN DE CCOs DE Citrus clementina

Mediante la búsqueda tBlastn utilizando como secuencia Query a la CCD4 de *Arabidopsis taliana* (Uniprot id: <u>049675</u>) en la base de datos *Phytozome* restringiendo la búsqueda al genoma de *Citrus Clementina* se indentificaron 12 genes candidatos a CCOs, de los cuales 5 serían miembros de la subfamilia CCD4: la a, b1, b2, c y d (Tabla 1).

La cantidad de pares de base (pb) que constituye la región codificante varía de acuerdo al subtipo de CCOs y con una longitud variable que va desde 1200 a 1800 pb. La gran mayoría no posee intrones. La longitud de la secuencia predicha de aminoácidos varía en un rango de 400 a 600 aminoácidos. El grado de similitud de la secuencia de aminoácidos es variable a lo largo de las distintas subfamilias de CCOs presentes en *Citrus Clementina* con valores que van desde el 30

al 95%. Si bien, en la gran mayoría se ha podido detectar mediante análisis de predicción de secuencias cpTP, no se han detectado para la CCD1 y la CCD7 con los métodos evaluados.

Tabla 1.	Enzimas CCOs	identificadas e	en individuos	de Citrus	clementina	en la b	ase de	datos
	Phytozome v13.	Miembros de l	la subfamilia (CCD4 se e	ncuentran re	saltados	s en na	ranja.

Gen	Código genómico ^a	GeneID ^b	N° de nucl.º	Intrones	Uniprot ^d	N° de aa ^e	cpTP ^f
CCD1	CICLE_v10031039mg	<u>18043785</u>	1233	13	V4SWQ9	589	No
CCD1*	CICLE_10031014mg	<u>18043784</u>	1794	13	<u>V4TR28</u>	597	Si
NCED2	CICLE_v10014639mg	<u>18050641</u>	1830	0	<u>A0SE34</u>	609	Sí
NCED3	CICLE_v10019364mg	<u>18046011</u>	1821	0	<u>A0SE37</u>	606	Sí
CCD4a	CICLE_v10031003mg	<u>18043465</u>	1812	0	<u>A0SE35</u>	603	Sí
CCD4b1	CICLE_v10028113mg	<u>18034103</u>	1692	0	V4SJR7	563	Sí
CCD4b2	CICLE_v10030384mg	<u>18034858</u>	1527	0	<u>V4SA09</u>	508	Sí
CCD4c	CICLE_v10011335mg	<u>18038224</u>	1794	0	<u>V4STW3</u>	597	Sí
CCD4d	CICLE_v10013726mg	<u>18038241</u>	1236	2	<u>V4STW5</u>	412	Sí
NCED6	CICLE_v10006710mg	<u>18032346</u>	1779	0	<u>V4S0R4</u>	592	Si
CCD7	CICLE_ 10027500mg	<u>18035847</u>	1866	6	V4SDY8	590	No
CCD8	CICLE_10010609mg	<u>18055012</u>	1671	5	<u>V4UI82</u>	556	Sí

a Haploid Clementine Genome, International Citrus Genome Consortium, Phytozome v13 (Goodstein et al., 2011) b Gene assigns an identifier. National Center for Biotechnology Information (NCBI) (Murphy et al., 2016)

c Número de nucleótidos

d Código de acceso Uniprot (The UniProt Consortium, 2018)

e Número de aminoácidos

f Predicho por los servidores servidor ChloroP 1.1 (Emanuelsson et al., 1999), TargetP-2.0(Emanuelsson et al., 2000) y LOCALIZER (Sperschneider et al., 2017)

* Probable gen duplicado de la CCD1

El análisis de similitud mediante la clasificación en clústers muestra la existencia de cuatro grupos bien identificados, el primero que agrupa a todos los miembros de la subfamilia CCD4 (CCD4a, b1, b2, c y d), el segundo conformado por las NCEDs (NCED2, 3 y 6), el tercero formado por CCD1 y CCD1* (probablemente un gen duplicado de CCD1 dada su ubicación en el genoma y su alto grado de similitud) y por último el grupo de los más disimiles formados por la CCD7 y CCD8 que logran una similitud no mayor del 40% respecto a las demás secuencias. Los resultados del análisis de *clustering* se muestran en la Fig. 1.



Figura 1. *Heatmap* del análisis de similitud de las secuencias de aminácidos de las CCDs de *Citrus Clementina*. Se señalan los clusters encontrados mediante los colores de las ramas del dendograma. En recuadro azul se indican las secuencias pertenecientes a la subfamilia CCD4.

La CCD4 en *Citrus clementina* se muestra como el grupo más numeroso y claramente diferenciado de las demás CCDs presentes, además se observa la estrecha relación entre la CCD4b1 y b2 como así también entre la CCD4c y la d.

PREDICCIÓN DE MOTIVOS CONSERVADOS

La descripción de los motivos conservados en las secuencias de CCD4 estimados mediante el software MEME suite, se muestran en la Tabla 2. Estos tienen una longitud de secuencia que varía en un rango de 30 a 50 aminoácidos y están ampliamente distribuidos en todas las CCD4 de plantas.

Tabla 2. Motivos conservados en secuencias CCD4 de origen vegetal. Secuencia consenso, valoresperado (*E-value*) y largo de la secuencia (aa).

\mathbf{N}°	Secuencia consenso	E-value	Largo
Motif 1	HLFDGDGMLHSJRISQGRATFCSRYVKTYKY	6,6e-985	31
Motif 2	EDDGYVVSYVHBENTGESRFLVMDAKSPNLDIVAAVKLPRRVPYGFHGLF	7,4e-1450	50
Motif 3	KVPRIGIIPRYAKDESEMRWFDVPGFNAIHAINAWEEEDDE	2,2e-1144	41
Motif 4	GFGLANTSLAFFGNKLFALGESDLPYAIRLTSNGDIETJGR	1,9e-1092	41

El diagrama LOGO de la secuencia consenso (Fig. 2) muestra constituciones similares del entono a nivel de composición de los aminoácidos en los motivos encontrados, lo que sugiere que se tratan de elementos comunes de una estructura simétrica.

Para visualizar mejor la relevancia estructural de estos motivos alineamos las secuencias consenso sobre la estructura de la VP14, observando que los mismas se sitúan en las láminas β hacia la cara interna de la arquitectura en β -propeller, característica de todos los miembros de la familia de las CCDs.

Dos de los motivos encontrados (*Motif 2* y *Motif 3*) incluyen a residuos de histidina fuertemente conservados a lo largo de todas las especies vegetales, ya que estos intervienen directamente en la coordinación del Fe⁺² en el centro catalítico de la enzima (Fig. S1). En esta misma región se localizan los residuos glutámicos que fueron descritos como una segunda esfera de coordinación del Fe⁺² (Kloer y Schulz, 2006).

El análisis de la distribución de los motivos también mostró la ausencia del motivo *Motif* 2 en la secuencia de la CCD4d, esto constituye una pérdida importante en la estructura de la enzima que afecta a la esfera de coordinación del Fe^{+2} en el centro catalítico, y sugiere que la misma no sería activa en la actualidad por lo que se podría tratar de un pseudogen.


Figura 2. Motivos conservados en genes CCD4. Izquierda arriba: secuencias motivo encontradas y representadas mediante el esquema LOGO. Izquierda abajo: P-valor y distribución de motivos (en rectángulos de colores) en las secuencias de aminoácidos (líneas grises) en genes CCD4 de *Citrus clementina*. Derecha: Posición de las secuencias motivo en la estructura de la VP14. *Motif* 1 en rojo, *motif* 2 en cyan, *motif* 3 en verde y *motif* 4 en violeta.

PREDICCIÓN DEL DESORDEN DE LAS SECUENCIAS

El análisis de predicción de desorden reveló la existencia de regiones desestructuradas que se ubican al inicio y al final de las secuencias, y que son de longitudes variables. El límite de probabilidad de desorden está fijado en el nivel 50% (Fig. 3).

Si bien en general todas las secuencias presentan una región de máxima probabilidad de desorden antes de los primero 100 aminoácidos, en particular, la CCD4a es la que presenta la región más amplia por encima de ese límite, extendiéndose hasta los primeros 86. La CCD4b1 resulta afectada solamente en los primeros 10 aminoácidos, y vuelve a presentar un pico cercano al límite de corte de predicción en el aminoácido número 45 de la secuencia.

La presencia de estas regiones desestructuradas coincide con lo observado por Rodrigo et al., (2013), en la que describen la presencia del cpTP en esta región. Proteínas con la señalización cpTP una vez traslocadas al interior del cloroplasto sufren la eliminación de esta región debido a la acción mediada por la peptidasa de procesamiento del cloroplasto durante el proceso de maduración (Lee y Hwang, 2018).



Predicción de desorden

Figura 3. Predicción de desorden de la secuencia. CCD4a (naranja), CCD4b1 (rojo), CCD4b2 (azul), CCD4c (verde), CCD4d (cyan). El punto de corte se señala con la línea horizontal punteada. Observe que la secuencia CCD4d tiene un largo de solo 400 aminoácidos.

PREDICCIÓN DE LA ESTRUCTURA SECUNDARIA

La predicción de estructura secundaria en todas las secuencias CCD4 analizadas, muestra una marcada presencia de regiones no estructuradas al inicio, seguidas por α -hélices. Por otro lado, se observa el predominio del plegamiento en láminas- β a lo largo de la secuencia (Fig. 4).

Las α -hélices que se encuentran al inicio de la secuencia fueron identificadas como probables dominios de interacción con las membranas tilacoidales. Estos resultados concuerdan con los obtenidos por Ma et al., (2013) en CCD4 provenientes de *Citrus unshui*. Las α -hélices se han propuesto como responsables de la formación de un domo hidrofóbico predominante en CCDs

de origen vegetal (Kloer y Schulz, 2006). Con estas interacciones las proteínas quedarían hipotéticamente ancladas a la membrana tilacoidal y cabe resaltar que sus sustratos (carotenoides) están principalmente acumulados en dichas membranas.

Una descripción más detallada de la predicción de la estructura secundaria junto con la secuencia se puede observar en las Fig. S2-S6.



Predicción de la estructura secundaria

Figura 4. Predicción de la estructura secundaria de las secuencias de aminoácidos de las CCD4 provenientes de *Citrus clementina*. Las regiones en rojo indican α-hélices, amarillo para las láminas-β. Las flechas muestran las α-hélices identificadas como probables dominios de interacción con las membranas.

ALINEAMIENTO Y ANÁLISIS DE LA CONSERVACIÓN DE LAS SECUENCIAS

La comparación de las secuencias completas de los miembros de la subfamilia CCD4 de *Citrus clementina* se muestra en la Fig. 5. En ella se evidencia que la región más variable es la que involucra a los primeros 100 aminoácidos (región N-terminal) en la que se encuentra la secuencia cpTP.

5 10 15 20 25 30 35 40 45 50 55 60
CCD4a MDSFSSSFFSSVLPPKLLTS LVIIPHKSTKPTSFSVSSVRIEEKPAVNQDPKTPTTTTSTPT
CCD4b2 MDHMNYSSLKISYNHP PKPKITYYN
68 73 78 83 88 93 98 103 108 113 118 123
CCD4a RTVSQPQPKPQPRTQSSYLPKKRAEPTIPTIILNACDDIINNFIDPPLKYSVDPRHVLSDNFA
CCD4b2 HTNEQIIPFLPFQKPKRCVPIQSSMKKNSSCDTKSSPSLFTNIIDRPLHPSVDPKHVFTGNFA
CCD4c ETTEENNKQQTKTTQKPSSSVSTTTNVSAVILNTLNKLIKPKLQAQPL SLDPSRAFSNNLA
CCD4d TSSSSSSARTMIV ···· NIINNLFN·NLMMINS·N··YK··LQ··P···SLDPKRALSNNFA
131136141146151156161166171176181186
CCD4a PVDELPPTECEVVQGSLPSCLDGAYIRNGPNPQYLPRG PYHLFDGDGMLHSIKISRGRATLC
CCD4b1 PVDELGPTECPVVDGKLPDSLTGTYTRNGPNPQHMPRG PLHFFEGDGMLHSMQLSKGRATYS
CCD4c PVDELHPTQCQVTQGSLPSCLEGAYIRNGPNPQYPPRSEPYSPFDGDGMLHCIKISQGQATFC
CCD4d PVDELPPTQCQVIQGSLPSCLDGAYI
194 199 204 209 214 219 224 229 234 239 244 249
CCD4a SRYVRTYKYTIENEAGSPILPNVFSGFNGLTASAARGALSAARLLAGQFNPVNGIGLANTNLA
CCD4b1 SRYVKTYKYKLERDAGGQIFPNVFSGFYGLVDMVQCVAS TARVLMGHVNLKKGFGLANTSLA
CCD4c SRYVKTYKYTIENTVGSPIVPRPFSSFSGPLAFLTRAGVLAARVLTGQLSINSGFGLANTSLA
CCD4d SRCVKTYKYTIENTVGSPFVSRAFSSFSGPLA
257 262 267 272 277 282 287 292 297 302 307 312
CCD4a FFGNRLYALGESDLPYAIRLTPNGDIETIGRHDFDGKLAMSMTAHPKLDSDTGEAFAFRYGPV
CCD4b1 FFSSKLLALGESDLPYIIKCTREGDIETLGRWDFDEELFASMTAHPKTDMTTKETFAFKFSPF
CCD462 FFSSKLLALGESDLPYTINSTREGDVETVGRWDFDEALFASMTAHPKTDTTTKETFAFKFSPV
CCD4d FLGDHLFAICEGDLPYAVRLTSNGDTETLGHYDFDEKLGMNMTAHPKKDPATDETFAFRYGPI
320 325 330 335 340 345 350 355 360 365 370 375
CCD4b1 FSPHLTFSRFDANGVKQKDVPILSIN RPTFIHDFAITKRFAIFSETQLAVSAANVMLGK
CCD4d P · PFVTHFYFDASGEKQIDVHIAMAG · · · RHPLFMHDFAITKKHTVSVDTQLNMYYTEMILRG
383 388 393 398 403 408 413 418 423 428 433 438
CCD4b1 GMPTVFDPEKVPRIGIILKYAKSDAEMKWFNVPGFNAMHVFNAW ENGEDEVVLIATNATSIE
CCD4b2 GMPTVFDPEKIPRIGIILKYATSDAEMKWFNVPGFNAMHVFNAW ENGDDEIVLIATTATSIE
CCD4C_EPILHSDPTKVSRTGVMPRYAKNSSQMRWFDAPGINNTHAVNAWDEEDGNVTVMVAPNTLSVE
CCD4a HILDRIDLVHALVERVRIDLRIGIVIRKPMSARNLDFGVINPAYMAKKSRYVYAAVGDPMPKI CCD4b1 NLFLKVDMVHISLEKVRINLRIGNVSRNILSARNLELGSINSSYIGKKNRYVFVGVGKEIPKM
CCD4b2 NL FHKIDMVHFSLEKVRINLRTGHVFRNILSTRNLELGSINSSYIGKKNRYVFMGVGKEIPKM
CCD4c QVFERMELIHCLMEKVRIDLRTGVVSRHPISARNLDMAVINPGYVAKKNRYVYAAVGDPMPKM
509 514 519 524 529 534 539 544 549 554 559 564
CCD4b2 EGVVKIDLEKEIEVSR RFYGPSCFGGEPLFVPRNEDHLDAADEDDGFVVTYIHDEIHE
CCD4c TGLVKLDLKKG ERRECTVATRIYGPDCYGGEPFFVARDAADKE ADEDDGYILSYVHDEKAG
572 577 582 587 592 597 602 607 612 617 622 627
CCD4a ESKFLVMDAKSPRLDIVAAVKLPRRVPYGFHGLFVRQAELDKLL
CCD4b2 ESKFLVMDAKSPNLDIVAAVKLPRRVPYGLHGLFVHKDNH
CCD4c ESKFLVMDAKSPHLDIVAEVKLPQRVPAGF <mark>H</mark> GLFVREIDFCKL
CCD4d

Figura 5. Alineamiento de las secuencias CCD4 de *Citrus clementina*. Los motivos estás subrayados: *motif 1* (rojo), *motif 2* (cyan), *motif 3* (verde) y *motif 4* (violeta). Histidinas responsables de la coordinación del Fe⁺² están señaladas mediante rectángulos amarillos. Regiones conservadas se indican en gris oscuro.

También se puede observar la distribución de los motivos de la secuencia y las histidinas conservadas que intervendrían en la esfera de coordinación del Fe^{+2} en el centro catalítico (Fig. S1).

Analizando el alineamiento de las CCD4 de *Citrus clementina* en la Fig. 5, se observa que la CCD4d se diferencia de las restantes en su longitud por la ausencia de los 100 aminoácidos en la región C-terminal que todas las demás poseen. Esta región corresponde a uno de los motivos conservados que se ha asociado a una de las láminas- β del centro del β -propeller (Fig. 2) que contiene a una histidina de la esfera de coordinación del Fe⁺². En consecuencia, se puede afirmar que ese es el motivo de la pérdida de su capacidad catalítica.

Por otro lado, estudios a niveles de expresión génica han observado que CCD4b2 no se expresa en ningún tejido/órgano (Rodrigo et al., 2013). Dada que la duplicación en tándem de genes CCD4 es una característica común en especies vegetales (Ahrazem et al., 2010), se puede concluir que CCD4b2 y CCD4d se tratarían de copias de CCD4b1 y CCD4c respectivamente. Este fenómeno permitiría acumular mutaciones dada la menor presión evolutiva sobre las mismas, resultando en secuencias de mayor variabilidad que las existentes entre ortólogos de esta subfamilia.

ANÁLISIS DE LA VARIABILIDAD DE LAS SECUENCIAS

El análisis de entropía de las secuencias (entropía de *Shannon*) para el alineamiento de las CCDs de origen vegetal muestra que la región más variable involucra a los primeros 100 aminoácidos (bit-score > 2), esto concuerda con lo observado en el alineamiento de las secuencias CCD4 provenientes de *Citrus clementina*. En esta región se puede observar una ocupación variable del alineamiento lo que indica un largo variable, es decir la ausencia de parte de estas regiones en algunos genes miembros de esta subfamilia (inserciones-deleciones) (Fig. S7).

El análisis de información mutua (MIp_{norm}) muestra que esta variabilidad en la región que va desde el aminoácido número 15 al 75, tomando como molde a la CCD4a (la más extensa de las CCD4 en *Citrus clementina*), pero los cambios en las posiciones no son independientes entre sí y está caracterizada por la coocurrencia de pares de aminoácidos en este sitio lo que da evidencias de coevolución de las posiciones que involucran a la secuencia cpTP. (Fig. 6)



Figura 6. Información Mutua (MIp_{norm}) en el alineamiento múltiple de las secuencias de CCD4 de origen vegetal filtradas por columnas (posiciones de aminoácidos) de CCD4a. En rojo se resalta la región en la que sustituciones de aminoácidos ocurren de manera conjunta (posición 15 a 75).

ANÁLISIS FILOGENÉTICO

La relación entre distintas proteínas de la subfamilia CCD4 provenientes de distintas especies vegetales se analizó en base a la secuencia y con la que se generó el árbol filogenético (Fig. 7). En el árbol se puede observar claramente que las secuencias provenientes de *Citrus*

clementina más cercanamente relacionadas corresponden a los pares CCD4b1/b2 y CCD4c/d lo que fortalece la hipótesis de que estas secuencias se tratarían de genes duplicados. También se puede observar que la CCD4a junto con otra especie del género citrus (*Citrus unshui*) está más cercanamente emparentada al grupo formado por las secuencias CCD4c/d mientras que CCD4b1/b2 se encuentran en ramas separadas y más lejanamente relacionadas a los demás miembros CCD4 de *Citrus clementina*.



Figura 7. Arbol filogenético de secuencias de aminoácidos del tipo CCD4 de *Citrus clementina* calculado por máxima verosimilitud (Log likelihood: -11419.306373). Se muestran los valores porcentuales del bootstrap (1000 réplicas). Se señalan los grupos de CCD4 de *Citrus Clementina* cercanamente emparentados mediante esferas de colores: CCD4a (verde); CCD4b1 y CCD4b2 (naranja) y; CCD4c y CCD4d (azul). Las siglas representan a las especies: At (*Arabidopsis thaliana*), Bn (*Brassica napus*), Cc (*Citrus clementina*), Cu (*Citrus unshui*), Cmor (*Chrysanthemum x morifolium*), Cros (*Crocus sativus*). Md (*Malus domestica*), Of (*Osmanthus fragrans*), Rd (*Rosa damascena*), St (*Solanum tuberosum*) y Vv (*Vitis vinífera*). Información de los identificadores de secuencia en la Tabla S1.

Finalmente, para analizar con una perspectiva más general a las CCOs de origen vegetal homólogas a las presentes en *Citrus clementina* se construyó un árbol filogenético (Fig. 8), en él se observan las relaciones entre las secuencias claramente definidas en dos grandes grupos o familias bien diferenciadas: las NCEDs y las CCDs. Se identificaron 3 subfamilias NCEDs homólogas a las presentes en *Citrus clementina* (NCED2, NCED3 y NCED6) y 4 subfamilias CCDs (CCD1, CCD4, CCD7 y CCD8). Estos resultados sugieren que cada familia de las CCOs analizadas evolucionan de forma independiente.



Figura 8. Árbol filogenético de 75 proteínas CCOs de distinto origen vegetal, homólogas a las presentes en *Citrus clementina*, obtenido por el método de máxima verosimilitud (ML). Los colores indican las distintas subfamilias de CCOs. Se indican con un círculo las hojas correspondientes a las especies de *Citrus clementina*. Información de los identificadores de secuencia en la Tabla S2.

CONCLUSIONES

Se han identificado 12 genes de enzimas de clivaje de carotenoides (CCDs) en *Citrus clementina* (CCD1, CCD1*, NCED2, NCED3, CCD4a, CCD4b1, CCD4b2, CCD4c, CCD4d, NCED6, CCD7, CCD8) y específicamente 5 miembros de la subfamilia CCD4. El análisis de las secuencias revela que las CCD4s forman un grupo claramente separado de las demás.

Los motivos conservados identificados de las secuencias CCD4s de origen vegetal presentan un largo variable entre 30 a 50 aminoácidos. Las secuencias consenso de tales motivos poseen una composición similar, lo que sugiere que estas regiones se organizan en plegamientos similares conformando láminas- β en la arquitectura típica de β -propeller características de las enzimas de origen vegetal que clivan carotenoides, como se comprobó al alinear los motivos sobre la estructura de la VP14.

La CCD4b2 y la CCD4d fueron predichas como duplicaciones de la CCD4b1 y CCD4c, respectivamente. Estas duplicaciones en tándem explican la mayor variabilidad de las mismas debido a la menor presión selectiva y la pérdida de un motivo fuertemente conservado en la CCD4d que afecta a una de las histidinas conservadas en la esfera de coordinación del Fe^{+2} sugiere la pérdida de su capacidad catalítica.

Se ha predicho la presencia de péptidos de destinación al cloroplasto (cpTP) en la mayoría de las CCOs de *Citrus clementina* con las excepciones de las CCD1 y CCD7 en las cuales no se ha podido identificar por ninguno de los métodos empleados.

El análisis de la región N-terminal de los miembros de la CCD4 que abarca aproximadamente unos ~100 aminoácidos muestran una región de alta variabilidad con un alto grado de desorden y estas posiciones marcan una clara tendencia a coevolucionar. En esta región se encuentra la señal cpTP y una vez que la enzima es traslocada al interior del cloroplasto duramente la maduración es eliminada de la estructura (Lee y Hwang, 2018).

La predicción de la estructura secundaria muestra un claro predominio de las láminas- β a lo largo de la secuencia, resultado consistente con la arquitectura de esta familia de proteínas cuya arquitectura se describe como β -propeller. Las α -hélices aparecen al inicio de la secuencia luego de la región N-teminal desestructurada (> 100 aa) y las mismas fueron identificadas como probables dominios de interacción con la membrana tilacoidal. Este dominio de α -hélices se postula como el domo hidrofóbico de interacción con las membranas consistente con lo observado por Messing et al., (2010) en la VP14 de *Zea mays*.

Finalmente, el análisis filogenético de las CCD4s demostró la cercana relación entre la CCD4a, CCD4c y CCD4d mientras que la CCD4b1 y CCD4b2 fueron las más lejanamente emparentadas con los demás miembros CCD4 de *Citrus clementina*.

REFERENCIAS

- Ahrazem, O., Gómez-Gómez, L., Rodrigo, M. J., Avalos, J., & Limón, M. C. (2016). Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: Features and functions. International Journal of Molecular Sciences, 17(11). <u>https://doi.org/10.3390/ijms17111781</u>
- Ahrazem, O., Trapero, A., Gómez, M. D., Rubio-Moraga, A., & Gómez-Gómez, L. (2010). Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: A deeper study in Crocus sativus and its allies. Genomics, 96(4), 239–250. <u>https://doi.org/https://doi.org/10.1016/j.ygeno.2010.07.003</u>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. <u>https://doi.org/10.1016/S0022-2836(05)80360-2</u>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. Nucleic Acids Research, 37(SUPPL. 2), 202–208. <u>https://doi.org/10.1093/nar/gkp335</u>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25(15), 1972–1973. <u>https://doi.org/10.1093/bioinformatics/btp348</u>
- Dunn, S. D., Wahl, L. M., & Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics, 24(3), 333–340. <u>https://doi.org/10.1093/bioinformatics/btm604</u>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. <u>https://doi.org/10.1093/nar/gkh340</u>

- Emanuelsson, O., Nielsen, H., Brunak, S., & von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology, 300(4), 1005–1016. <u>https://doi.org/10.1006/jmbi.2000.3903</u>
- Emanuelsson, O., Nielsen, H., & von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science: A Publication of the Protein Society, 8(5), 978–984. <u>https://doi.org/10.1110/ps.8.5.978</u>
- Gerts, E. M., Yu, Y. K., Agarwala, R., Schäffer, A. A., & Altschul, S. F. (2006). Compositionbased statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biology, 4, 1–14. <u>https://doi.org/10.1186/1741-7007-4-41</u>
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2011). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research, 40(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research, 40(D1), 1178–1186. https://doi.org/10.1093/nar/gkr944
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–321. <u>https://doi.org/10.1093/sysbio/syq010</u>
- Huang, F.-C., Molnár, P., & Schwab, W. (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. Journal of Experimental Botany, 60(11), 3011– 3022. <u>https://doi.org/10.1093/jxb/erp137</u>
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology, 292(2), 195–202. https://doi.org/10.1006/jmbi.1999.3091
- Jones, D. T., & Cozzetto, D. (2015). DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics, 31(6), 857–863. https://doi.org/10.1093/bioinformatics/btu744
- Kato, M. (2012). Mechanism of carotenoid accumulation in citrus fruit. Journal of the Japanese Society for Horticultural Science, 81(3), 219–233. <u>https://doi.org/10.2503/jjshs1.81.219</u>

- Kato, M., Matsumoto, H., Ikoma, Y., Okuda, H., & Yano, M. (2006). The role of carotenoid cleavage dioxygenases in the regulation of carotenoid profiles during maturation in citrus fruit. Journal of Experimental Botany, 57(10), 2153–2164. https://doi.org/10.1093/jxb/erj172
- Kloer, D. P., & Schulz, G. E. (2006). Structural and biological aspects of carotenoid cleavage. Cellular and Molecular Life Sciences, 63(19–20), 2291–2303. https://doi.org/10.1007/s00018-006-6176-6
- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics, 30(22), 3276–3278. <u>https://doi.org/10.1093/bioinformatics/btu531</u>
- Lashbrooke, J. G., Young, P. R., Dockrall, S. J., Vasanth, K., & Vivier, M. a. (2013). Functional characterisation of three members of the Vitis vinifera L. carotenoid cleavage dioxygenase gene family. BMC Plant Biology, 13, 156. <u>https://doi.org/10.1186/1471-2229-13-156</u>
- Lee, D. W., & Hwang, I. (2018). Evolution and design principles of the diverse chloroplast transit peptides. Molecules and Cells, 41(3), 161–167. <u>https://doi.org/10.14348/molcells.2018.0033</u>
- Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution, 34(9), 2422–2424. <u>https://doi.org/10.1093/molbev/msx149</u>
- Liang, M.-H., Zhu, J., & Jiang, J.-G. (2018). Carotenoids biosynthesis and cleavage related genes from bacteria to plants. Critical Reviews in Food Science and Nutrition, 58(14), 2314–2333. <u>https://doi.org/10.1080/10408398.2017.1322552</u>
- Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., Wahyudi, A., Motohashi, R., & Kato, M. (2013). Enzymatic formation of β-citraurin from β-cryptoxanthin and zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. Plant Physiology, 163(2), 682–695. <u>https://doi.org/10.1104/pp.113.223297</u>
- Martin, L. C., Gloor, G. B., Dunn, S. D., & Wahl, L. M. (2005). Using information theory to search for co-evolving residues in proteins. Bioinformatics, 21(22), 4116–4124. <u>https://doi.org/10.1093/bioinformatics/bti671</u>
- Messing, S. A. J., Mario Amzel, L., Gabelli, S. B., Echeverria, I., Vogel, J. T., Guan, J. C., Tan, B. C., Klee, H. J., & McCarty, D. R. (2010). Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. Plant Cell, 22(9), 2970– 2980. <u>https://doi.org/10.1105/tpc.110.074815</u>
- Murphy, M., Brown, G., Wallin, C., Tatusova, T., Pruitt, K. D., Murphy, T., & Maglott, D. R. (2016). Gene Help: Integrated Access to Genes of Genomes in the Reference Sequence Collection.
- Nugent, T., & Jones, D. T. (2009). Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics, 10, 1–11. <u>https://doi.org/10.1186/1471-2105-10-159</u>

- Nugent, T., & Jones, D. T. (2012). Detecting pore-lining regions in transmembrane protein sequences. BMC Bioinformatics, 13(1). <u>https://doi.org/10.1186/1471-2105-13-169</u>
- Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., & Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiology, 142(3), 1193–1201. <u>https://doi.org/10.1104/pp.106.087130</u>
- Rodrigo, M. J., Alquézar, B., Alís, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461– 4478. https://doi.org/10.1093/jxb/ert260
- Rodrigo, M. J., Alquézar, B., Alós, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461– 4478. https://doi.org/10.1093/jxb/ert260
- Rubio-Moraga, A., Rambla, J. L., Fernández-De-Carmen, A., Trapero-Mozos, A., Ahrazem, O., Orzez, D., Granell, A., & Gimez-Gimez, L. (2014). New target carotenoids for CCD4 enzymes are revealed with the characterization of a novel stress-induced carotenoid cleavage dioxygenase gene from Crocus sativus. Plant Molecular Biology, 86(4–5), 555–569. <u>https://doi.org/10.1007/s11103-014-0250-5</u>
- Schneider, T. D., & Stephens, R. M. (1990). Sequence logos: A new way to display consensus sequences. Nucleic Acids Research, 18(20), 6097–6100. https://doi.org/10.1093/nar/18.20.6097
- Sethi, A., O'Donoghue, P., & Luthey-Schulten, Z. (2005). Evolutionary profiles from the QR factorization of multiple sequence alignments. Proceedings of the National Academy of Sciences of the United States of America, 102(11), 4045–4050. https://doi.org/10.1073/pnas.0409715102
- Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N., & Taylor, J. M. (2017). LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Scientific Reports, 7(1), 44598. <u>https://doi.org/10.1038/srep44598</u>
- Strait, B. J., & Dewey, T. G. (1996). The Shannon information entropy of protein sequences. Biophysical Journal, 71(1), 148–155. <u>https://doi.org/10.1016/S0006-3495(96)79210-X</u>
- The UniProt Consortium. (2018). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506–D515. <u>https://doi.org/10.1093/nar/gky1049</u>

- Vallabhaneni, R., Bradbury, L. M. T., & Wurtzel, E. T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. Archives of Biochemistry and Biophysics, 504(1), 104–111. https://doi.org/10.1016/j.abb.2010.07.019
- Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., King, G. J., & Liu, K. (2015). Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in Brassica species. New Phytologist, 206(4), 1513–1526. <u>https://doi.org/10.1111/nph.13335</u>



Figura S1. Esfera de coordinación del Fe⁺² (esfera rosa). Histidinas en varillas observadas en cristales VP14 de Zea mays (PBD: <u>3NPE</u> en verde), ACO de Synechocystis sp. (PDB: <u>2BIW</u> en naranja) y RPE65 de Bos taurus (PDB: <u>3FSN</u> en amarillo).



Figura S2. Estructura secundaria predicha para CCD4a. Láminas-β (amarillo) y α-hélices (rosa)



Figura S3. Estructura secundaria predicha para CCD4b1. Láminas- β (amarillo) y α -hélices (rosa)



Figura S4. Estructura secundaria predicha para CCD4b2. Láminas- β (amarillo) y α -hélices (rosa)



Figura S5. Estructura secundaria predicha para CCD4c. Láminas-β (amarillo) y α-hélices (rosa)



Figura S6. Estructura secundaria predicha para CCD4d. Láminas-β (amarillo) y α-hélices (rosa)



Figura S7. Información del alineamiento múltiple de las secuencias de CCD4s de origen vegetal filtradas por columnas (posiciones de aminoácidos) de CCD4a de *Citrus clementina* a) Ocupación por residuos del alineamiento de las CCDs de origen vegetal. Región de menor ocupación indicada mediante la barra púrpura. b) Entropía de Shannon de las posiciones del alineamiento. Región de mayor variabilidad en el alineamiento indicada mediante la barra púrpura. La estructura secundaría predicha se indica en la barra inferior: α-hélice (rojo), lámina-β (amarillo) y *loops* (negro).

 Tabla S1. Identificadores de acceso e información de las secuencias utilizadas para la construcción del árbol filogenético de secuencias de aminoácidos del tipo CCD4 de *Citrus clementina*.

N°	Nombre de la especie	Etiqueta	Identificador
1	Arabidopsis thaliana	AtCCD4	NP_193652.1
2	Brassica napus	BnCCD4	XP_013656050.1
3	Citrus unshiu	CuCCD4	GAY43306.1
4	Malus domestica	MdCCD4	ABY47995.1
5	Osmanthus fragrans	OfCCD4	ABY60887.1
6	Rosa x damascena	RdCCD4	ABY60886.1
7	Solanum tuberosum	StCCD4	NP_001275103.1
8	Chrysanthemum x morifolium	CmorCCD4a	ABY60885.1
9	Citrus clementina	CcCCD4a	XP_006437482.1
10	Crocus sativus	CrosCCD4a	ACD62476.1
11	Vitis vinifera	VvCCD4a	AGT63321.1
12	Citrus clementina	CcCCD4b1	XP_006424046.1
13	Citrus clementina	CcCCD4b2	XP_006424047.2
14	Crocus sativus	CrosCCD4b	ACD62477.1
15	Vitis vinifera	VvCCD4b	AGT63322.1
16	Citrus clementina	CcCCD4c	XP_006430912.1
17	Crocus sativus	CrosCCD4c	AEO50759.1
18	Citrus clementina	CcCCD4d	ESR44157.1

 Tabla S2-1.
 Identificadores de acceso e información de las secuencias utilizadas para la construcción del árbol filogenético de secuencias de aminoácidos de CCDs de origen vegetal.

N°	Nombre de la especie	Etiqueta	Identificador
1	Solanum tuberosum	StCCD7	XP_006347925.1
2	Sesamum indicum	SiCCD7	XP_011094476.1
3	Daucus carota	DcCCD7	XP_017228448.1
4	Trema orientale	ToCCD7	PON85588.1
5	Cicer arietinum	CaCCD7	XP_004513935.1
6	Tripterygium wilfordii	TwCCD7	XP_038684300.1
7	Arachis hypogaea	AhCCD7	RYR05300.1
8	Prosopis alba	PaCCD7	XP_028753454.1
9	Hibiscus syriacus	HsCCD7	XP_039040729.1
10	Carica papaya	CpCCD7	XP_021903305.1
11	Ricinus communis	RcCCD7	EEF50298.1
12	Corchorus olitorius	CoCCD7	OMP00710.1
13	Citrus clementina	CcCCD7	XP_006425527.2
14	Rhamnella rubrinervis	RrCCD7	KAF3440065.1
15	Malus domestica	MdCCD7	XP_028961438.1
16	Tripterygiu mwilfordii	TwCCD8	KAF5748266.1
17	Vigna radiata	VrCCD8	XP_014500485.2
18	Populus deltoides	PdCCD8	KAF9827265.1
19	Citrus clementina	CcCCD8	XP_006450690.2
20	Quercus lobata	QINCED2	XP_030972679.1
21	Citrus clementina	CcNCED2	XP_006449708.1
22	Durio zibethinus	DzNCED2	XP_022777272.1
23	Gossypium barbadense	GbNCED2	PPR90777.1
24	Nyssa sinensis	NsNCED3	KAA8524088.1
25	Citrus clementina	CcNCED3	XP_006442236.1
26	Populus deltoides	PdNCED3	KAF9840133.1
27	Vitis riparia	VitrNCED6	XP_034685154.1
28	Syzygium oleosum	SoNCED6	XP_030462960.1
29	Cucumis sativus	CucNCED6	XP_004150345.1
30	Brassica rapa	BrNCED6	XP_033130359.1
31	Morus notabilis	MnNCED6	EXB76918.1
32	Cannabis sativa	CansNCED6	KAF4370844.1
33	Rhododendron simsii	RsNCED6	KAF7112535.1
34	Jatropha curcas	JcNCED6	XP_012092670.1
35	Prunus yedoensis	PyNCED6	PQP99424.1
36	Rosa chinensis	RcNCED6	XP_024195455.1
37	Hibiscus syriacus	HsNCE6D	XP_039050921.1
38	Morella rubra	MrNCED6	KAB1203761.1

Tabla S2-2. (Continuación). Identificadores de acceso e información de las secuencias utilizadaspara la construcción del árbol filogenético de secuencias de aminoácidos de CCDsde origen vegetal.

N°	Nombre de la especie	Etiqueta	Identificador
39	Citrus clementina	CcNCED6	XP_006420602.2
40	Manihot esculenta	MeNCED6	XP_021594070.1
41	Salix brachista	SbNCED6	KAB5564874.1
42	Citrus clementina	CcCCD1	XP_024041179.1
43	Rosa chinensis	RcCCD1	XP_024188836.1
44	Bixa orellana	BoCCD4b	AMJ39500.1
45	Bixa orellana	BoCCD4c	AMJ39501.1
46	Bixa orellana	BoCCD4d	AMJ39502.1
47	Populus trichocarpa	PtZCD	XP_024464565.1
48	Citrus clementina	CcCCD4b1	XP_006424046.1
49	Citrus clementina	CcCCD4b2	XP_006424047.2
50	Theobroma cacao	TcZCD	XP_017982919.1
51	Herrania umbratica	HuCCD4	XP_021278762.1
52	Corchorus capsularis	CorcCCD4a	OMO51635.1
53	Gossypium raimondii	GrCCD4	XP_012485195.1
54	Durio zibethinus	DzCCD4	XP_022753724.1
55	Nyssa sinensis	NsCCD4a	KAA8519447.1
56	Nyssa sinensis	NsCCD4b	KAA8544388.1
57	Rhododendron simsii	RhosCCD4	KAF7138155.1
58	Actinidia rufa	ArCCD4	GFZ02936.1
59	Vitis vinifera	VvCCD4	XP_002269538.2
60	Nelumbo nucifera	NnCCD4a	XP_010251938.1
61	Cephalotus follicularis	CfCCD4a	GAV84147.1
62	Nelumbo nucifera	NnCCD4b	XP_010275133.1
63	Corchorus capsularis	CorcCCD4b	OMO88378.1
64	Raphanus sativus	RsCCD4	XP_018447153.1
65	Macleaya cordata	McCCD4	OVA01701.1
66	Vigna angularis	VaCCD4	KOM45343.1
67	Cephalotus follicularis	CfCCD4b	GAV77659.1
68	Nelumbo nucifera	NnCCD4c	XP_010275132.1
69	Manihot esculenta	NnCCD4d	XP_021595193.1
70	Citrus clementina	CcCCD4a	XP_006437482.1
71	Prunus dulcis	PrundCCD4	VVA13456.1
72	Gossypium barbadense	GbCCD4	PPS01874.1
73	Theobroma cacao	TrCCD4	EOY25343.1
74	Citrus clementina	CcCCD4d	ESR44157.1
75	Citrus clementina	CcCCD4c	XP 006430912.1

CAPÍTULO III

MODELADO MOLECULAR Y SIMULACIÓN DE LA CCD4a Y CCD4b DE Citrus clementina

INTRODUCCIÓN

En los citrus, el color de la cáscara y el jugo está asociado por los consumidores con la calidad de la fruta y es uno de los principales parámetros para su aceptación. Las frutas cítricas constituyen una de las principales producciones frutales del mundo con más de 120 millones de toneladas (FAO, 2017). Por este motivo las investigaciones enfocadas en la bioquímica y bases moleculares de la coloración de las frutas cítricas han ganado interés en los últimos años y han constituido un área de investigación activa. (Rodrigo et al., 2013).

Se ha reportado una serie de CCDs de la subfamilia CCD4 en citrus responsables de la generación de apocarotenoides C_{30} (8'-apo- β -carotenal y β -citraurina) que proveen la característica pigmentación naranja-rojiza a las cáscaras de algunas variedades de mandarinas y naranjas (Ma et al., 2013b).

La actividad de muchas CCD4 provenientes de plantas han mostrado que el sustrato preferente es el β -caroteno y el punto de corte preferencial es la posición 9=10 (9'=10') produciendo β -ionona y 10'-apo- β -carotenal (Huang et al., 2009; Gonzalez-Jorge et al., 2013; Lätari et al., 2015) (Fig. 1)



Figura 1. Corte en la posición 9=10 (9'=10') de las CCD4 de plantas del β -caroteno y sus productos β -ionona y 10'-apo- β -carotenal. Los posibles sitios de corte se señalan con flecha y con la barra roja.

La CC4b1 (en adelante CCD4b) de citrus fue el primer caso reportado de una dioxigenasa capaz de cortar carotenoides C₄₀ en la posición de doble enlace 7=8 o 7'=8'. Se han sugerido como potenciales sustratos de la enzima CCD4b en citrus la zeaxantina y la β -criptoxantina, carotenoides di y monohidroxilados, respectivamente (Rodrigo et al., 2013). Ambas pertenecen al grupo de las

xantófilas sintetizadas en frutas cítricas de color naranja durante la maduración. Ensayos *in vitro* han observado una ligera preferencia en la posición 7'=8' cercana al anillo hidroxilado de la β -criptoxantina, produciendo 8'-apo- β -carotenal y β -citraurina (Kato, 2012) (Fig. 2)



Figura 2. (a) Zeaxantina y los dos productos resultantes de la escisión en el doble enlace 7=8:
 β-citraurina y 3-hidroxi-β-ciclocitral. (b) La estructura 8'-apo-β-carotenal, producto apocarotenoide cuando el sustrato es β-caroteno o β-criptoxantina con el anillo hidroxilado dentro del sitio activo.

Ensayos *in vitro* de las enzimas CCD4b con otros carotenoides como el β -caroteno y α caroteno, luteína y licopeno muestran que el β -caroteno puede ser sustrato produciendo 8'-apo- β - carotenal, mientras que la luteína y el α -caroteno fueron solamente clivados del lado del anillo β . La actividad de la CCD4b en carotenos lineales fue probada en dos trabajos utilizando como sustrato el licopeno, para estos casos, se detectó actividad nula o muy baja (Ma et al., 2013; Rodrigo et al., 2013).

En resumen, para la mayoría de las CCD4 provenientes de plantas analizadas mediante ensayos *in vitro* e *in vivo* apuntan al β -caroteno como el sustrato de preferencia y como puntos de corte a las posiciones de doble enlace 9=10 y/o 9'=10'. De cualquier manera, las CCD4 provenientes de especies de citrus constituyen una relevante excepción en cuanto a la preferencia por el sustrato y la posición de clivaje, y que puede constituir una clave en el entendimiento de la selectividad de las enzimas CCD4. (Ohmiya et al., 2006; Bruno et al., 2015)

La estrategia de estudio *in silico* se aplica en la construcción de modelos de estructura tridimensional de las enzimas y de los complejos enzima-sustrato como herramientas de racionalización de resultados experimentales, explorar nuevas hipótesis y realizar predicciones. Los conocimientos de bioinformática estructural en términos de *docking* y simulaciones de dinámica molecular pueden dar una visión más profunda de las fuerzas intermoleculares que gobiernan la interacción de los complejos proteína-ligando. Esta información puede ayudar a entender las bases moleculares de la reactividad y selectividad de las enzimas CCD4 con sus sustratos carotenoides, y conjeturar sobre su influencia en el color y otras características organolépticas *in vivo* de las especies de citrus. Ya que el comportamiento estructural y dinámico de estas enzimas y sus sustratos permanecen desconocidos, analizarlos son los objetivos de esta sección del trabajo.

En este capítulo presentamos las estructuras tridimensionales de los subtipos a y b de la subfamilia CCD4, modelados por homología y validados por dinámica molecular. Adicionalmente, los complejos moleculares de estas enzimas fueron construidos y se estudió el modo de unión y sus interacciones moleculares con 7 carotenoides distintos (α -caroteno, β -caroteno, licopeno, zeaxantina, luteína, β -criptoxantina y 3-OH-8'- β -apocaroteno]).

El presente capítulo fue publicado bajo el título "An in silico study of the citrus dioxygenases CCD4 family substrates" en la revista Journal of Biomolecular Structure and Dynamics. (Vega-Teijido et al., 2018). El artículo completo está disponible en el Anexo 1.

OBJETIVOS

- Obtener un modelo de la estructura tridimensional de la CCD4a y CCD4b1 mediante modelado por homología y validar mediante Dinámica Molecular.
- Evaluar los requerimientos estructurales para la interacción receptor-ligando mediante *docking* molecular y analizar sus interacciones puntuales.
- Realizar simulaciones de Dinámica Molecular de los complejos y evaluar las interacciones moleculares que gobiernan la unión de complejo receptor-ligando.

MATERIALES Y MÉTODOS

SOFTWARE, HARDWARE, CAMPO DE FUERZA Y PARAMETRIZACIÓN

Los modelos y simulación se realizaron en una estación de trabajo con un procesador *Quad Core Intel i7.* Los cálculos se realizaron con el programa *Molecular Operating Environment* (MOE) versión 2013.10 (Chemical Computing Group Inc, 2014), en una suite Linux (OpenSUSE 12.3) y Windows 7.

El campo de fuerza seleccionado para la simulación fue Amber96 implementado en MOE. El programa MOE es capaz de parametrizar a los ligandos mediante un algoritmo de reconocimiento de patrones, identificando a los átomos del ligando y asignando los parámetros dependiendo del elemento (sea el caso C, H, O), la conectividad, la valencia y la geometría. Finalmente, las cargas atómicas parciales fueron asignadas siguiendo la definición de Amber96. Como resultado, las interacciones no covalentes entre el ligando y el receptor se evaluaron a lo largo de la dinámica molecular.(Weiner et al., 1986; Wang et al., 2000)

MODELADO POR HOMOLOGÍA DE LAS PROTEÍNAS

La selección del *template* se definió utilizando la estructura de la CCD con mayor similitud de secuencia que contaba con un carotenoide en el sitio activo. La enzima oxigenasa de clivaje de apocarotenoides (ACO) es una estructura cristalográfica proveniente de *Synechocystis sp.* expresada en *Escherichia coli* (PDB ID: <u>2BIW</u>). La ACO es una enzima capaz de tomar como sustrato 8'-apo- β -carotenol (3ON) y clivar en la posición 15=15' para producir retinal. La resolución del cristal es de 2,39 Å, con un R-factor = 0,182 y R-free = 0,224. (EC <u>1.13.11.75</u>) (Kloer et al., 2005)

Modelos por homología de la CCD4a y CCD4b fueron construidos luego de un alineamiento con el *template*, se reemplazaron las cadenas laterales, reteniendo las coordenadas espaciales de la cadena principal de la ACO, el cofactor Fe^{+2} y el apocarotenoide 8'-apo- β -carotenol (3ON).

Debido al nivel porcentaje de similitud de las secuencias entre las CCD4 y el *template* (37,1% con la CCD4a), los modelos por homología de la CCD4a resultaron una aproximación preliminar. Estos modelos fueron refinados y sometidos a un proceso de validación. En la primera etapa se procedió a una inspección de la esfera de coordinación del Fe⁺², del sitio activo y de la posición del apocarotenoide cocristalizado, definiendo el sitio activo como el sitio ocupado por el apocarotenoide cocristalizado, no se observaron *clashes*. La siguiente etapa consistió en un refinamiento del modelo por minimización de su energía. Se procedieron a varias optimizaciones locales y finalmente a una dinámica molecular de corta duración, para asegurar la estabilidad del modelo. Para modelar la CCD4b se utilizó como *template* el modelo final de la CCD4a siguiendo el mismo protocolo.

REFINAMIENTO Y DINÁMICA MOLECULAR

Se construyó una caja de agua con solvente explícito definiendo 6 Å de margen entre el átomo más externo y el borde de la caja en cada eje cartesiano, resultando una caja con las dimensiones: $90 \times 80 \times 75$ Å conteniendo un total de 12.000 moléculas de agua. Se utilizaron como contraiones átomos de Na⁺ y Cl⁻ a fin de asegurar la neutralidad global de cargas del sistema. Posteriormente se realizó una etapa de minimización de energía hasta un gradiente de 1 kcal/mol/Å² y una dinámica molecular para equilibrar termodinámicamente todo el sistema a 300 K de temperatura constante con una duración de 5 ns en condiciones periódicas con un *cutoff* para las interacciones de largo alcance de 10 Å. El tamaño de los pasos de la simulación (*time step*) fue de 2 femtosegundos (fs), las moléculas de agua se consideraron como cuerpos rígidos, permitiendo su rotación y traslación a lo largo de la caja. Se utilizó el ensamble canónico NVT y para resolver las ecuaciones de movimiento, se utilizó la formulación Nosé-Poincaré-Andersen (NPA) (Bond et al., 1999). La estructura cristalográfica utilizada como *template* junto con su ligando cocristalizado fueron utilizadas como moléculas de prueba para poder comparar a los efectos de la validación, las trayectorias realizadas en las mismas condiciones con todos los modelos, todas ellas ejecutadas siguiendo el mismo protocolo.

PROCEDIMIENTO DE DOCKING: BASE DE DATOS DE LIGANDOS

Las geometrías de los ligandos (α -caroteno, β -caroteno, licopeno, zeaxantina, luteína, β criptoxantina y 3-OH-8'- β -apocarotenol) se construyeron manualmente, y refinadas mediante minimización de energía utilizando el campo de fuerza MMFF94x (Halgren, 1999). Posteriormente, debido al amplio espacio conformacional que pueden adoptar los carotenoides por sus múltiples enlaces simples rotables, se realizó una búsqueda conformacional construyendo una base de datos del mapa conformacional de cada uno de los ligandos que luego se utilizó en la fase de *docking*.

VALIDACIÓN DEL DOCKING

Un procedimiento inicial de *docking* de validación fue realizado para encontrar las mejores poses del apocarotenoide 3ON en el sitio activo del *template*, a modo de lograr obtener mediante el anclaje la pose observada en la estructura cristalográfica. En una primera etapa, se definió como receptor a toda la proteína, como ligando la base de datos del mapa conformacional de la 3ON y como sitio de unión al espacio ocupado por la esfera de 4,5 Å de la pose cristalográfica del 3ON. En la segunda etapa se evaluaron distintos métodos de *placement* y funciones de *score*, de manera de determinar el método más adecuado para este tipo de receptor.

El proceso de anclaje seleccionado incluyó al método de *placement* "Proxy triangle", la función de *score* "ASE", posteriormente una fase de refinamiento del ligando en el sitio utilizando el campo de fuerza Amber99 y finalmente una evaluación ulterior del score utilizando la función "London Δ G". Todas estas funciones se incluyen dentro del programa MOE.

Durante el *placement* se permitió a los enlaces rotables girar libremente, generando una colección de poses, siguiendo el método *Proxy triangle*. Este método fue desarrollado para manipular ligandos con muchos grados de libertad conformacional, como es el caso de los carotenoides. Luego se evalúan las poses con la función de score ASE, para luego almacenar las 30 mejores poses para cada ligando.

La función de score ASE (Chemical Computing Group Inc, 2014), evalúa la contribución entálpica de la energía libre de unión como una descripción del empaquetamiento de la molécula en el sitio, definiendo a partir de los pares de átomos ligando-receptor una suma de funciones gaussianas multiplicado por una constante de proporcionalidad, siguiendo la Ecuación 1:

Ecuación 1.
$$ASE = c * \sum R_i R_j e^{-d^2/2}$$

siendo,

 R_i los radios atómicos del ligando en Å,

 R_i los radios atómicos del receptor en Å,

d es la distancia entre los pares de átomos $R_i R_i$ en Å,

c es la constante de proporcionalidad igual a 0,146 kJ/mol.

La función London ΔG es una estimación más detallada de las contribuciones entálpicas de la energía libre de unión entre el ligando y el receptor (Chemical Computing Group Inc, 2014), descritas en la Ecuación 2 e involucra a los siguientes términos:

Ecuación 2. London
$$\Delta G = c + E_{flex} + \sum_{enlaces-H} C_{HB}f_{HB} + \sum_{enlaces-M} C_M f_M + \sum_{\acute{a}tomos-i} \Delta D_i$$

donde,

c es el promedio de la ganancia/pérdida de entropía rotacional y traslacional,

 E_{flex} es la energía debido a la pérdida de flexibilidad del ligando, calculado solamente por la topología del ligando,

 f_{HB} mide las imperfecciones geométricas de los enlaces de hidrógeno y toma valores en el rango [0-1],

 C_{HB} es la energía del enlace de hidrógeno ideal,

 f_M mide las imperfecciones geométricas de los enlaces metálicos y toma valores en el rango [0-1], C_M es la energía del enlace metálico ideal,

 D_i es la energía de desolvatación del átomo i.

La diferencia en la energía de desolvatación se calculó siguiendo la Ecuación 3:

Ecuación 3.
$$\Delta D_i = c_i R_i^3 \left\{ \iiint_{u \notin A \cup B} |u|^{-6} du - \iiint_{u \notin B} |u|^{-6} du \right\}$$

donde,

 $A \cup B$ son los volúmenes del ligando y/o receptor con el átomo i que corresponden al volumen B. R_i es el radio de solvatación del átomo i (usando el parámetro σ más 0.5 Å de la función de Van der Waals del campo de fuerza OPLS-AA)

 c_i es el coeficiente de desolvatación del átomo i.

Los valores de los coeficientes c, C_{HB} , C_M , C_i están determinados empíricamente de una base de datos con aproximadamente 400 de complejos cristalográficos de proteína-ligando para

los que se conocen los valores experimentales de *pKi*. El volumen está estimado usando el método de integración de Bohr generalizado.

Finalmente, dada la similitud estructural entre la CCD4a y la CCD4b, y aprovechando el desarrollo realizado para la estructura CCD4b, la posición inicial de los ligandos en la CCD4b se estableció alineando ambas estructuras y tomando la posición del carotenoide en el modelo CCD4a obtenido por *docking*.

DINÁMICA MOLECULAR DE LOS COMPLEJOS

La simulación de dinámica molecular se realizó siguiendo el protocolo de la sección de REFINAMIENTO Y DINÁMICA MOLECULAR. La simulación se dividió en tres etapas: calentamiento de 0 a 300 K en 100 ps, 900 ps adicionales a 300 K restringiendo la posición de los átomos de la cadena principal (*backbone*). Luego, se obtuvo una trayectoria de muestreo (sin ningún tipo de restricción) de 10 ns.

MEDIDAS DE LA ENERGÍA DE INTERACCIÓN

A lo largo de la simulación de dinámica molecular se evaluaron las interacciones entre los ligandos y el resto del sistema (receptor, solvente e iones), la misma se definió como Uab y se describe en la Ecuación 4:

Ecuación 4.
$$\langle Uab \rangle = \langle U - Ua - Ub \rangle$$

donde,

U = La energía potencial total del sistema. Ua = La energía potencial del receptor, el solvente e iones y sus interacciones (excluyendo al ligando).

Ub = La energía potencial interna del ligando.

ENTORNO DE INTERACCIONES DEL LIGANDO

Adicionalmente se generaron diagramas en 2D que resumen las componentes más importantes de la interacción entre los ligandos y sus receptores, listando los enlaces de hidrógeno, el entorno que establecen los residuos en la esfera de interacción con el ligando, sean estos ácidos, básicos o hidrofóbicos, así como también las interacciones aromáticas que se podrían presentar.

El proceso computacional utilizado consistió en una serie de etapas que se muestran en la Fig. 3 en la cual se muestra una estrategia que implica las etapas de modelación y posterior uso de los modelos para medir la interacción con diferentes carotenoides.



Figura 3. Resumen de las etapas del proceso de modelado, *docking* molecular, simulación de dinámica molecular y análisis de las trayectorias utilizadas en el estudio.

La etapa de modelación incluye modelado por homología, minimización de energía (refinamiento), dinámica molecular, análisis y selección del mejor modelo. Luego, en la etapa de *docking*, cada modelo seleccionado dio lugar a un modelo de cada carotenoide anclado a cada proteína. Luego, cada complejo fue validado mediante simulación de dinámica molecular en la cual se analizó la estabilidad y se estableció el tiempo a partir del cual se puede considerar que cada complejo CCD4-carotenoide ha alcanzado el equilibrio termodinámico. A partir de ese tiempo se establece la etapa de producción en equilibrio. Finalmente, una vez determinado el tiempo en el que los modelos se muestran equilibrados, se colectaron las estructuras y propiedades durante las etapas de producción de las simulaciones con el fin de realizar medidas de energía de interacción Uab, RMSD, distancias en la esfera de coordinación del Fe⁺² y posibles sitios de clivaje de cada carotenoide

RESULTADOS Y DISCUSION

ALINEAMIENTO DE SECUENCIAS Y SELECCIÓN DEL TEMPLATE

Los alineamientos de las secuencias muestran un porcentaje de identidad de 22,9% entre la CCD4a y la ACO y una similitud de 37,1% mientras que, para la CCD4b, se observa un 23,1% de identidad y 38,9% de similitud (Tabla 1).

Tabla 1. Porcentaje de identidad (diagonal inferior) / similitud (diagonal superior), para las secuencias de la CCD4a, CCD4b comparando con el *template* ACO. Para el alineamiento se utilizó la matriz de sustitución BLOSUM62, para abrir un gap = 7, extender un gap = 1, utilizando el programa MOE-Align.

Identidad/similitud	CCD4a	CCD4b	ACO
CCD4a	100,0	62,7	37,1
CCD4b	48,8	100,0	38,9
ACO (template)	22,9	23,1	100,0

Los valores observados son relativamente bajos, pero se espera una buena conservación de la estructura secundaria y terciaria en la superfamilia de las CCDs, debido a esto se consideró

aceptable tomar a la ACO como un *template* para modelado para proceder a la fase de modelado por homología. Adicionalmente, considerando el gran número de grados de libertad que poseen en su estructura los carotenoides, es de crucial importancia contar con una referencia cristalográfica del ligando en el sitio de unión. La estructura de la ACO cumple con estos requisitos, debido a esto se seleccionó como *template* frente a otros posibles candidatos.

MODELADO MOLECULAR

El modelo por homología de la CCD4b, incluyendo al ligando apocarotenoide y su estructura secundaria, así como la esfera de coordinación del Fe^{+2} en el sitio activo, están resaltados en la Fig. 4. La superposición de los cuatro modelos CCD4 iniciales (a, b, c) sobre el *template* de la ACO se muestra en la Fig. 5.



Figura 4. Diagrama de cinta de la cadena principal de la proteína CCD4b modelada antes y después de 10 ns de simulación de dinámica molecular. El apocarotenoide co-cristalizado en la estructura ACO se indica en esferas grises. Los átomos de oxígeno en el apocarotenoide se muestran en color rojo. El átomo de Fe⁺² se muestra en una esfera cian y algunos detalles del entorno catalítico se indican con varas.


Figura 5. Diagrama de cinta de la cadena principal de las proteínas CCD4 modeladas: CCD4a en rojo, CCD4b en amarillo y CCD4c en verde, superpuestos con el *template* ACO en rosa. En azul están representados: el átomo de Fe⁺² en el centro catalítico y el apocarotenoide co-cristalizado en la estructura de la ACO.

Las proteínas modeladas pueden describirse como un β -propeller formada por siete láminas- β arregladas en un eje formando una estructura toroidal, análoga a lo descrito para el *template*. Cuatro residuos de histidina altamente conservadas se encuentran formando la esfera de coordinación de la Fe⁺². La estructura terciaria contiene una serie de cadenas antiparalelas que están conectadas por *loops*. Las láminas- β forman un cilindro espiralizado como la parte de una estructura rígida, mientras que los *loops* juntos con algunas α -hélices cortas conforman la parte flexible de la estructura. (Kloer y Schulz, 2006)

El sitio activo conformado por la esfera de coordinación de la estructura de la CCD4a modelada y del *template* se muestran en la Fig. 6.



Figura 6. Esfera de coordinación del Fe⁺² (esfera cian). Las moléculas de agua están marcadas con un "W". (a) Esfera de coordinación del *template* ACO. (b) Esfera de coordinación del modelo CCD4a.

RESULTADOS OBSERVADOS EN EL DOCKING

En la Tabla 2 se muestran los 5 mejores scores de *docking* de las mejores poses correspondientes a conformaciones de los tres carotenoides estudiados con CCD4a. Todos estos valores se encuentran en el rango de -85 a -65 kJ/mol, lo que resulta razonable por ser ligandos semejantes, si bien la función de score utilizada es lo suficientemente sensible como para detectar las bajas diferencias en la polaridad y la formación de enlaces de hidrógeno de estos tres ligandos.

Poses	Zeaxantina	β-criptoxantina	β-caroteno	
1	-85,14	-73,43	-67,40	
2	-82,30	-73,30	-66,19	
3	-81,13	-72,84	-65,40	
4	-80,25	-72,51	-64,94	
5	-78,83	-71,76	-64,73	

Tabla 2. Valores observados del score London ΔG en el procedimiento de *docking* para la Zeaxantina, β -criptoxantina y β -caroteno en el sitio activo de la CCD4a

El score logrado se puede asociar con el número de grupos hidroxilos ubicados en los anillos de los ligandos, donde se observa que ambos siguen la misma tendencia: zeaxantina > β -criptoxantina > β -caroteno. Estos resultados confirman la importancia de los patrones de enlaces de hidrógeno entre los carotenoides y las oxidasas.

Una estimación más precisa de las energías de interacción entre los ligandos y el receptor fue calculada con la función Uab y en la Tabla 3 se compararon con los resultados experimentales. En las dinámicas moleculares, los valores de la Uab reflejan la capacidad de establecer los enlaces de hidrógenos de los ligandos, siendo correlativo al número de grupos OH presentes: Luteína y Zeaxantina > β -criptoxantina > (α , β)-caroteno y licopeno.

Por otro lado, en la misma Tabla 3 se informan las distancias entre el Fe^{+2} y los átomos de carbono de los ligandos mostrándose que los dobles enlaces susceptibles a ser clivados corresponderían a las posiciones 7'=8' y 9'=10'.

Tabla 3. Distancias (en Å) entre el centro metálico Fe⁺² de la CCD4b y los átomos de carbono 15, 15', 10', 9', 8' y 7' en los correspondientes carotenoides estudiados. Coeficiente de partición octanol/agua (logP_{o/w}), energía de interacción del ligando Uab (kJ/mol) y bioactividad reportada por la literatura.

CCD4b	15	15'	10'	9'	8'	7'	logP _{w/o}	Uab (kJ/mol)	Bioactividad(+)
Luteína	11,70	12,48	7,05	6,19	5,70	5,82	8,77	-450,53	Sustrato*
Zeaxantina	10,90	9,68	5,51	5,38	5,97	6,13	8,57	-407,73	Sustrato
β -criptoxantina	11,68	11,16	5,24	5,81	5,28	6,22	8,65	-339,11	Sustrato
Licopeno	11,87	10,75	4,84	4,49	5,27	5,47	17,65	-254,43	Baja actividad
a-caroteno	12,84	11,69	7,45	8,17	7,65	8,18	17,28	-264,81	Sustrato*
β -caroteno	12,91	11,92	8,02	8,53	7,68	8,19	17,44	-253,17	Sustrato*

*Compuestos presentes solamente en frutas cítricas inmaduras.

+ (Rodrigo et al., 2013)

Con fines comparativos, en la Tabla 4 se presentan las distancias entre el Fe^{+2} y los átomos de carbono de la CCD4a y se observa que son ligeramente mayores que las observadas para CCD4b (Tabla 3), aun siendo distancias adecuadas para un clivaje mediado por Fe-O₂.

Tabla 4. Distancias (en Å) entre el centro metálico Fe⁺² de la CCD4a y los átomos de carbono 15, 15', 10', 9', 8' y 7' en los correspondientes carotenoides estudiados. Coeficiente de partición octanol/agua (logPo/w), energía de interacción del ligando Uab (kJ/mol) y biactividad reportada por la literatura.

CCD4a	15	15'	10'	9'	8'	7'	logP _{w/o}	Uab (kJ/mol)	Bioactividad(+)
Zeaxantina	11,32	10,93	8,11	7,23	7,95	7,69	8,57	-424,76	Sustrato
β -criptoxantina	9,80	8,02	7,85	7,07	7,70	7,48	8,65	-322,08	Sustrato
β-caroteno	12,30	11,62	8,14	6,90	6,67	7,75	17,44	-247,02	Sustrato*

*Compuestos presentes solamente en frutas cítricas inmaduras.

+ (Rodrigo et al., 2013)

En la Tabla 4 se muestran los valores de la Uab para tres ligandos estudiados en ambas enzimas CCD4, con el objetivo de compararlas. En ambos casos, la diferencia en los valores de la Uab entre la zeaxantina y la β -criptoxantina es aproximadamente de 103 kJ/mol y entre la zeaxantina y el β -caroteno es de 178 kJ/mol.

En las Tablas 3 y 4 se presentaron los valores del logaritmo del coeficiente de partición octanol/agua (logP_{o/w}) como referencia sobre las características hidrofílicas/hidrofóbicas de los ligandos. En este caso todos los ligandos se pueden clasificar como hidrofóbicos, tomando como referencia que el logP_{o/w} de la molécula de agua es de 3,34 y que los valores positivos mayores corresponden a moléculas hidrofóbicas. Se observan diferencias significativas en las xantofilas que se presentan como menos hidrofóbicas que los carotenos por poseer hidroxilaciones en sus ciclos.

En la Tabla 5 se presentan los aminoácidos detectados y el número de moléculas de agua en la esfera de coordinación del átomo de Fe^{+2} para los distintos complejos proteína-ligando estudiados, tomando como distancia de corte 4,5 Å, por ser esta una distancia en la cual se conforman la mayoría de las interacciones no enlazantes. Como se puede observar, el Fe^{+2} mantienen los contactos con los cuatro residuos His y dos moléculas de aguas alineadas, que se encuentran en la dirección del ligando. En el caso los complejos CCD4, durante la dinámica molecular, el Asp268 entra en contacto con la esfera de coordinación del Fe⁺², cancelando parte de la carga positiva.

Complejos	His218	His267	Asp268	His333	His513	H ₂ O
CCD4b - Zeaxantina	Х	Х	Х	Х	Х	2
CCD4b - β -criptoxantina	Х	Х	Х	Х	Х	2
CCD4b - β-caroteno	Х	Х	Х	Х	Х	2
CCD4a - Zeaxantina	Х	Х	Х	Х	Х	2
CCD4a - β -criptoxantina	Х	Х	Х	Х	Х	2
CCD4a - β-caroteno	Х	Х	Х	Х	Х	2
ACO-30N	Х	Х		Х	Х	1

Tabla 5. Residuos aminoácidos detectados en la esfera de 4,5 Å alrededor del centro metálico Fe⁺² y número de moléculas de agua detectadas

Para completar el análisis, se presenta a continuación una visión gráfica en dos dimensiones (2D) de la interacción entre ligandos y CCD4a (Fig. 7).

En la Fig. 7, se muestran los gráficos de interacción 2D entre ligandos y el receptor de los complejos CCD4a con β -caroteno, β -criptoxantina y zeaxantina. Se puede observar la importancia de las interacciones hidrofóbicas con el sitio de unión, representadas por nubes esféricas azules.

Es de notar que los enlaces de hidrógenos presentes en este diagrama están todos mediados por moléculas de agua en el interior del sitio de unión, lo cual indica que el solvente está jugando un rol de mediador con la enzima.



Figura 7. Gráfico de interacción 2D de los complejos de CCD4a con β -caroteno, β -criptoxantina y zeaxantina obtenidos con el módulo *Ligand-Interaction* del programa MOE. Los residuos que hacen contactos se muestran como círculos en colores correspondientes a la hidrofobicidad / hidrofobicidad (rosa= residuos polares; verde = residuos hidrofóbicos). Las áreas azules representan la accesibilidad del solvente y la línea punteada representa el sitio de unión del ligando. En el caso de las xantófilas β -criptoxantina y zeaxantina, Se representan la red de enlaces de hidrógeno formada con las moléculas de agua dentro del sitio de unión, que juegan un papel de puente con los residuos de la enzima.

Finalmente, para completar la descripción de la evolución de la estructura de los diferentes complejos en una situación en la que se simulan las mismas condiciones termodinámicas que en los procesos biológicos/bioquímicos observados, se hizo uso de las trayectorias de dinámica molecular y se evaluó la desviación cuadrática media (RMSD) para dos muestras seleccionadas: una trayectoria duplicada de 10 ns cada una del complejo CCD4a-zeaxantina, y una simulación de 20 ns de trayectoria del complejo CCD4b- β-caroteno. Por un lado, el RMSD fue graficado y se obtuvo un promedio de 0,286 y 0,3 Å para las dos trayectorias duplicadas del complejo CCD4a-zeaxantina. En el caso del complejo CCD4b-β-caroteno el RMSD promedio fue de 0,380 Å. Esto demuestra una situación de equilibrio estructural en la simulación. El resultado se muestra en la Fig. 8.





Por otro lado, se realizó un estudio de la evolución del patrón de enlaces de hidrógeno a lo largo del tiempo en la esfera de 4,5 Å centrada en cada carotenoide estudiado y se midieron la



cantidad de enlaces de hidógeno que forman todos los átomos dadores y aceptores presentes en dicha esfera (Fig. 9).

Figura 9. Evolución del patrón de enlaces de hidrógeno en trayectorias de dinámica molecular observadas en una esfera de 4,5 Å centrada en cada ligando. Superior: Se muestran dos réplicas (r1 y r2) en trayectorias de 10 ns de CCD4a y zeaxantina donde el número de enlaces de hidrógeno promedio observado fue de 177 Inferior: Se muestra una trayectoria de 20 ns de CCD4b1 y β-caroteno, donde el número de enlaces de hidrógeno promedio observado fue de 154.

Para el complejo CCD4a-zeaxantina en 10 ns de simulación se observaron promedios de 176 y 178 enlaces de hidrógeno, mientras que para para el complejo CCD4a-zeaxantina en 10 ns de simulación mientras que se obtuvo un total de 154 para el complejo CCD4b-β-caroteno en 20 ns de simulación. En primer lugar, es de destacar que el β-caroteno no participa de ninguno de esos enlaces de hidrógeno, tal como se pone en evidencia gráficamente en la Fig. 7. Sin embargo, puede jugar un rol estructurante del solvente (y sus interacciones de enlace de hidrógeno) de manera diferente a cualquier otro carotenoide hidroxilado. Por otro lado, tal como es esperable, en el caso de las xantofilas se observa un patrón de enlaces de hidrógeno mayor (176 y 178) y que las mismas participan en el mismo a través de sus grupos hidroxilados. Eso también se observa en la Fig. 7. Esto agrega otra evidencia en que las interacciones de las xantofilas son más intensas que en el caso de los carotenos. En conclusión, el promedio de enlaces de hidrógeno mayor para el caso de las xantofilas se corresponde con la presencia de un *cluster* de moléculas de agua configurado en el entorno de las hidroxilaciones de las diferentes xantofilas (Fig. 7).

ESFERA DE COORDINACIÓN DEL Fe $^{+2}$ Y LA INFLUENCIA EN LA REACTIVIDAD DE LAS ENZIMAS CCD4

En el cristal de la ACO y en los correspondientes modelos de las CCD4s, una molécula de agua toma el lugar del oxígeno en la catálisis real, en la cual el Fe⁺² activa al oxígeno para que pueda llevarse a cabo el ataque sobre alguno de los carbonos de los dobles enlaces. Por eso es esencial en los modelos, ubicar a tal molécula de agua y medir su distancia al Fe⁺². Tal como se anota en la Fig. 10, construida mediante la sobreposición de todas las moléculas de carotenoides ancladas, la distancia entre el Fe⁺² y una de las moléculas de agua es de 2,54 Å, y entre ella y el doble enlace es de 2,63 Å. Podemos concluir que existe un espacio adecuado para que una molécula de oxígeno, posicionada en el lugar de esta molécula de agua, participe de la catálisis en donde alguno de los dobles enlaces de los carotenoides será oxidado.



Figura 10. Representación en varas y esferas superpuestas de carotenoides acoplados en CCD4a. Magenta: zeaxantina; Verde: β-caroteno; Naranja: β-criptoxantina. En el gráfico se muestra la distancia entre la molécula de agua (que ocupa la posición del oxígeno) y el Fe⁺² y la distancia de la molécula de agua al enlace 7=8.

CONCLUSIONES

En el presente trabajo presentamos modelos moleculares validados para la CCD4a y CCD4b de *Citrus clementina* con una serie de seis diferentes ligandos carotenoides. Estimaciones de las energías de interacción se presentaron por *scores* de *docking* y cálculos de dinámica molecular. Ambos modelos moleculares, en su primera versión (la acá presentada) fueron obtenidos por modelado por homología partiendo de la estructura cristalográfica de la ACO (PDB ID: <u>2BIW</u>) y resultaron energéticamente estables luego de 5 ns de simulación de dinámica molecular. Los modelos construidos confirman la estabilidad de la típica estructura terciaria de los homólogos de la CCD4. En ella se pueden describir una parte flexible conformada por *loops* largos y α -hélices cortas y una parte rígida conformada por láminas- β que contienen al Fe^{+2.} El átomo metálico está coordinado en una primera esfera por cuatro residuos de histidina. Una segunda esfera de coordinación, a una distancia mayor que tales histidinas, se puede describir envolviendo

y manteniendo la estabilidad general de la primera. En ella se encontró el residuo Asp268. Durante la simulación de dinámica molecular, los residuos y moléculas de agua alternaban posiciones cercanas al Fe^{+2} lo que sugiere un entorno flexible, con la variabilidad y adaptabilidad suficiente como para que la estructura y dinámica de la esfera de coordinación acompañe adecuadamente el evento catalítico.

Los resultados de *docking* de los carotenoides en el sitio activo de la CCD4a sugieren que la unión es más favorable a la zeaxantina que la β -criptoxantina y menos favorable para el β -caroteno. Los cálculos de dinámica molecular de la CCD4a muestran que la energía de interacción promedio sobre las trayectorias están en el mismo orden de unión zeaxantina > β -criptoxantina > β -criptoxantina > β -criptoxantina > β -caroteno.

En el caso de la CCD4b, se encontró de acuerdo con el resultado experimental para todos los carotenoides con anillos β , el orden de la preferencia de unión fue: Luteína > zeaxantina > β -criptoxantina > α -caroteno > licopeno > β -caroteno. Estos resultados muestran mejores comportamientos de las xantófilas comparadas con los carotenos, siendo la polarizabilidad y la presencia de anillos β hidroxilados importantes para la unión con la enzima.

La isomería geométrica de los carotenoides a quienes se les asignaron una configuración con todos sus enlaces trans (*all trans*) se conservó durante las simulaciones de dinámica molecular en la mayoría de los ligandos. De cualquier manera, el licopeno (caroteno alifático) sufrió una isomerización geométrica de *trans* a *cis* de uno de sus dobles enlaces conjugados, resultado que amerita una investigación más profunda. Una isomerización *in silico* es un evento inesperado durante una dinámica molecular de equilibrio dada la alta energía relativa requerida para una rotación de un doble enlace y el espacio restringido que ocupa el ligando en el sitio activo. De cualquier manera, isomerizaciones *in vivo* pueden ocurrir, a través de la actividad catalítica de estas enzimas, cuando ellas promueven cambios en los enlaces de los ligandos (Kloer y Schulz, 2006). Adicionalmente en otros trabajos, la geometría y el espacio conformacional se estudió para licopeno y zeaxantina (Meléndez-Martínez et al., 2014).

Las distancias entre el centro catalítico Fe^{+2} y los ligandos son adecuados para clivar en la posición del doble enlace 7'=8'. En todos los casos se puede descartar la posible actividad de corte simétrico 15=15', que concuerda con los resultados experimentales (Rodrigo et al., 2013).

Alternativamente, no se puede descartar un posible punto de corte en el doble enlace 9'=10' debido a que las distancias son cercanas a las observadas en 7'=8'.

Una vez realizado un estudio detallado del sitio de unión y el comportamiento de dos miembros de la familia CCD4 provenientes de *Citrus clementina* a través de los modelos propuestos en este capítulo, nos dispusimos a analizar la estructura de estas proteínas en un contexto más realísta. El mismo incluye la interacción con membranas por lo que fue necesario reconstruir el modelo con sus respectivos dominios de interacción con membranas, el cual será presentado en el capítulo siguiente, en una investigación que brindará una visión *in silico* más cercana a los ensayos *in vivo*.

REFERENCIAS

- Bond, S. D., Leimkuhler, B. J., & Laird, B. B. (1999). The Nosé–Poincaré Method for Constant Temperature Molecular Dynamics. Journal of Computational Physics, 151(1), 114–134. https://doi.org/https://doi.org/10.1006/jcph.1998.6171
- Bruno, M., Beyer, P., & Al-Babili, S. (2015). The potato carotenoid cleavage dioxygenase 4 catalyzes a single cleavage of beta-ionone ring-containing carotenes and non-epoxidated xanthophylls. Archives of Biochemistry and Biophysics, 572, 126–133. https://doi.org/10.1016/j.abb.2015.02.011
- Chemical Computing Group Inc. (2014). Molecular Operating Environment (MOE) (2014.09).
- FAO. (2017). Citrus fruit. Fresh and processed- Statistical Bulletin 2016. Trade and Markets Division, 47. <u>http://www.fao.org/3/a-i8092e.pdf</u>
- Gonzalez-Jorge, S., Ha, S.-H., Magallanes-Lundback, M., Gilliland, L. U., Zhou, A., Lipka, A. E., Nguyen, Y.-N., Angelovici, R., Lin, H., Cepela, J., Little, H., Buell, C. R., Gore, M. A., & DellaPenna, D. (2013). CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds. The Plant Cell, 25(12), 4812–4826. <u>https://doi.org/10.1105/tpc.113.119677</u>
- Halgren, T. A. (1999). MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. Journal of Computational Chemistry, 20(7), 730–748.

https://doi.org/https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2-T

- Huang, F.-C., Molnár, P., & Schwab, W. (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. Journal of Experimental Botany, 60(11), 3011– 3022. <u>https://doi.org/10.1093/jxb/erp137</u>
- Kato, M. (2012). Mechanism of carotenoid accumulation in citrus fruit. Journal of the Japanese Society for Horticultural Science, 81(3), 219–233. <u>https://doi.org/10.2503/jjshs1.81.219</u>
- Kloer, D. P., & Schulz, G. E. (2006). Structural and biological aspects of carotenoid cleavage. Cellular and Molecular Life Sciences, 63(19–20), 2291–2303. https://doi.org/10.1007/s00018-006-6176-6
- Kloer, Daniel P., Ruch, S., Al-Babili, S., Beyer, P., & Schulz, G. E. (2005). The structure of a retinal-forming carotenoid oxygenase. Science, 308(5719), 267–269. https://doi.org/10.1126/science.1108965
- Lätari, K., Wüst, F., Hübner, M., Schaub, P., Beisel, K. G., Matsubara, S., Beyer, P., & Welsch, R. (2015). Tissue-Specific Apocarotenoid Glycosylation Contributes to Carotenoid Homeostasis in Arabidopsis Leaves. Plant Physiology, 168(4), 1550 LP 1562. https://doi.org/10.1104/pp.15.00243
- Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., Wahyudi, A., Motohashi, R., & Kato, M. (2013). Enzymatic formation of beta-citraurin from betacryptoxanthin and Zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. Plant Physiology, 163(2), 682–695. <u>https://doi.org/10.1104/pp.113.223297</u>
- Meléndez-Martínez, A. J., Paulino, M., Stinco, C. M., Mapelli-Brahm, P., & Wang, X.-D. (2014). Study of the Time-Course of cis/trans (Z/E) Isomerization of Lycopene, Phytoene, and Phytofluene from Tomato. Journal of Agricultural and Food Chemistry, 62(51), 12399– 12406. <u>https://doi.org/10.1021/jf5041965</u>
- Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., & Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiology, 142(3), 1193–1201. <u>https://doi.org/10.1104/pp.106.087130</u>
- Rodrigo, M. J., Alquézar, B., Alís, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461– 4478. <u>https://doi.org/10.1093/jxb/ert260</u>

- Rodrigo, M. J., Alquézar, B., Alós, E., Lado, J., & Zacarías, L. (2013). Biochemical bases and molecular regulation of pigmentation in the peel of Citrus fruit. Scientia Horticulturae, 163, 46–62. https://doi.org/https://doi.org/10.1016/j.scienta.2013.08.014
- Vega-Teijido, M., Cantero, J., Rodrigo, M. J., López, C., & Paulino Zunini, M. (2018). An In Silico Study of the Citrus Dioxygenases CCD4 family Substrates. Journal of Biomolecular Structure and Dynamics, 0(0), 1–28. <u>https://doi.org/10.1080/07391102.2018.1477619</u>
- Wang, J., Cieplak, P., & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry, 21(12), 1049–1074. <u>https://doi.org/https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F</u>
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all-atom force field for simulations of proteins and nucleic acids. Journal of Computational Chemistry, 7(2), 230– 252. <u>https://doi.org/https://doi.org/10.1002/jcc.540070216</u>

CAPÍTULO IV

CCD4 DE Citrus clementina Y SU INTERACCIÓN CON MEMBRANAS BIOLÓGICAS

INTRODUCCIÓN

El cloroplasto es un organelo celular característico de las especies vegetales y en el interior de el se desarrolla la fotosíntesis. El cloroplasto está conformado por un par de membranas concéntricas que forman la envoltura. La membrana de la envoltura interna es altamente selectiva y contiene numerosos transportadores que regulan los flujos de metabolitos e iones entre el cloroplasto y el citosol. La membrana interna también interviene en algunas enzimas de las rutas biosintéticas y entre ellas las de carotenoides.(Rascio, 2013) El tilacoide es un sistema de membranas al interior del cloroplasto, que contiene los complejos multiprotéicos por lo general involucrados en las reacciones de la fase luminosa de la fotosíntesis. El sistema de membranas del cloroplasto consta de lamelas acopladas (membranas tilacoides) separadas por un estrecho espacio intermembrana, que se intercomunican para constituir un sistema de membrana continuo y cerrado, cuyo espacio interno (lumen) está totalmente aislado del entorno (estroma). (Rascio, 2013)

La composición de la membrana tilacoidal está integrada principalmente por glicolípidos, sulfolípidos, fosfolípidos, las clorofilas y los carotenoides, siendo los principales carotenoides presentes la luteína, la violaxantina, la neoxantina y el β -caroteno. (Jeffrey et al., 1974; Britton, 1990)

La localización de las enzimas que intervienen en la ruta de los carotenoides, específicamente con su asociación a las membranas es determinante para su actividad enzimática. La existencia de estas enzimas explica la acumulación y transporte de varios carotenoides, productos intermedios y finales en las membranas tilacoidales, muchas de estas enzimas están relacionadas a los fotosistemas en las plantas. (Rodríguez-Villalón et al., 2009).

Como se ha señalado en el Capítulo 3, la estructura en forma de β -propeller se conserva en toda la familia de enzimas CCD y también el parche hidrofóbico en el domo de la proteína que cubre al canal de entrada del sustrato (Kloer y Schulz, 2006). En las plantas esta región es aún mayor y está formada por dos α -hélices antiparalelas que se han propuesto como involucradas en la penetración de la membrana (Ahrazem et al., 2016).

Un estudio previo sugirió que los residuos de aminoácidos de las α -hélices de la CCD4 de citrus cuentan con la capacidad de interactuar con la membrana, sin embargo, la herramienta utilizada en ese estudio estaba desarrollada para identificar específicamente proteínas de transmembrana. Esta identificación se realizó únicamente en base a la predicción de la constitución de la secuencia sin datos estructurales adicionales como la profundidad de penetración y orientación de los residuos que intervendrían en el anclaje a la membrana.(Ma et al., 2013a)

Estudios mutacionales muestran que con la remoción del dominio α -hélice la enzima pierde su capacidad de interactuar con la membrana tilacoidal llevando a una perdida de su actividad enzimática. (Tan et al., 2001) Otro estudio mostró que en la VP14, una CCD proveniente de *Zea mays*, gran parte del dominio comprendido por las α -hélices podrían penetrar la superficie de la membrana tilacoidal y extraer el sustrato de la membrana a través de un túnel formado principalmente por residuos hidrofóbicos y que se extiende hacia el centro catalítico de la enzima (Rubio-Moraga et al., 2014)

La primera estructura cristalográfica cristalizada de una CCD proveniente de planta corresponde a *Zea mays* y sus autores sugieren como la VP14 se asocia a la membrana de manera tal que su orientación y profundidad de penetración permitiría acceder a su sustrato soluble en membrana. (Messing et al., 2010). Es importante destacar que 9-cis-neoxantina y 9-cis-violaxantina, los sustratos primarios de VP14, son 10 veces más abundantes en la membrana tilacoide que en otras membranas vegetales (Parry y Horgan, 1991). Además, estos autores sugieren que es un mecanismo heredado consistente entre todos las CCDs de origen vegetal.

En las CCDs provenientes de *Arabidopsis thaliana* se han observado diferencias en la unión a la membrana entre los distintas miembros de esta familia, demostrando que diferentes CCDs de una misma especie pueden diferir en su afinidad a la membrana. (Tan et al., 2003)

Dada la importancia biotecnológica de las dioxigenasas CCD4, ampliamente detallada en el Capítulo 1, y la ausencia de evidencia estructural suficientes sobre el anclaje y la interacción con las membranas, el presente capítulo se centró en las CCD4 de especies de *Citrus clementin*a que se expresan (CCD4a, b y c) combinando diferentes técnicas bioinformáticas para modelar y validar sus estructuras tridimensionales, la construcción de complejos con ligandos carotenoides putativos para dilucidar su selectividad de sustrato y posibles sitios de corte junto con un análisis detallado de los residuos que participan en las interacciones receptor-ligando, así como su interacción con membranas biológicas mediante técnicas de simulación molecular.

El presente capítulo fue remitido a la revista "International Journal of Molecular Sciences (IJMS)" bajo el título "Citrus carotenoids cleavage dioxygenases type CCD4 family and their interaction with membranes". Autores: Jorge Cantero, Fabio Polticelli y Margot Paulino. El manuscrito se encuentra actualmente en revisión. El documento completo se adjunta en el Anexo A2.

OBJETIVOS

- Obtener modelos de la estructura tridimensional de proteínas de la familia CCD4 provenientes de especies de *Citrus clementina* en interacción con un modelo de membrana biológica.
- Caracterizar los aminoácidos que intervienen en la interacción con la membrana.
- Evaluar la influencia de la membrana en la estabilidad estructural del dominio αhélice de las CCD4 mediante simulación de dinámica molecular.
- Identificar la selectividad y posibles sitios de corte sugeridos por el análisis de las trayectorias de dinámica molecular.

MATERIALES Y MÉTODOS

SOFTWARE, HARDWARE, CAMPO DE FUERZA Y PARAMETRIZACIÓN

Los modelos y simulación se realizaron en una estación de trabajo equipada con un procesador Intel® Core[™] i7-6950X Extreme Edition corriendo bajo el sistema operativo Linux (Ubuntu 18.04), una tarjeta gráfica NVIDIA GTX 1080 para aceleración de cálculos utilizando la tecnología CUDA Toolkit 9.0. Todos los cálculos fueron llevados a cabo mediante el software NAMD2 versión 2.12 optimizado para trabajar en núcleos CUDA (Phillips et al., 2020).

El campo de fuerza seleccionado para la simulación fue el Charmm36 (Best et al., 2012) compatible de forma nativa con el programa NAMD2. La parametrización de los ligandos se realizó a través del programa CGenFF (*CHARMM General Force Field*) (Vanommeslaeghe et al., 2010), compatible con el campo de fuerzas Charmm36. En este programa la asignación de parámetros y cargas se realiza por analogía, estudiando los patrones de conectividad y el entorno químico de cada átomo. Para el átomo de Fe⁺² en coordinación se utilizó el esquema no enlazante. En consecuencia, la interacción Fe⁺²-proteína se modeló mediante términos electrostáticos y

VDW, permitiendo de esta manera el estudio de posibles cambios en el centro de coordinación. Las interacciones de VDW se modelaron mediante un potencial 12–6 Lennard-Jones (LJ) siguiendo el modelo del campo de fuerza Charmm36. Los parámetros R_{min} y ε se obtuvieron de trabajos previos publicados en parametrización de cationes metálicos con carga +2 (Li et al., 2013) y se optimizaron para este sistema.

CONSTRUCCIÓN DE MODELOS 3D DE LAS PROTEÍNAS

Los modelos de las proteínas CCD4a, CCD4b y CCD4c, incluyendo a los dominios $\alpha 1$ y $\alpha 3$ se construyeron utilizando el software PHYRE2 (*Protein Homology / analogY Recognition Engine*) (Kelley et al., 2016). Este software combina el método de modelado por *threading* con técnicas *ab initio* con modelado de partes de la secuencia por homología de múltiples *templates*.

Se tomaron algunos elementos estructurales del modelo presentado en el capítulo anterior, como punto de partida (Vega-Teijido et al., 2018). En ese sentido, luego de una inspección de la esfera de coordinación del Fe^{+2} , la posición del Fe^{+2} se definió por superposición de ambos modelos, verificando la esfera de coordinación del Fe^{+2} con las histidinas (His257, His307, His373 y His548).

Por otro lado, en estos nuevos modelos, la inclusión de los dominios $\alpha 1$ y $\alpha 3$ (presentes únicamente en proteínas provenientes de plantas) es especialmente importante dado que son quienes interactúan con las membranas tilacoidales de los cloroplastos.

Los modelos obtenidos de la CCD4a, CCD4b y CCD4c fueron comparados con la estructura cristalográfica de la enzima vegetal dioxigenasa de clivaje de 9-cis-epoxicarotenoides (VP14), una estructura proveniente del maíz (*Zea mays*) expresada en *Escherichia coli* (PDB ID: <u>3NPE</u>) siendo esta la única estructura proveniente de plantas con estructura cristalográfica conocida que posee los dominios α 1 y α 3 que interactúan con las membranas tilacoidales. VP14 es una enzima capaz de tomar como sustrato a los carotenoides con la configuración 9-cis y clivar en la posición 11 = 12 (11 '= 12') de los dobles enlaces, formando parte de la primera etapa de la

ruta de biosíntesis del ABA (EC 1.13.11.51). La resolución del cristal es de 3.2 Å, con un R-factor = 0.242 y un R-free = 0.272.

REFINAMIENTO Y DINÁMICA MOLECULAR

La validación de los modelos se realizó mediante simulaciones de dinámica molecular, para lo cual se construyó una caja de agua con solvente explícito definiendo 15 Å de margen entre el átomo más externo y el borde de la caja en cada eje cartesiano, resultando celdas periódicas de dimensiones de 88 x 100 x 86 Å. Se utilizaron contraiones de NaCl a fin de asegurar la neutralidad global de cargas del sistema. Las cajas de agua se construyeron utilizando el modelo de agua TIP3P, en el cual las moléculas de agua se consideran rígidas, permitiendo su rotación y traslación a lo largo de la caja.

Para todas las simulaciones se consideraron sistemas en condiciones periódicas, con un time step de 2 fs. Las etapas de muestreo se realizaron cada 2 ps. La electrostática del sistema se trató mediante el método Particle Mesh Ewald (PME con un tamaño de grilla de 1 Å y una distancia de *cutoff* para las interacciones de largo alcance de 12 Å.

Cada simulación consistió en cuatro etapas:

- Refinamiento mediante minimización de energía con 2000 steps utilizando gradientes conjugados y aplicando restricciones armónicas sobre la cadena principal de la proteína y sobre la espera de coordinación del Fe⁺² para evitar posibles deformaciones en etapas iniciales.
- 2 Etapa de calentamiento de 0 a 300 K en 100 ps de simulación, manteniendo la presión constante a 1 atm y las restricciones armónicas sobre la cadena principal y la esfera de coordinación del Fe⁺².
- 3 Etapa de simulación de dinámica molecular a temperatura constante de 300 K durante
 1 ns, manteniendo las restricciones armónicas de las etapas anteriores.
- 4 Etapa de simulación de dinámica molecular a temperatura constante (300 K) durante
 20 ns, sin restricciones armónicas.

La ecuación de movimiento se resolvió utilizando la ecuación de dinámica de Langevin (Ecuación 1) para simular un ensamble isotérmico-isobárico (NPT):

Ecuación 1.
$$m_i \frac{d^2 x_i(t)}{dt^2} = F_i x_i(t) - \gamma_i \frac{dx_i(t)}{dt} m_i + R_i(t)$$

donde,

 m_i es la masa del átomo i $x_i(t)$ es la posición del átomo i al tiempo t $F_i x_i(t)$ es la fuerza que actúa sobre el átomo i al tiempo t γ_i es el coeficiente de fricción que actúa sobre el átomo i (5 ps⁻¹)

 $R_i(t)$ son las fuerzas aleatorias que actúan sobre el átomo i al tiempo t

Luego de finalizar las etapas de simulación de los modelos de la CCD4a, CCD4b y CCD4c, se analizó su comportamiento energético y estructural en el equilibrio y como resultado, se consideraron adecuadas para utilizarlas en el estudio de los complejos CCD4-ligando y sus interacciones con modelos de membranas biológicas.

CONSTRUCCIÓN DE LOS COMPLEJOS CCD4-LIGANDOS

Dada la similitud estructural entre la CCD4a, la CCD4b, y CCD4c obtenidas mediante el método de *threading* con los obtenidos por homología, la posición inicial de los ligandos en el sitio activo se estableció alineando estructuralmente ambos modelos y tomando la posición del carotenoide en los modelos obtenidos y validados en el Capítulo III (Vega-Teijido et al., 2018). Posteriormente, todos los complejos fueron sometidos a simulaciones de dinámica molecular.

DINÁMICA MOLECULAR DE LOS COMPLEJOS

La simulación de dinámica molecular de los complejos se realizó siguiendo el protocolo de la sección de **REFINAMIENTO Y DINÁMICA MOLECULAR**. La simulación se dividió en cuatro etapas: minimización de energía, calentamiento de 0 a 300 K en 100 ps, dinámica molecular de equilibrio a 300 K durante 1 ns restringiendo la posición de los átomos de la cadena principal (cadena principal). Luego, se obtuvo una trayectoria de muestreo (sin ningún tipo de restricción) de 20 ns. Todas las simulaciones se realizaron por triplicado para lograr un muestreo estadístico de los parámetros fisicoquímicos a ser medidos.

CONSTRUCCIÓN DE LOS COMPLEJOS EN MEMBRANA

Para proponer la disposición espacial de las estructuras de proteínas en las bicapas lipídicas utilizamos el programa OPM (*Orientations of Proteins in Membranes*) (Lomize et al., 2012) en el cual cada proteína se considera como cuerpo rígido que puede interactuar libremente con una capa hidrofóbica de ancho ajustable. La orientación de la proteína se optimiza mediante una simulación de dinámica molecular de cuerpo rígido, considerando siempre la proteína como el soluto cuya posición relativa respecto a la membrana se debe optimizar y a la membrana mediante una representación implícita (no estructural). Como primera etapa, la energía libre ($\Delta G_{transfer}$) de transferencia de la proteína a la membrana desde un medio de solución acuosa se minimiza mediante un cálculo a diferentes niveles de penetración, moviendo a la proteína a una velocidad de 0,2 Å /step. Finalmente se reporta la profundidad de penetración óptima, el ángulo de rotación del centro de masas respecto al plano conformado por la bicapa fosfolipídica, los residuos en contacto con la membrana y los residuos fuera de ella y el valor óptimo del $\Delta G_{transfer}$

Para la construcción de la membrana utilizamos un modelo pre-equilibrado de fosfatidilcolina (POPC), formado por unidades de diacilglicerol y fosfolípido, las cuales inicialmente abarcaron una superficie de 102 x 140 Å. El ancho inicial de las membranas, se midió

como la distancia promedio entre las capas de los grupos fosfato, y tuvo un valor promedio de 39 Å para los tres modelos (CCD4a, b y c).

Las estructuras de las proteínas se acoplaron a las capas de membrana, respetando la profundidad y orientación obtenidas en la etapa anterior y removiendo las unidades de POPC que se superponían con las mismas.

El modelo de los complejos proteína-membrana se sometió a simulaciones de dinámica molecular siguiendo el protocolo de la sección de REFINAMIENTO Y DINÁMICA MOLECULAR, manteniendo la superficie de la membrana constante (formada por el plano X-Y). La simulación se dividió en cuatro etapas: minimización de energía, calentamiento de 0 a 300 K en 100 ps, dinámica a 300 K durante 1 ns restringiendo la posición de los átomos de la cadena principal (*backbone*) y finalmente la etapa de producción para todos los complejos modelados con trayectorias triplicadas de 20 ns.

Como última etapa, al complejo CCD4b- β -criptoxantina para modelar un paisaje biológico más realísta, considerando que la β -criptoxantina es uno de los carotenoides más activos reportado en la literatura y para el cual se tiene evidencia experimental (Rodrigo et al., 2013), se amplió el muestreo de observación, extendiendo la trayectoria a 1 μ s.

ENERGÍA DE INTERACCIÓN DEL LIGANDO (Uab)

La energía de interacción del ligando con todo el sistema molecular modelado (proteína, solventes e iones) se evaluó a lo largo de la dinámica molecular, de acuerdo a la Ecuación 2:

Ecuación 2.
$$\langle Uab \rangle = \langle U - Ua - Ub \rangle$$

donde,

U es la energía potencial total del sistema

Ua es la suma de la energía potencial del receptor, membrana, iones, solventes y sus

interacciones (el ligando es excluido del sistema)

Ub es la energía potencial del ligando

La *Uab* está compuesta por las componentes no enlazantes del campo de fuerza, que se separan en dos términos: las contribuciones de Van der Waals (VDW) y la componente electrostática (ELECT).

INTERACCIONES PROTEINA-MEMBRANA

Para el estudio de la influencia de la membrana fosfolipídica en los dominios de las CCD4 a lo largo de las simulaciones de dinámica molecular, se identificarán los residuos que participan directamente de la interacción, la evolución temporal de la superficie de contacto entre la proteína y la membrana y la caracterización de las interacciones no enlazantes presentes (tales como el patrón de enlaces de hidrógeno, las componentes electrostáticas y de VDW que caracterizan a la interacción no enlazante entre ambas).

Además, se estudiarán los cambios de la estructura de la membrana relacionados a la inserción de estas proteínas.

INFLUENCIA DE LA MEMBRANA EN LA MOVILIDAD DE RESIDUOS

El cálculo de la movilidad de los residuos que conforman a una proteína es una medida muy importante de su estabilidad. Una de las maneras de representarla es la fluctuación cuadrática media o RMSF (*Root Mean Square Fluctuation*), que es una medida de la desviación de las posiciones que ocupa el carbono alfa de cada residuo respecto a una ubicación espacial de referencia como puede ser el promedio de las coordenadas del ensemble (Ecuación 3):

Ecuación 3. RMSF_i =
$$\sqrt{\frac{1}{T}\sum_{t_j=1}^{T} (\Gamma_i(t_j) - \overline{\Gamma}_i)^2}$$

donde,

T es tiempo total de la simulación

 $\Gamma_i(t_i)$ es la posición del centro de masas de la cadena principal del residuo Γ_i al tiempo t_i

 $\overline{\Gamma_i}$ es el promedio de las posiciones del centro de masas de la cadena principal del residuo Γ_i , como posición de referencia.

Mientras mayor es el valor de RMSF de un residuo, este será más móvil.

Para estimar las diferencias de movilidad y estabilidad de los distintos dominios de las CCD4 en solvente respecto al modelo con interacción con las membranas, se diseñó *ad hoc* una manera de calcular las diferencias entre las fluctuaciones cuadráticas medias (Δ RMSF), siguiendo la siguiente ecuación (Ecuación 4):

Ecuación 4.
$$\Delta RMSF_i = \langle RMSF_i \rangle_{memb} - \langle RMSF_i \rangle_{sol}$$

donde,

 $(RMSF_i)_{memb}$ es el valor de la fluctuación cuadrática media del residuo *i* en el modelo de interacción con las membranas.

 $\langle RMSF_i \rangle_{sol}$ es el valor de la fluctuación cuadrática media del residuo i en el modelo en solvente.

Si el valor de $\Delta RMSF_i$ es positivo, el residuo i presenta mayor movilidad en el modelo de interacción con membranas cuando se compara con el modelo en solvente. Si el valor de $\Delta RMSF_i$ es negativo, el residuo i presenta menor movilidad en el modelo de interacción con membranas cuando se compara con el modelo en solvente. Si el valor de $\Delta RMSF_i$ es cero, no se observan diferencias en las fluctuaciones del residuo i entre el modelo de interacción con membranas y el modelo en solvente.

Para facilitar la visualización, se representarán los valores del $\Delta RMSF_i$ mediante la coloración similar del diagrama de cintas de la que se utiliza para visualizar las diferencias en el factor B cristalográfico. En ese caso, hemos adaptado esa visualización de manera que cuando se

visualice la cinta en un color azul, esto indicará una mayor movilidad en fase acuosa, mientras que los colores rojizos en las cintas estarán indicando regiones de mayor movilidad em membrana.

PENETRACIÓN EN LA MEMBRANA

La penetración de la proteína dentro de la membrana se estimó mediante la medida de tres parámetros a lo largo de la simulación, para lo cual desarrollamos "*ad hoc*" un conjunto de algoritmos. Estos parámetros describen la evolución del espesor de la membrana en el tiempo, la profundidad de penetración de la proteína dentro de la membrana y la penetración relativa de la proteína en función al grosor de la membrana.

Para medir el grosor de la membrana, los grupos fosfato alineados en el plano XY y proyectados en el eje Z se agruparon en dos conjuntos: un primer conjunto MP2 se definió como todos los grupos fosfato en el lado de la membrana en el que se encontraba insertada la proteína. Los MP1 son los fosfatos que pertenecen a la otra capa de la membrana.

El método con el que se midió el *grosor* medio de la membrana se presenta en la Ecuación 5:

Ecuación 5.
$$Grosor_t = \frac{\sum_i \left(Z_{MP1(i)}\right)_t}{n_{(i)}} - \frac{\sum_j \left(Z_{MP2(j)}\right)_t}{n_{(j)}}$$

donde,

 $Z_{MP1(i)}$ es el conjunto de valores de las coordenadas cartesianas en el eje Z de los *i* fosfatos que pertenecen al conjunto MP1 al tiempo *t*,

 $Z_{MP2(j)}$ es el conjunto de valores de las coordenadas cartesianas en el eje Z de los *i* fosfatos que pertenecen al conjunto MP2 al tiempo *t*,

 $n_{(i)}$ es el número de fosfatos que pertenecen al conjunto MP1,

 $n_{(i)}$) es el número de fosfatos que pertenecen al conjunto MP2.

El segundo parámetro medido es la *penetración* de la proteína dentro de la membrana para cuyo cálculo se desarrolló el algoritmo presentado en la Ecuación 6:

Ecuación 6. Penetración_t = Max
$$(Z_{prot})_t - \frac{\sum_j (Z_{MP2(j)})_t}{n(j)}$$

donde,

 $Max(Z_{prot})_t$ es el máximo valor observado de la coordenada cartesiana Z de la proteína (no se consideran átomos de hidrógeno).

Dado que el grosor de la membrana es variable a lo largo de la simulación, el tercer parámetro desarrollado fue la *penetración relativa* de la proteína en la membrana (RPD) en función al ancho de la misma, el que se evaluó a lo largo de la simulación mediante el algoritmo presentado en la Ecuación 7:

Ecuación 7. $RPD_t = \frac{Penetración_t}{Grosor_t} * 100$

ESTIMACIÓN DE LA ENERGÍA LIBRE DE UNIÓN (MMPBSA)

MMPBSA (*Molecular Mecanics Poisson-Boltzmann Surface Area*) es un método útil para la estimación de la energía libre de unión que combina mecánica molecular y modelos de solvente implícito continuo. Los procesos de unión receptor ligando se descomponen en dos etapas: asociación en fase gaseosa y solvatación en fase acuosa. Mientras que la energía en fase gaseosa se calcula con campos de fuerza de mecánica molecular, los términos de solvatación se cuantifican usando modelos de solvente implícito. El muestreo de conformaciones se tomó de las trayectorias generadas durante las dinámicas moleculares de equilibrio. Se calcularon tres contribuciones energéticas siguiendo el método MMPBSA. Primero, en fase gaseosa se calculó la diferencia energética entre el complejo y la suma de las energías del receptor y el ligando por separado, utilizando NAMD y el campo de fuerzas de la MD (MM). Luego, la contribución polar a la energía libre de solvatación se calcula numéricamente mediante la ecuación de Poisson-Boltzmann (PB) implementado en el programa APBS (*Adaptive Poisson-Boltzmann Solver*) (Jurrus et al., 2018). Posteriormente, se miden los cambios (diferencias) de la superficie del área accesible al solvente (SASA) y a partir de ella se estima la contribución no polar a la energía libre de solvatación, mediante la relación lineal con SASA. Por último, la energía libre de unión $\Delta G_{unión}$ se estima mediante la Ecuación 8:

Ecuación 8.
$$\Delta G_{binding} = \Delta H - T\Delta S = \langle \Delta M M_{gas} + \Delta G_{polar} + \Delta G_{nopolar} - T\Delta S \rangle$$

donde,

H es la entalpía *T* es la temperatura absoluta *S* es la entropía del sistema ΔMM_{gas} es la diferencia energética entre el complejo y la suma de las energías del receptor y el ligando por separado, calculada por mecánica molecular en fase gaseosa ΔG_{polar} es la contribución polar a la energía libre de solvatación $\Delta G_{nopolar}$ es la contribución no polar a la energía libre de solvatación $T\Delta S$ es la contribución entrópica

El grupo de compuestos estudiados es estructuralmente similar, por pertenecer todos a la clase estructural "carotenoides". Se puede asumir que su contribución a las diferencias en la entropía es similar y que por lo tanto los cambios de la energía libre debido al factor entrópico se anulan entre sí, cuando se estudian las diferencias entre las energías libres (expresadas como $\Delta\Delta G$). En consecuencia, el término T Δ S puede ser ignorado tal como es usual en los métodos MMBPSA.

RESULTADOS Y DISCUSIÓN

Los modelos tridimensionales alineados resultados del *threading* de la CCD4a, CCD4b y CCD4c se muestran en la Fig. 1, superpuestos a la estructura de la VP14, una CCD proveniente del maíz que posee los dominios de α -hélices que caracterizan a las CCDs de origen vegetal. Tal como se explicó antes, estas α -hélices tendrían como función interactuar con las membranas tilacoides. En consecuencia, incluirlas en el modelo nos permitirá estudiar el efecto de la interacción de nuestras proteínas con membranas biológicas.



Figura 1. Esquema de cintas de los modelos CCD4 alineados: CCD4a (verde), CCD4b (azul), CCD4c (rosa) y VP14 (rojo). El dominio α1 / α3 está señalado con flechas rojas.

Las estructuras obtenidas por PHYRE 2 describen una estructura característica del tipo β propeller de los CCDs. Este dominio está formado por siete hojas- β que se disponen alrededor de un eje central y dentro del cual se encuentra el centro catalítico compuesto por una esfera de coordinación de Fe⁺² a través de 4 histidinas altamente conservadas en los miembros de esta familia. El alineamiento estructural de los modelos con la estructura cristalográfica de la VP14 muestra una diferencia estructural en un rango de RMSD 0,85-1,38 Å, mostrando una gran similitud entre ellos (Tabla 1). Esto nos permite sugerir que los modelos tridimensionales obtenidos fueron muy cercanos a la ya conocida en otra especie vegetal, similitud esperada ya que comparte una gran similitud de secuencia con la VP14.

Tabla 1.Valores de RMSD (Diagonal inferior, en blanco) y porcentaje de similitud (diagonal
superior, naranja claro) calculados para los modelos de CCD4a, CCD4b y CCD4c,
comparados con la estructura cristalográfica de la VP14.

RMSD/Similitud	CCD4a	CCD4b	CCD4c	VP14
CCD4a		62,70	72,00	51,90
CCD4b	1,14		64,30	52,60
CCD4c	1,38	0,85		49,60
VP14 (X-Ray)	0,95	0,69	0,85	

Las principales diferencias en la estructura tridimensional observadas en los modelos se encuentran en los *loops*, que constituyen las regiones más móviles de la estructura como se puede apreciar en la Fig. 2.



Figura 2. Diagrama de cinta de los modelos obtenidos de la CCD4a, b, c y la VP14 indicados mediante un esquema de colores RGB en el diagrama de cinta, coloreados en una escala continua desde 0 a 1,5 Å, en la cual se colorean según tres rangos: valores bajos (en verde) medios (en amarillo) y altos (en rojo).

INSERCIÓN DEL DOMINIO α-HÉLICE EN LAS MEMBRANAS

El dominio de α -hélices de las CCD4s está formado por dos α -hélices antiparalelas, $\alpha 1$ y $\alpha 3$. En la Fig. 3, las secuencias alineadas de las estructuras secundarias helicoidales $\alpha 1$ y $\alpha 3$ se muestran y se colorean por su característica polar/carga. La CCd4b se utiliza como ejemplo, pero todos los modelos muestran una disposición similar. Para analizar en detalle, en la parte inferior de la figura, se alinearon las secuencias de todas las regiones helicoidales para todos los modelos. Es de notar que, en la CCD4b, la región helicoidal $\alpha 1$ comienza en el 45° aminoácido, mientras que en la CCD4a y CCD4c en el 86° y 78° respectivamente. Sin embargo, la estructura terciaria está muy conservada entre los miembros de la subfamilia CCD4, los primeros aminoácidos siendo eliminados en la etapa postraduccional.

Por un lado, $\alpha 1$ está conformado por 19-20 aminoácidos y predominan los aminoácidos polares, Por otro, $\alpha 3$ está conformado por 17 aminoácidos con predominio de hidrofóbicos, lo que hace posible una mejor interacción con el núcleo hidrofóbico de la membrana. Se hace evidente que, dada una mayor superficie polar o cargada del dominio $\alpha 1$, sus aminoácidos interactúan con la membrana tilacoide a través de interacciones polares con las cabezas de los grupos fosfolípidos.



Figura 3. a) Localización del dominio de α -hélices en la estructura de la CCD4a. El dominio $\alpha 1/\alpha 3$ está señalado con flechas rojas y diagrama de cintas del mismo color.

α1																						
Receptor	init	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	end
CCD4a	86	Α	Е	Ρ	Т	1	Ρ	Т	1	1	L	Ν	А	С	D	-	D	1	1	Ν	Ν	104
CCD4b	45	Р	Ι	Q	S	L	Μ	G	Т	Ν	S	S	Y	Ν	Т	К	S	А	Р	S	L	64
CCD4c	78	Т	Т	Т	Ν	V	S	А	V	1	L	Ν	Т	L	Ν	K	-	L	1	К	Р	96
α3																						
Receptor	init	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	en	d			
CCD4a	218	Т	А	S	Α	Α	R	G	А	L	S	Α	Α	R	L	L	Α	23	3			
CCD4b	180	L	V	D	Μ	V	Q	С	V	А	S	Т	A	R	V	L	M	19	5			
CCD4c	210	L	А	F	L	Т	R	А	G	V	L	Α	Α	R	V	L	Т	22	5			

Figura 3 (continuación). b) Secuencias alineadas del dominio helicoidal (α1 y α3) en los receptores CCD4a, b y c. La polaridad de los aminoácidos se representa en colores: cargados negativos (rojo); cargados positivos (azul); hidrofóbicos (verde); polares neutros (violeta). *Init* y *fin*: indican la posición inicial y final de las secuencias.

En la Fig. 4 se muestran dos representaciones del modelo de la proteína CCD4b insertada en la membrana. Por un lado, una representación en cinta del modelo insertado en la membrana permite ver las dos hélices $\alpha 1$ y $\alpha 3$ insertadas en la membrana. Por otro lado, las superficies accesibles al solvente están coloreadas por la distribución de carga y se muestran en dos perspectivas. En ambas representaciones, los grupos fosfato de la membrana están representados mediante esferas grises. Se muestra la interacción de los grupos hidrofílicos (superficies rojas y azules) de las α -hélices, y la de las regiones hidrofóbicas que interactuarán con el núcleo hidrofóbico al interior de la membrana.

Se puede observar que la penetración de la proteína en la membrana tilacoidal es tal que los grupos hidrófobos de las α -hélices interactúan con los ácidos grasos dentro de la membrana. Dependiendo del receptor, la penetración inicial en la membrana fosfolipídica fue de 5,2 a 7 Å, y con un valor promedio del $\Delta G_{transfer}$ de -58,1 kJ / mol (Tabla 2). Estos valores son consistentes con los observados para VP14, en la que se reportó una de penetración de 7 Å (Messing et al., 2010).

En consecuencia, el modelo propuesto no sugiere que sean proteínas transmembrana, sino posibles proteínas periféricas ancladas a las membranas tilacoidales. La proteína quedará anclada a la membrana tilacoidal y la mayor parte de su superficie se orientará hacia el estroma (la matriz al interior de los cloroplastos).



Figura 4. Superior: Representación en diagrama de cintas del modelo CCD4b insertado en la membrana (grupos fosfatos indicados mediante la superficie gris punteada). Inferior: una representación de la superficie molecular coloreada por carga/hidrofobicidad: hidrófobo (verde), negativo (rojo) y azul (positivo).

Una comparación entre todos los modelos de CCD4s anclados a la membrana, indica una orientación similar dentro de la membrana. En la Tabla 2 se muestran el ángulo, la profundidad y la energía libre de transferencia del solvente a la membrana ($\Delta G_{transfer}$). En la Fig. 5, se representan como cintas todos los modelos alineados de complejos de membrana-CCD4.



Figura 5. CCD4a (rosa), CCD4b (naranja) y CCD4c (cian) en complejos con la membrana representada por un plano de esferas grises definidas por las cabezas de fosfolípidos en las que se orientan los dominios $\alpha 1$ y $\alpha 3$. La flecha roja apunta al centro catalítico de Fe⁺² (esfera en cian).

Es evidente a partir de esta figura y de los valores que se muestran en la Tabla 2 que no existe una gran diferencia entre estos modelos.

Tabla 2. Estimación de los valores del ángulo de inclinación del primer eje principal de inercia con respecto al plano formado por los grupos fosfatos de la membrana, la profundidad de penetración en la membrana y la energía libre de transferencia del solvente a la membrana ($\Delta G_{transfer}$).

Proteína	Ángulo	Profundidad	$\Delta G_{transfer}(kJ/mol)$
CCD4a	32° (±4°)	7,0 (±0,6)	-64,4
CCD4b	39° (±4°)	5,8 (±1,3)	-61,5
CCD4c	28° (±4°)	5,2 (±0,5)	-48,5
Promedio	33°	6,0	-58,1

La visión de estos modelos nos permite proponer que la penetración de la región hidrofóbica formada por las α -hélices de las CCD4s coloca el canal de entrada del sustrato en las proximidades de la membrana tilacoidal, facilitando el acceso de los ligandos al receptor. Es

importante señalar que uno de los componentes más abundantes de la membrana tilacoidal son los carotenoides, observación que permite apoyar nuestra hipótesis.

ANÁLISIS ENERGÉTICO Y EQUILIBRIO TERMODINÁMICO

En la Tabla 3 se muestran los valores promedios y las desviaciones estándar correspondientes a las componentes de las energías de interacción: Van der Waals (VDW) y electrostática (ELECT), y la energía total de interacción (Uab) de los ligandos, para las trayectorias de la dinámica molecular de los complejos CCD4 a, b, c y todos los carotenoides aquí estudiados.

Tabla 3. Energía de interacción total (Uab) y sus componentes de Van der Waals (VDW) y electrostáticos (ELECT) promediados para todos los complejos. Los promedios se evalúan en función del tiempo para todas las trayectorias. Todos los valores están en kJ/mol. Valores menores de la Uab (Total) se resaltan en naranja.

Receptor	Ligando	Uab (ELECT)	Uab (VDW)	Uab (Total)		
		Media (±SD)	Media (±SD)	Media (±SD)		
	30N	-116,50 (±23,93)	-274,37 (±16,58)	-390,87 (±23,61)		
	ACR	-33,11 (±10,16)	-344,21 (±18,75)	-377,32 (±22,07)		
	BCR	-21,03 (±8,75)	-344,23 (±17,66)	-365,26 (±19,51)		
CCD4a	LYC	-32,66 (±10,96)	-351,83 (±15,51)	-384,49 (±18,94)		
	LUT	-110,95 (±21,48)	-342,63 (±17,54)	-453,58 (±23,83)		
	RRX	-77,24 (±16,85)	-333,07 (±17,81)	-410,30 (±21,77)		
	ZEX	-135,94 (±25,19)	-316,85 (±18,62)	-452,79 (±26,42)		
	30N	-124,99 (±27,10)	-271,27 (±18,53)	-396,27 (±24,13)		
	ACR	-33,41 (±10,80)	-351,57 (±16,37)	-384,97 (±21,24)		
	BCR	-31,76 (±10,65)	-342,07 (±14,83)	-373,83 (±18,16)		
CCD4b	LYC	-32,05 (±10,86)	-367,96 (±15,46)	-400,02 (±19,02)		
	LUT	-119,87 (±23,65)	-331,35 (±17,78)	-451,22 (±24,61)		
	RRX	-76,22 (±12,89)	-350,70 (±15,71)	-426,93 (±20,45)		
	ZEX	-121,58 (±25,71)	-348,27 (±18,43)	-469,85 (±26,72)		
	30N	-137,52 (±25,04)	-257,66 (±16,61)	-395,18 (±24,23)		
	ACR	-29,88 (±10,07)	-355,33 (±14,78)	-385,21 (±17,95)		
	BCR	-32,91 (±11,40)	-346,46 (±16,16)	-379,36 (±21,21)		
CCD4c	LYC	-36,42 (±11,65)	-363,98 (±17,23)	-400,40 (±19,95)		
	LUT	-137,28 (±28,76)	-318,32 (±18,68)	-455,60 (±26,78)		
	RRX	-83,85 (±18,59)	-329,94 (±18,81)	-413,79 (±23,63)		
	ZEX	-132,90 (±23,44)	-360,72 (±18,69)	-493,62 (±24,86)		

Los carotenoides son moléculas altamente hidrófobas. De manera complementaria, el sitio de unión es un canal lineal altamente hidrofóbico que favorece las interacciones con el ligando que está fuertemente empaquetado con interacciones gobernadas por interacciones de tipo VDW de corto alcance. Las interacciones polares también están presentes, pudiéndose entonces comprender y explicar que la contribución a la energía de interacción es mayor en las xantofilas en comparación con los carotenos alifáticos. Las xantofilas en sus regiones polares (sitios caracterizados por la presencia de oxígeno) tienden a formar enlaces de hidrógeno con moléculas de agua, o interactuar con otras moléculas polares del receptor.

En la Fig. 6 se muestra gráficamente la evolución de las energías de β -criptoxantina, el carotenoide más activo, en función del tiempo.



Figura 6. Evolución de la energía de interacción (Uab) y sus componentes (Elec y VDW) de la βcriptoxantina en complejo con el CCD4b a lo largo de la simulación de dinámica molecular. Se muestran las contribuciones electrostaticas (rojo), Van der Waals (verde) y la energía de interacción Uab total (azul).
ANÁLISIS ESTRUCTURAL (RMSD Y RMSF)

En la Fig. 7 se representa el análisis de los complejos RMSD de los modelos CCD4 a, b y c y los siete carotenoides en estudio.



Figura 7. Valores de RMSD en función al tiempo de simulación de dinámica molecular para los modelos de los complejos CCD4a (a), CCD4b (b) and CCD4c (c) y siete ligandos carotenoides: 3ON: 3-hidroxi-8'-apocarotenol, ACR: α-caroteno; BCR: β-caroteno; LUT: luteína; LYC: licopeno; RRX: β-criptoxantina: ZEX: zeaxantina.

Como se desprende de estos gráficos, los diferentes complejos presentaron una estabilidad global (energética y estructural).

Para evaluar la influencia de la membrana en la estabilidad estructural, se realizó una evaluación del RMSF de modelos con y sin membrana y se evaluaron las diferencias (Δ RMSF) entre ambos grupos de modelos. Los resultados se muestran en la Fig. 8.



Figura 8. Izquierda: Valores de Δ RMSF para CCD4a, CCD4b y CCD4c. El primer cuadrado azul punteado está señalando hélices α 1 y el segundo cuadrado de puntos azules α 3. Derecha: Estructuras 3D de todos los modelos insertados en la membrana. Regiones más móviles, corresponden a Δ RMSF > 0 se indican con el color rojo. Las hélices coloreadas en azul corresponden a regiones menos móviles y con Δ RMSF < 0.

Los dominios $\alpha 1$ y $\alpha 3$, al tener una interacción directa con la membrana, son las regiones que reciben el efecto de la presencia de la membrana, evidenciado una disminución de las fluctuaciones promedio (Δ RMSF negativo). Además, es posible observar que $\alpha 1$ tiene una coloración azul más intensa lo que indica una menor movilidad en comparación con la $\alpha 3$. Con respecto a las regiones que aumentan la movilidad (Δ RMSF positivo y de color rojo en la imagen 3D), son *loops* expuestos a la matriz líquida del estroma en el cloroplasto.

DISTANCIAS AL Fe⁺² Y ENERGÍAS LIBRES DE UNIÓN

Las energías libres de unión de los complejos CCD4 a, b y c con todos los carotenoides (ACR, BCR, LUT, RRX) en estudio y el apocarotenoide cocristalizado (3ON), se estimaron mediante el método MMPBSA a partir de conformaciones generadas en las simulaciones de dinámica molecular y se calcularon como promedios temporales y sus desviaciones estándar correspondientes.

Se muestran los valores promediados y las desviaciones estándar correspondientes de la energía libre de unión $\Delta G_{\text{binding}}$ estimada mediante MMPBSA, calculada para las estructuras extraídas de las trayectorias de dinámica molecular equilibradas de los complejos CCD4 a, b, c y todos los carotenoides aquí estudiados. Los resultados se muestran en la Tabla 4.

Como se puede observar en la Tabla 4, el anclaje y todo el evento catalítico se podría centrar principalmente alrededor de una región cerca del doble enlace C7 = C8 y C9 = C10 y sugiere que la catálisis ocurre en un sitio asimétrico.

Cuando las energías libres de unión de MMPBSA se consideran junto a la observación de distancias, se podría concluir que la tendencia del sitio catalítico de las oxidasas del tipo CCD4 de citrus se ubican cerca del doble enlace C9 = C10, promoviendo una ruptura no simétrica en esa región.

Con respecto a la especificidad del sustrato, se podría conjeturar que para el caso de CCD4a se favorecerá una oxidación asimétrica en C9 = C10 del β -caroteno como posible mejor sustrato, con la mejor energía de unión estimada de -162,54 kJ/mol, y separándose del resto, mientras que su afinidad por las xantofilas es menor (LUT, RRX y ZEX), dando cierta preferencia a los carotenos.

Tabla 4. Distancias (A) entre los átomos de carbono C7, C8, C9, C10, C10 y C15' de los carotenoides y el átomo de Fe⁺². Primera columna: carotenoides ensayados: ACR = α -caroteno; BCR = β -caroteno; LUT = luteína; RRX = β -criptoxantina y ZEX = zeaxantina. Las celdas que contienen los valores mínimos de distancias están resaltadas en naranja. Las energías libres de unión evaluadas en kJ/mol por el método MMPBSA se presentan en la última columna. La trayectoria de RRX * en CCD4b se extendió a 1 μ s.

Receptor	Ligando	C7	C8	С9	C10	C15	C15'	MMPBSA (±SD)
	ACR	9,3	8,7	8,4	8,2	10,6	11,9	-148,50 (±14,88)
	BCR	8,2	7,1	6,3	6,0	8,0	9,2	-162,54 (±17,60)
CCD4a	LUT	9,5	8,5	8,6	8,3	11,2	11,9	-136,63 (±17,78)
	RRX	7,4	7,3	7,3	8,4	10,9	11,2	-133,16 (±18,15)
	ZEX	7,8	7,04	6,5	7,4	10,2	11,1	-124,20 (±17,85)
	ACR	11,8	12,3	12,2	11,4	11,7	11,7	-123,18 (±17,91)
	BCR	11,1	11,2	10,7	9,5	9,4	10	-142,60 (±19,08)
	LUT	7,5	7,1	6,5	7,3	9,5	10,6	-119,18 (±29,12)
CCD4b	RRX	10,1	9,9	8,9	7,8	5,8	6,8	- 150,19 (±16,83)
	RRX*	8,9	8,0	6,8	6,9	7,9	8,6	-164,57 (±15,74)
	ZEX	7,0	6,7	6,2	7,1	10	11	-88,43 (±22,31)
	ACR	12,3	12	11,3	10	7,9	7,3	-106,32 (±19,08)
	BCR	10,9	9,9	8,5	8,2	5,8	6,3	-92,58 (±27,12)
CCD4c	LUT	11,4	12,2	11,9	11,6	11,6	11,3	-105,77 (±28,35)
	RRX	10,8	11,2	10,5	9,8	8,7	9,12	-128,21 (±20,68)
	ZEX	12,2	11,14	9,93	8,8	6,4	6,8	-99,74 (±19,03)

CCD4b-RRX* corresponde al complejo extendido a un tiempo de simulación de dinámica molecular de 1 µs.

En el caso de la CCD4c, si nos referimos a distancias con el Fe⁺² en el sitio catalítico, podría ser favorable una rotura no simétrica (C9 = C10) así como simétrica (C15 = C15²). En este caso, la β -criptoxantina tiene una energía de enlace de -128,21 kJ/mol obtenida como promedio de la trayectoria de 20 ns.

Prestamos especial atención a la enzima CCD4b en la que la energía libre de unión apunta a la β - criptoxantina como el mejor ligando. Como no existe una definición clara con respecto a la simetría de rotura, se obtuvo una trayectoria de microsegundos para confirmar este resultado y se presenta anotado como RRX *. Con la extensión de la simulación se observa una disminución en las distancias C9 = C10 con Fe⁺² sugiriendo a esta posición como posible sitio de corte. Para conocer con más detalle cómo interactúa el ligando RRX con el receptor CCD4b, se utilizó la herramienta de "*Ligand Interaction*" en MOE (Fig. 9) en la que se observan los principales aminoácidos que intervienen en la interacción, en una visión bidimensional. En la segunda parte de la Fig. 9, se aporta la visión tridimensional para el mismo complejo.



Figura 9. a) Representación gráfica 2D de la interacción de la β-criptoxantina en complejo con la CCD4b. Se muestran los residuos que entran en contacto con el carotenoide a menos de 4,5 Å. El significado de los colores y las sombras dibujadas como esferas sobre cada aminoácido se indica en la leyenda de la figura. b) Representación 3D en la que se señaliza el centro catalítico y especialmente el átomo de Fe⁺², mostrando la proximidad al doble enlace C9=C10.

ANÁLISIS DE LA INTERACCIÓN PROTEÍNA – MEMBRANA

En la Fig. 10, se muestra el modelo de la CCD4b insertada en la membrana, luego de 20 ns de dinámica molecular. La evolución a lo largo de la simulación del espesor medio de la membrana, la profundidad de penetración de la proteína dentro de la membrana y la penetración relativa de la proteína en la membrana se muestran en las Fig. 11 - 13.

El área que comprende la superficie de contacto observado a lo largo de la simulación de dinámica molecular entre los distintos complejos y la membrana muestra que la interacción es significativa y abarca una superficie no despreciable que va típicamente desde los 2250 a los 3250 Å². (Fig. 14). Esta superficie de contacto está dominada principalmente por los residuos que involucran al dominio α -hélice, al igual que lo observado en la estructura de la VP14 por Messing et al. (2010), lo que nos permite señalar como un mecanismo consistente también con las CCD4 de citrus, sugeridos previamente en los estudios de Ma et al. (2013), estableciendo una interacción estable que debido a la orientación y profundidad de penetración en la superficie de la membrana tilacoidal permitiría a las CCD4 a extraer a los carotenoides solubles en la membrana a través de un túnel formado principalmente por residuos hidrofóbicos y que se extiende hacia el centro catalítico de las enzimas, este túnel hidrofóbico fue previamente caracterizado en el Capítulo III.

Por último, para complementar el análisis de las interacciones entre las proteínas y la membrana, se registró y graficó la evolución de los enlaces de hidrógeno a lo largo de la simulación (Fig. 15). El promedio de enlaces de hidrógenos observados fueron: 9 para el CCD4a, 7 para CCd4b y 10 para CCD4c. Por tanto, no existe una diferencia significativa a este respecto entre las diferentes proteínas que indican que el mecanismo de unión proteína-membrana es similar en todas ellas.



Figura 10. CCD4b insertado en una bicapa de fosfatidilcolina después de la etapa de 20 ns. Los grupos fosfato están alineados en el plano XY y proyectados en el eje Z. MP2 es el conjunto de todos los grupos fosfato en el lado de la membrana en el que se insertó la proteína. MP1 es el conjunto de todos los grupos fosfato que pertenecen a la otra capa de la membrana. $Max(Z_{prot})$ es la máxima posición atómica observada de la proteína a lo largo del eje Z (sin considerar los átomos de hidrógeno). Z1 y Z2 son los valores promedios de las coordenadas en el Z de MP1 y MP2, respectivamente.



Figura 11. Espesor promedio de la membrana para los complejos CCD4a, CCD4b y CCD4c en el tiempo, coloreados según un código que se indica en el lado derecho de la figura.



Profundidad de penetración en la membrana

Figura 12. Profundidad de inserción de las proteínas en la membrana medida para los tres complejos estudiados, coloreados según el código indicado a la derecha



Figura 13. Penetración relativa de la proteína en la membrana (RPD) evaluada a lo largo de las trayectorias de dinámica molecular para la CCD4a, CCD4b y CCD4c.



Superficie de contacto

Figura 14. Evolución de la superficie de contacto (Å²) de los complejos con la membrana: CCD4a (rojo), CCD4b (verde) y CCD4c (azul). CCD4a obtuvo una media de 2797.15 (±153.89), CCD4b: 2674.08 (±163.75) y CCD4c: 2885.87 (±118.13).



Enlaces de Hidrógeno entre proteínas y membranas

Figura 15. Evolución temporal de los enlaces de hidrógeno entre proteínas y la membrana. CCD4a (rojo), CCD4b (verde) y CCD4c (azul). El promedio de los enlaces de hidrógeno fue nueve para CCD4a, siete para CCD4b y 10 para CCD4c.

CONCLUSIONES

Se han construido nuevos modelos de los miembros de la subfamilia CCD4 provenientes de *Citrus clementina*, incluyendo dominios que interaccionan con las membranas, para estudiar la posible interacción con las membranas tilacoidales. Los modelos han sido validados mediante procesos de simulación de dinámica molecular en agua en trayectorias que han mostrado estabilidad estructural a lo largo de la simulación.

Se aportan evidencias de que los dominios de las hélices $\alpha 1$ y $\alpha 3$ incluidos en estos modelos son esenciales para la interacción con las membranas tilacoidales y que en presencia de estas la fluctuación promedio en estos dominios disminuye estabilizando la estructura. Mientras que $\alpha 1$ se caracteriza por el dominio de aminoácidos polares que interactúan con los grupos fosfatos de la membrana, $\alpha 3$ se caracteriza por la abundancia de aminoácidos apolares que favorecen la interacción con el núcleo hidrofóbico de la membrana, logrando una penetración de 10 Å que constituye una penetración relativa de aproximadamente 30-35% de la membrana (Fig. 12-13).

Mientras que en la energía de interacción Uab se observa una preferencia en la unión con las xantofilas (ZEX, LUT y RRX), cuando se mide por el método MMPBSA se observa un panorama similar pero que también incluye a los carotenos como BCR como posibles sustratos, en concordancia con lo que se sabe a partir de datos experimentales que las xantofilas y los carotenos se unen a CCD4 (Ahrazem et al., 2016).

Finalmente, el posible sitio de escisión oxidativa muestra una tendencia claramente asimétrica en las posiciones C7 = C8 y C9 = C10 que es consistente con la de otros miembros ya conocidos del tipo CCD4 (Ahrazem et al., 2016; Rodrigo et al., 2013).

REFERENCIAS

- Ahrazem, O., Gómez-Gómez, L., Rodrigo, M. J., Avalos, J., & Limón, M. C. (2016). Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: Features and functions. International Journal of Molecular Sciences, 17(11). <u>https://doi.org/10.3390/ijms17111781</u>
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ , ψ and side-chain $\chi 1$ and $\chi 2$ Dihedral Angles. Journal of Chemical Theory and Computation, 8(9), 3257–3273. <u>https://doi.org/10.1021/ct300400x</u>
- Britton, G. (1990). Biosynthesis of Chloroplast Carotenoids BT Current Research in Photosynthesis: Proceedings of the VIIIth International Conference on Photosynthesis Stockholm, Sweden, August 6–11, 1989 (M. Baltscheffsky (ed.); pp. 2733–2740). Springer Netherlands. <u>https://doi.org/10.1007/978-94-009-0511-5_620</u>
- Jeffrey, S. W., Douce, R., & Benson, A. A. (1974). Carotenoid transformations in the chloroplast envelope. Proceedings of the National Academy of Sciences of the United States of America, 71(3), 807–810. <u>https://doi.org/10.1073/pnas.71.3.807</u>
- Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., ... Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. Protein Science, 27(1), 112–128. <u>https://doi.org/10.1002/pro.3280</u>

- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2016). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10(6), 845–858. https://doi.org/10.1038/nprot.2015-053
- Kloer, D. P., & Schulz, G. E. (2006). Structural and biological aspects of carotenoid cleavage. Cellular and Molecular Life Sciences, 63(19–20), 2291–2303. https://doi.org/10.1007/s00018-006-6176-6
- Li, P., Roberts, B. P., Chakravorty, D. K., & Merz, K. M. (2013). Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. Journal of Chemical Theory and Computation, 9(6), 2733–2748. <u>https://doi.org/10.1021/ct400146w</u>
- Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. Nucleic Acids Research, 40(D1), 370–376. <u>https://doi.org/10.1093/nar/gkr703</u>
- Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., Wahyudi, A., Motohashi, R., & Kato, M. (2013). Enzymatic formation of β-citraurin from β-cryptoxanthin and zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. Plant Physiology, 163(2), 682–695. <u>https://doi.org/10.1104/pp.113.223297</u>
- Messing, S. A. J., Mario Amzel, L., Gabelli, S. B., Echeverria, I., Vogel, J. T., Guan, J. C., Tan, B. C., Klee, H. J., & McCarty, D. R. (2010). Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. Plant Cell, 22(9), 2970– 2980. <u>https://doi.org/10.1105/tpc.110.074815</u>
- Parry, A. D., & Horgan, R. (1991). Carotenoids and abscisic acid (ABA) biosynthesis in higher plants. Physiologia Plantarum, 82(2), 320–326. <u>https://doi.org/10.1034/j.1399-3054.1991.820228.x</u>
- Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak, B. K., Skeel, R. D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., ... Tajkhorshid, E. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. Journal of Chemical Physics, 153(4). https://doi.org/10.1063/5.0014475
- Rascio, N. (2013). Chloroplasts. Encyclopedia of Biological Chemistry: Second Edition, 506–510. https://doi.org/10.1016/B978-0-12-378630-2.00141-9
- Rodrigo, M. J., Alquézar, B., Alís, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461– 4478. <u>https://doi.org/10.1093/jxb/ert260</u>

- Rodríguez-Villalón, A., Gas, E., & Rodríguez-Concepción, M. (2009). Colors in the dark: A model for the regulation of carotenoid biosynthesis in etioplasts. Plant Signaling and Behavior, 4(10), 965–967. <u>https://doi.org/10.4161/psb.4.10.9672</u>
- Rubio-Moraga, A., Rambla, J. L., Fernández-De-Carmen, A., Trapero-Mozos, A., Ahrazem, O., Orzíez, D., Granell, A., & Gimez-Gimez, L. (2014). New target carotenoids for CCD4 enzymes are revealed with the characterization of a novel stress-induced carotenoid cleavage dioxygenase gene from Crocus sativus. Plant Molecular Biology, 86(4–5), 555–569. https://doi.org/10.1007/s11103-014-0250-5
- Tan, B. C., Cline, K., & McCarty, D. R. (2001). Localization and targeting of the VP14 epoxycarotenoid dioxygenase to chloroplast membranes. Plant Journal, 27(5), 373–382. <u>https://doi.org/10.1046/j.1365-313X.2001.01102.x</u>
- Tan, Bao Cai, Joseph, L. M., Deng, W. T., Liu, L., Li, Q. B., Cline, K., & McCarty, D. R. (2003). Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant Journal, 35(1), 44–56. <u>https://doi.org/10.1046/j.1365-313X.2003.01786.x</u>
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & MacKerell Jr., A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry, 31(4), 671–690. https://doi.org/10.1002/jcc.21367.CHARMM
- Vega-Teijido, M., Cantero, J., Rodrigo, M. J., López, C., & Paulino Zunini, M. (2018). An In Silico Study of the Citrus Dioxygenases CCD4 family Substrates. Journal of Biomolecular Structure and Dynamics, 0(0), 1–28. <u>https://doi.org/10.1080/07391102.2018.1477619</u>

CAPÍTULO V CONCLUSIONES Y PERSPECTIVAS

CONCLUSIONES

Del conjunto de resultados de esta Tesis se pueden extraer las siguientes conclusiones:

Se han identificado 12 genes de enzimas de clivaje de carotenoides (CCDs) en *Citrus clementina* (CCD1, CCD1*, NCED2, NCED3, CCD4a, CCD4b1, CCD4b2, CCD4c, CCD4d, NCED6, CCD7, CCD8), que contiene entonces 5 miembros de la subfamilia CCD4. El análisis de las secuencias revela que las CCD4s forman un grupo claramente separado de las demás subfamilias.

Los motivos conservados identificados de las secuencias CCD4s de origen vegetal, presentan un largo variable entre 30 a 50 aminoácidos. Las secuencias consenso de tales motivos poseen una composición similar, lo que sugiere que estas regiones se organizan en plegamientos similares conformando láminas- β en la arquitectura típica de β -propeller. Tales plegamientos son

característicos de las enzimas de origen vegetal que clivan carotenoides, como se comprobó al alinear los motivos sobre la estructura de la VP14 de *Zea mays* (PDB: <u>3NPE</u>).

La predicción de la estructura secundaria muestra un claro predominio de las láminas- β a lo largo de las secuencias de las CCD4, resultado consistente con la arquitectura de esta familia de proteínas cuya arquitectura se describe como β -propeller. Las α -hélices aparecen al inicio de la secuencia luego de la región N-teminal desestructurada (> 100 aa) y las mismas fueron identificadas como probables dominios de interacción con la membrana tilacoidal. Este dominio α -hélices se postula como el domo hidrofóbico de interacción con las membranas consistente con lo observado por Messing et al., (2010) en la VP14 de *Zea mays*.

La CCD4b2 y la CCD4d fueron predichas como duplicaciones de la CCD4b1 y CCD4c, respectivamente. Estas duplicaciones en tándem explican la mayor variabilidad de las mismas debido a la menor presión selectiva y a la pérdida de un motivo fuertemente conservado en la CCD4d que afecta a una de las histidinas conservadas en la esfera de coordinación del Fe⁺², sugiriendo la pérdida de su capacidad catalítica.

Se ha predicho la presencia de péptidos de destinación al cloroplasto (cpTP) en la mayoría de las CCDs de *Citrus clementina* con las excepciones de las CCD1 y CCD7 en las cuales no se ha podido identificar por ninguno de los métodos empleados.

El análisis de la región N-terminal de los miembros de la CCD4 que abarca aproximadamente unos ~100 aminoácidos muestran una región de alta variabilidad con un alto grado de desorden y estas posiciones marcan una clara tendencia a coevolucionar. En esta región se encuentra la señal cpTP y posteriormente, una vez que la enzima es traslocada al interior del cloroplasto duramente la maduración, es eliminada de la estructura (Lee y Hwang, 2018).

El análisis filogenético de las CCD4s demostró la cercana relación entre la CCD4a, CCD4c y CCD4d mientras que la CCD4b1 y CCD4b2 fueron las más lejanamente emparentadas con los demás miembros CCD4 de *Citrus clementina*.

Se obtuvieron modelos moleculares validados para la CCD4a y CCD4b de *Citrus clementina* construidos por homología partiendo de la estructura cristalográfica de la ACO (PDB ID: 2BIW) con una serie de seis diferentes ligandos carotenoides anclados por métodos de *docking* y posterior cálculo de dinámica molecular que confirman la estabilidad de la típica estructura terciaria como los descritos para los homólogos de la CCD4.

Se identificó una región flexible conformada por *loops* largos y α -hélices cortas y una región rígida conformada por láminas- β que contienen al Fe⁺² en el centro del sitio catalítico. El centro catalítico está definido el átomo metálico, que en una primera esfera de coordinación contiene cuatro residuos de histidina, que se puede ampliar a una segunda esfera conteniendo ácidos glutámicos y aspárticos.

Durante la simulación de dinámica molecular, los residuos y moléculas de agua alternan posiciones cercanas al Fe^{+2} lo que sugiere un entorno flexible, con la variabilidad y adaptabilidad suficiente como para que la estructura y dinámica de la esfera de coordinación acompañe adecuadamente el evento catalítico.

Los resultados de *docking* de los carotenoides en el sitio activo de la CCD4a sugieren que la unión es más favorable a la zeaxantina que la β -criptoxantina y menos favorable para el β -caroteno. Los cálculos de dinámica molecular de la CCD4a muestran que la energía de interacción promedio sobre las trayectorias están en el mismo orden de unión: zeaxantina > β -criptoxantina >

En el caso de la CCD4b se observó que, de acuerdo con el resultado experimental, para todos los carotenoides con anillos β , el orden de la preferencia de unión fue: luteína > zeaxantina > β -criptoxantina > α -caroteno > licopeno > β -caroteno. Estos resultados muestran mejores comportamientos de las xantófilas comparadas con los carotenos, siendo la polarizabilidad y la presencia de anillos β hidroxilados importantes para la unión con la enzima.

La isomería geométrica de los carotenoides a quienes se les asignaron una configuración inicial con todos sus enlaces trans (*all-trans*) se conservó durante las simulaciones de dinámica molecular en la mayoría de los ligandos. De cualquier manera, el licopeno (caroteno alifático) sufrió una isomerización geométrica de trans a cis de uno de sus dobles enlaces conjugados.

Modelos de las CCD4 (a, b y c) provenientes de *Citrus clementina*, incluyendo dominios α -hélices que interaccionan con las membranas, han mostrado estabilidad estructural a lo largo de la simulación de dinámica molecular que, sumando todas las simulaciones realizadas, acumula un tiempo de 4,5 µs.

Se aportan evidencias de que los dominios de las hélices $\alpha 1$ y $\alpha 3$ incluidos en estos modelos, son esenciales para la interacción con las membranas tilacoidales y que en presencia de estas la fluctuación promedio en estos dominios disminuye estabilizando la estructura.

La región que involucra a α 1 se caracteriza por el dominio de aminoácidos polares que interactúan con los grupos polares de la membrana. La hélice α 3 se caracteriza por la abundancia de aminoácidos apolares que favorecen la interacción con el núcleo hidrofóbico de la membrana, logrando una penetración de 10 Å que representa una penetración relativa de aproximadamente 30-35% del grosor de la membrana.

El análisis energético señala la preferencia en la unión con las xantofilas (ZEX, LUT y RRX), sin excluir como posibles sustratos a carotenos como el BCR, en concordancia con lo que se sabe a partir de datos experimentales (Ahrazem et al., 2016).

Finalmente, el posible sitio de escisión oxidativa muestra una tendencia claramente asimétrica en las posiciones 7=8 (7'=8') y 9=10 (9'=10'), resultado consistente con otros miembros ya conocidos del tipo CCD4 (Ahrazem et al., 2016; Rodrigo et al., 2013).

PERSPECTIVAS

De la experiencia adquirida durante las investigaciones realizadas en este trabajo de tesis, señalamos las perspectivas y posibles campos que se abren dentro de esta área:

El estudio de las secuencias de la subfamilia CCD4 ha mostrado una alta variabilidad y heterogeneidad en la región N-terminal (<100 AA) que corresponde al péptido de destinación al cloroplasto (cpTP). Un estudio más exhaustivo de esta región podría revelar mayores detalles de cómo coevolucionan con el resto de la estructura de las CCDs, las señales de localización al cloroplasto.

Se ha observado en una de las simulaciones de los complejos proteína-ligando, la isomerización de un carotenoide (licopeno). Una isomerización *in silico* es un evento inesperado durante una dinámica molecular de equilibrio dada la alta energía relativa requerida para un cambio de configuración cis-trans de un doble enlace y también debido al espacio restringido que ocupa el ligando en el sitio activo. Las isomerizaciones son eventos comunes *in vivo*, y pueden estar mediados por enzimas con actividad de isomerasas, en las que ellas promueven cambios en los enlaces de los ligandos y que están muy cercanamente relacionadas a las CCDs en su estructura (Kloer y Schulz, 2006). Por todas estas razones una investigación más profunda de isomerasas de carotenoides se puede postular como una perspectiva interesante.

Se sugieren investigaciones más profundas sobre la geometría y el espacio conformacional de los carotenoides dado el alto grado de flexibilidad de los mismos. Así también la inclusión de isómeros geométricos en el estudio de la preferencia y selectividad de las CCDs, todos ellos siendo aspectos no cubiertos en esta tesis.

Debido a que el centro catalítico describe un entorno altamente cargado y las limitaciones de los campos de fuerza para describir este tipo de entornos son conocidos se recomienda la incorporación de métodos del tipo Mecánica Cuántica (QM) y métodos hídridos Mecánica Molecular/Mecánica Cuántica (QM/MM), los que aportarían más información del comportamiento electrónico y la catálisis enzimática.

La exploración de otros métodos de simulación basados en dinámica molecular tales como la *Steered Molecular Dynamics* (SMD) o Metadinámica en conjunto con la extensión de los tiempos de simulación podrían revelar más detalles de la interacción proteína-ligando y también de las interacciones proteína-membrana.

La fosfatidilcolina (POPC) es un componente común en los modelos de membranas biológicas, sin embargo, modelos más específicos de la composición de la membrana tilacoidal utilizando principalmente glicolípidos, previa validación de los mismos, podrían dar una descripción más realista de los fenómenos de interacción proteína-membrana.

REFERENCIAS

- Ahrazem, O., Gómez-Gómez, L., Rodrigo, M. J., Avalos, J., & Limón, M. C. (2016). Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: Features and functions. International Journal of Molecular Sciences, 17(11). <u>https://doi.org/10.3390/ijms17111781</u>
- Lee, D. W., & Hwang, I. (2018). Evolution and design principles of the diverse chloroplast transit peptides. Molecules and Cells, 41(3), 161–167. <u>https://doi.org/10.14348/molcells.2018.0033</u>
- Messing, S. A. J., Mario Amzel, L., Gabelli, S. B., Echeverria, I., Vogel, J. T., Guan, J. C., Tan, B. C., Klee, H. J., & McCarty, D. R. (2010). Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. Plant Cell, 22(9), 2970– 2980. https://doi.org/10.1105/tpc.110.074815
- Rodrigo, M. J., Alquézar, B., Alís, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., & Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of

Citrus fruit-specific apocarotenoid pigments. Journal of Experimental Botany, 64(14), 4461–4478. <u>https://doi.org/10.1093/jxb/ert260</u>



Anexo 1



Check for updates

An in silico study of the citrus dioxygenases CCD4 family substrates

Mauricio Vega-Teijido^{a,b}, Jorge Cantero^{a,d}, Maria J. Rodrigo^c, Carolina López^a and Margot Paulino Zunini^a

^aCentro de Bioinformática Estructural (CeBiolnfo), Departamento de Experimentación y Teoría de la Estructura de la Materia y sus Aplicaciones (DETEMA), Facultad de Química, Universidad de la República, Montevideo, Uruguay; ^bComputational Chemistry and Biology Group (CCBG), DETEMA, Facultad de Química, Universidad de la República, Montevideo, Uruguay; ^cLaboratorio Fisiología y Biotecnología Postcosecha, Instituto de Agroquímica y Tecnología de Alimentos, IATA-CSIC, Valencia, Spain; ^dDepartamento de Investigación, Facultad de Agronomía, Universidad Nacional del Este, Ciudad del Este, Paraguay

Communicated by Ramaswamy H. Sarma

ABSTRACT

The coloration of Citrus fruits is related with the concentration of carotenoids, isoprenoid pigments of 40 carbon atoms (C_{40}). Rodrigo et al. and Ma et al. reported a CCD4-type citrus dioxygenase responsible for the generation of C₃₀ apocarotenoids providing a reddish-orange pigmentation to the peel of many mandarins and oranges. Among them, CCD4b was the first case described of a dioxygenase that cleaves carotenoids C_{40} in the double bond 7', 8' or 7, 8, generating β -citraurin or 8- β -apocarotenal. Here we report the three-dimensional structures of CCD4a and CCD4b, modeled by sequence homology (2BIW) and validated by molecular dynamics (MD). Docking calculations were performed in CCD4a and CCD4b structures with thousands of rotated initial carotenoid conformations and all the possible poses in the active site were found. The interaction energy was measured by means of ASE scoring, Amber99 refinement and London ΔG rescoring. For the case of CCD4a model, the results showed London ΔG score of -19, -17 and -15 kcal/mol for zeaxanthin, β -cryptoxanthin and β -carotene, respectively. The same sequence in the estimated interaction strength for the three ligands was obtained using MD. The interaction energy of CCD4b indicated that, in agreement with experimental data, zeaxanthin and β -cryptoxanthin could be cleaved by the enzyme, β - and α -carotene have chances to be oxidized and lycopene has not good interaction energy to be predicted as substrate. These findings will be discussed considering the potential in vivo substrates and products, and the physiological role in Citrus fruits.

1. Introduction

In citrus, the color of the peels and juices is associated by the consumer with the fruit quality and a major parameter for its acceptance. Citrus fruits are at the top of fruit crops produced in the world with more than 120 MT (FAO, 2016) add guery. Therefore, investigations focused on the biochemical and molecular basis of coloration of citrus fruits has gained an increased interest in the last years and is an active area of research. In citrus, coloration is due to the accumulation of carotenoids, which are isoprenoid pigments of 40 carbon atoms (C40) with a high conformational flexibility. Recently, Rodrigo, Alquézar, Alós, Medina, et al. (2013) and Ma et al. (2013), reported a series of Carotenoid Cleavage Dioxygenases (CCD) 4-type in Citrus involved in the generation of C30 apocarotenoids (β -apo-8'-carotenal and β -citraurin (Figure 1b), which provide the characteristic reddish-orange pigmentation to the peel of mandarins and oranges. In citrus the CCD4 family contains five genes: CCD4a, b1, b2, c and d, although b2 and d can be considered pseudogenes since they are not expressed and the predicted proteins are truncated with key catalytic residues missing (Rodrigo, Alquézar, Alós, Medina, et al., 2013). In agreement, the allelic diversity and expression

ARTICLE HISTORY

Received 8 December 2017 Accepted 12 April 2018

KEYWORDS

carotenoids; dioxygenases; citrus; CCD4

level of the five CCD4 gene members in different citrus species including orange, mandarin and pummelo, supported a crucial role of CCD4 enzymes in fruit peel coloration in citrus (Zheng et al., 2015).

To date the activity of many plant CCD4 enzymes have been examined by using *in vitro* and *in vivo* assays and the main activity of CCD4 was shown using β -carotene as substrate at the preferential cleavage positions at 9, 10 (9', 10') producing C₁₃ β -ionone, while hydroxylated xanthophylls did not show detectable activity (Huang, Molnár, & Schwab, 2009). By using Arabidopsis plant mutants defective in CCD4 activity it has been suggested that the epoxy- β -xanthophylls (mainly violaxanthin) are candidates the be *in vivo* substrates in *Arabidopsis* leaves to produce C₁₃ apocarotenoids (cleavage occuring at 9, 10 or 9', 10' double bonds) (Gonzalez-Jorge et al., 2013; Lätari et al., 2015).

Analysis of carotenoids substrates and cleavage products in seeds and roots of CCD4 mutants also indicate that CCD4 may accept β -carotene as substrate (Gonzalez-Jorge et al., 2013; Lätari et al., 2015). In potato, *in vitro* and *in vivo* assays on CCD4 points to β -carotene as the main substrate cleaved at 9, 10 or 9', 10' positions (Bruno, Beyer, & Al-Babili, 2015). In saffron, *in vivo* assays with CCD4a/b showed that β -carotene and zeaxanthin are mainly cleaved at 9, 10(9', 10') double bonds while other member of the family, CCD4c, cleaves other xanthophylls such as lutein (Rubio et al., 2008).

In vivo assays of two different grapevine CCD4s indicate that these enzymes cannot process cyclic carotenoids, (being the monocyclic ε-carotene an exception), but have activity against linear carotenes (neuroporene, ζ--carotene) preferentially at 9, 10 (9', 10') position. In rape (Brassica napus) flowers the activity of CCD4 suggests cleavage activity on α -carotene or δ -carotene at 9, 10 position to render α --ionone (Zhang et al., 2015). In Chrysanthemum morifolium the activity of a specific CCD4 present in flower petals was associated with the carotenoids degradation and pointed as responsible for the white coloration of the flowers (Ohmiya Kishimoto, Aida, Yoshioka, & Sumitomo, 2006). Other possible ligands and products are still under study and is one aim of our work. One example of the CCD4 role in pigmentation has been described in peach (Prunus persica) cultivars. Brandi et al. (2011), reported that white-fleshed peach contains at least 10-times less concentration of carotenoids and higher concentration of volatile derivatives than the yellow ones (Brandi et al., 2011). Recently, other works reported mutations on the CCD4 gene that causes accumulation of carotenoids in the yellow varieties (Adami et al., 2013; Falchi et al., 2013; Fukamatsu, Tamura, Hihara, & Oda, 2013; Ma et al., 2014). The citrus CCD4b (formerly described as CCD4b1) was the first case reported of a dioxygenase that cleaves carotenoids C40 in the double bond 7', 8' or 7, 8.

As potential in vivo substrates for CCD4b enzyme in citrus has been suggested zeaxanthin and β-cryptoxanthin (Ma et al., 2013; Rodrigo, Alquézar, Alós, Medina, et al., 2013). Zeaxanthin and β -cryptoxanthin are di- and mono-hydroxylated carotenoids, respectively, and are members of the carotenoids class known as xanthophylls (Figure 1a). Both xanthophylls are actively synthetized in orange-colored citrus fruits during ripening (Kato, 2012; Rodrigo, Alquézar, Alós, Lado, & Zacarías, 2013). In the case of CCD4b, in assays with zeaxanthin and β-carotene (Rodrigo, Alquézar, Alós, Medina, et al., 2013) the unique product showed a retention time and spectrum that were coincident with the used standards of 3-hydroxy- β -cyclocitral and the non-OH analog β -apo-8'carotenal, respectively. Such results indicate a cleavage reaction at the 7', 8' or 7, 8 position of the carotenoid backbone. In Figure 1b is shown the cleavage for zeaxanthin. Two products were simultaneously identified when β -cryptoxanthin was assayed as substrate: β -apo-8'-carotenal and β -citraurin (Figure 1b). This mixture of products is resulting from an asymmetric cleavage at the 7', 8'or 7, 8 double bonds. However, it was detected a slight preference of the enzyme to produce β -apo-8'-carotenal, that means, a preferred cleavage at the 7', 8' bond near the hydroxylated ring. Besides this evidence, two C₃₀ derivatives are possible: the β -apo-8'-carotenal (when in the catalytic site is posed the hydroxylated ring) and the β-citraurin on the other case.

For citrus CCD4b enzyme other carotenoids, such as β and α -carotene, lutein and lycopene were tested as substrates in *in vitro* assays (Ma et al., 2013; Rodrigo, Alquézar, Alós, Medina, et al., 2013). The β -carotene (Figure 1a) was cleaved and the resulting product was β -apo-8'-carotenal while both lutein and α -carotene (Figure 1a) were only cleaved at the side of the β -ionone ring. The activity of CCD4b on linear carotenoids was tested in two works using lycopene. In this case, no activity or a very low activity was evidenced (Ma et al., 2013; Rodrigo, Alquézar, Alós, Medina, et al., 2013).

In summary, for most of the plant CCD4 analyzed the *in vitro* and *in vivo* assays point to β -carotene as the preferred substrates and 9, 10 or/and 9', 10' double bonds cleavage as the preferred positions. However, citrus CCD4b is a relevant exception regarding substrate or cleavage site (Ohmiya et al., 2006; Campbell et al., 2010; Rodrigo, Alquézar, Alós, Medina, et al., 2013; Bruno et al., 2015), which can be a key in the understanding of CCD4 enzymes' selectivity.

The *in silico* strategies applied to build the three-dimensional structure of both, enzymes and enzyme-substrate complexes are appreciated tools to complement experimental results. The structural knowledge in terms of docking and MD simulation can give a deeper viewing of the forces that rule the molecular complexes between ligands and receptors. This information can help to understand the bases of the reactivity and selectivity of CCD4 enzymes with carotenoid substrates, thus, allow conjecturing about the influence in the color and other sensorial *in vivo* characteristics of citrus.

In this communication we report unpublished threedimensional structures of dioxygenases of citrus belonging to the CCD4 family. Two subtypes, a and b, were modeled and validated by means Molecular Dynamics (MD). Additionally, molecular complexes of both enzymes were built and studied. The binding mode and their molecular interactions with carotenoid ligand (zeaxanthin, lycopene, α and β -carotene, lutein and β -cryptoxanthin) is analyzed and reported here.

In literature, recently were reported new crystallographic and homology models of another variant of these type of enzymes (5kja and 5j53 PDB codes, Sui et al., 2015 and McAndrew et al., 2016). Although, the dynamic behavior of the structure and their complexes with carotenoids remains unclear, these are the main targets of our research.

2. Materials and methods

2.1. Software, hardware, force field and parametrization

Models and simulations were done using workstations with a Quad Core Intel i7 processor. The calculations were performed by means Molecular Operating Environment (MOE) 2013 (Chemical Computing Group, 2009) suite on Linux OS (OpenSUSE 12.3) and Windows OS (Vista and 7).

The force field selected to the simulation was Amber96. Under the MOE environment, is able to assign consistent parameters using an algorithm of pattern recognition, identifying—and specially for the case of carotenoids—the atoms in the ligand and assigning the parameters depending on



Figure 1. (a) Six carotenoids submitted to in silico and experimental studies, (b) zeaxanthin substrate and the two products resulting of the cleavage in the 7 = 8 double bond: β -citraurin and 3-hidroxy- β -cyclocitral. Additionally the β -apo-8'-carotenal structure is depicted (inside the dashed box), which is the apocarotenoid product when the substrate is β -carotene or β -cryptoxanthin with the hydroxylated ring inside the active site.

the element (C, O or H in this case) the atom type (all yet assigned in Amber96), the connectivity, the valence and geometry of atoms. Finally, charges are assigned as the Amber96 definition. As a result, the non covalent interactions evaluated through the docking and molecular dynamics could be considered enough precise. Further validation could be considered the agreement with the experimental observations

2.2. Homology modeling of proteins

2.2.1. Sequence alignment, selection of the best template and homology modeling

Homology models of CCD4a and b were built after a sequence alignment with 2BIW template (PDB code), and then, by replacing side chains, keeping the main chain of 2BIW structure. The selection of the template was done by the modeling software from the Protein Data. The sequence



Figure 1. (Continued)

alignments were done using BLOSUM62 substitution matrix with default settings (Silvarrey et al., 2016).

Given the low identity percentage between model and template, the homology modelling step was considered an initial step to obtain a set of pre-mature models. After, a graphical inspection of Fe(II) coordination sphere and the active site and the pose of a carotenoid (over the co-crystallized apo-carotenoid) without clashes defined a model with which continue a refinement process. Then, the structure was submitted to many local optimization processes and short molecular dynamics to, finally, obtain a model with which make the docking step. In summary, the model building consisted in an homo/dynamic strategy. After docking, each docked model was validated through long molecular dynamics to assure the stability and adequacy of final models. The models were submitted to an energy minimization in an explicit box of water of *ca*. 6 Å radius, and then, relaxed by a 5 ns molecular dynamics simulation to test their stability. The Amber99 force field and periodic boundary conditions were set as simulation options (Weiner, Kollman, Nguyen, & Case, 1986; Wang, Cieplak, & Kollman, 2000).

2.2.2. Energy minimization and molecular dynamics relaxation of crystallographic template 2BIW and CCD4 holo models with apocarotenoid

To validate all models, the crystallographic structure of the template, together with it co-crystallized ligand, an apocarotenoid, were taken as test molecules.

The template is the crystal structure of *Synechocystis sp.* native apocarotenoid cleavage oxygenase (E.C.1.13.11.75,

PDB Code 2BIW), expressed in *Escherichia coli*. The resolution is of 2.39 Å, with a R-factor: 0.182 and R-free: 0.224. It is an all-trans-8'-apo- β -carotenal 15, 15'-oxygenase which catalyzed an oxidation of all-trans-8'-apo- β -carotenal to give all-trans-retinal and (2E,4E,6E)-2,6-dimethylocta-2,4,6-trienedial. It has a bound ligand (Het Group name =3ON) and a Fe(II) as non-heme cofactor.

The CCD4a-apocarotenoid model was build copying the initial structure of the ligand in the crystallographic structure. The 2BIW sequence matches with 37.1% similarity with respect to that of CCD4a. The CCD4a three-dimensional structure was built as the first stage of modeling and, in addition to performing the study of its interaction with carotenoids, its structure was used as template to build the of CCD4b model as will be explained later.

Starting with this crystallographic or modeled structure coordinates, a refinement through energy minimization was performed. The Amber99 force field was used to energy minimization and for further molecular dynamics to reach energetically equilibrated structures. Amber99 is a force field that includes the parameters for all the atoms commonly found in proteins and nucleic acids. MOE program uses Gasteiger partial charges (Gasteiger & Marsili, 1980) and/or a generic force field to optimize when is needed to set missing parameters. A solvent water box was built and refined by energy minimization again. In the water box, two crystallographic waters were retained because they are connecting the ligand with the Fe(II) in the catalytic site. In the final model, the water box includes ca. 12,500 molecules and NaCl counter-ions (to neutralize the charges). A distance of 6Å was setted between the atoms with the higher x, y or z coordinates and the edge of the box. As a result of this selection, a box of 90 A \times 80 A \times 75 A was built. Periodic conditions and a cut-off of 10 A was used. Taking in consideration the size of the box, this distance was considered enough large to take into account all electrostatic relevant interactions. After that, a molecular dynamic was performed using the mentioned conditions to simulate the solvent effects. A SVL (Scientific Vector Language) function inside MOE (oSetCollection) was used to assign atoms to the named set MD_INTERACTION to evaluate interaction energy averages between each ligand and all the environment (protein, solvent and ions). A time step of 2 fs was set, light bonds were constrained, water molecules was considered as rigid bodies. A microcanonical ensemble NVT was used and for solving the equations of motion, the NPA (Nosé-Poincaré-Andersen) method was used for solving the equations of motion.

Simulations were split into "heating," "equilibration" and "sampling" separate output database files (trajectories). They contain: the time t of the sample (in picoseconds); the value at time t of the full (extended system) Hamiltonian; the potential energy U(r) of the atomic system at time t in kcal/mol; the kinetic energy K(v) of the atoms in kcal/mol; the instantaneous temperature T (in Kelvin) of the atoms, the instantaneous pressure P (in kPa), of the system and the instantaneous volume V in cubic angstroms. Other rows in the output database were the position vector rt, the velocity

vector v, a vector xpt of coordinates and momenta (for internal use) of the extended Hamiltonian and the dimensions of the periodic box.

After the molecular dynamics equilibration of the CCD4a(b)-apocarotenoid model, the structure was considered adequate to be used in docking studies with the potential carotenoids substrates.

The coordinates of all built models will be kindly available by request in pdb or mol2 formats.

2.3. Docking procedure

2.3.1. The ligand to be docked: carotenoids database building

The ligands' optimized geometries (a- and B-carotene, lutein, zeaxanthin, β-cryptoxanthin and lycopene) were built employing the MMFF94x force field (Halgren, 1999a, 1999b). In previous research, the conformational space of lycopene, phytoene and phytofluene (Meléndez-Martínez et al., 2014) was studied. In solution, those carotenoids shown to have am ample conformational space, trending to a folding according their rotational single bonds freedom. Even if the all-trans configurations were one of less energetic conformations, the energetic barriers allowed deducing that the force needed to ensemble an alltrans configuration inside the binding pocket of oxygenases should be not very high. All-trans configurations of the carotenoids here studied (and the RR stereoisomer in the case of zeaxanthin) were modeled by means MOE program and the conformational maps for the carotenoids were built following the same strategy as for lycopenes, and the and offered as putative conformations to be docked.

2.3.2. Docking validation

A docking procedure was performed in order to find the better poses of apocarotenoid 3ON in the active site of 2BIW structure. As a first step, this method requires that a "receptor," a "binding site" and a "ligand" are defined. As a "receptor" all 2BIW atoms were selected, the "ligand" was the 3ON three-dimensional structure and as "Site," all protein atoms inside a 4.5 Å sphere around the ligand were set. As a second step, all available scoring and placement functions were tested (Alpha PMI, Alpha Triangle, Proxy Triangle and Triangle Matcher) in order to determine the more adequate placement method for this kind of receptor.

The complete docking process involved the placement (Proxy Triangle), the scoring (ASE), further refinement using the force field (Amber99) and finally, a re-scoring using London ΔG .

The enthalpy contribution (as the main component of the free energy) was calculated using the ASE scoring function. ASE equation is a sum of gaussian functions with a general formula: $R1 \cdot R2 \cdot \exp(-0.5d^2)$ for all the "ligand atom-receptor" pairs. R1 is the corresponding radii of the ligand atoms and R2 is the receptor atom radii or -1.85 Å for alpha spheres. *d* is the pair distance in Å. The proportionality constant is by default equal to 0.035 kcal/mol. The geometrical placement priority of ASE score aiming to fit very flexible ligands in the free space of the receptor demonstrated being the best starting scoring function.

During the placement, the ligand bonds were allowed to rotate freely, and then, the 30 best poses were retained.

London dG was set as the re-scoring method, which estimates the value of the ligand-receptor free energy of binding. The equation involve next terms:

$$\Delta G = c + E_{flex} + \sum_{h-bonds} C_{HB} f_{HB} + \sum_{h-bonds} C_{HB} f_{HB} + \sum_{m-lig} C_{M} f_{M} + \sum_{atoms \ i} \Delta D_{i}$$

where:

- c is the average of rotational and translational entropy
- E_{flex} is the lost energy due to flexibility of the ligand
- *f_{HB}* represents the deviation of the hydrogen bonds ideality taking a value in a range of 0–1
- C_{HB} is the energy of an ideal hydrogen bond
- *f_M* represents the deviation of the metal bond ideality taking a value in a range of 0–1
- c_M is the energy of an ideal metal bond
- D_i is the energy due to the desolvation of atom i.

The difference in desolvation energies is calculated according to the next equation:

$$\Delta D_i = c_i R_i^3 \left\{ \iiint_{u \notin A \cup B} |u|^{-6} du - \iiint_{u \notin B} |u|^{-6} du \right\}$$

where:

- A and B are the receptor and/or ligand volumes with atom i belonging to volume B
- R_i is the solvation radius of atom i (using OPLS-AA van der Waals sigma parameter plus 0.5 Å)
- c_i is the desolvation coefficient of atom i.

The coefficients { $c_{, C_{HB}}, c_{M}, c_{i}$ } are empirical values obtained from ~400 X-ray crystal structures of protein-ligand complexes with known experimental pKi data. The space integrals are estimated using the numerical Generalized Born integral method.

Referring to the placement method, a collection of poses was generated from the pool of ligand conformations using Proxy Triangle method. This method was developed to manage ligands with a large number of conformers, like is the case of carotenoids. Conformers are placed in the binding site prior fit the possible rotation of the bonds, thus saving computational time.

2.3.3 Docking of CCD4s and carotenoids

The docking procedure described in Section 2.3.2 was applied to β -carotene, β -cryptoxanthin and zeaxanthin in the CCD4a model. Additionally, the option "induced fit" of receptor was set in order to simulate flexibility in the active site residues.

The docking studies were done after a 2.0 ns MD of the initially optimized structures with the apocarotenoid substrate, in order to preserve the active site ready to allocate new substrates. The major modeling work was done after the docking studies, were the best pose(s) were submitted to longer MD simulation (10-20 ns, depending de case). Finally giving rise to stabilized protein-substrate complexes as it is presented in the following section.

2.3. MD of docked complexes

2.3.1. CCD4a complexes molecular dynamics

A series of 10 ns MD, in the same condition as in Section 2.2.2, were performed with the docked CCD4a- β -carotene complex in order to optimize it and collect multiple samples.

After that, 10 ns MD were run with the complexes CCD4azeaxanthin and CCD4a- β -cryptoxanthin derived of docking calculations.

2.3.2. CCD4b complexes modeling and molecular dynamics

Given that CCD4a and CCD4b1 are more related filogenetically each other than with *Synechocystis* sp. native apocarotenoid cleavage oxygenase (2BIW) and considering that the percentage of similarity between CCD4a and b is of 62.7%, molecular models of complexes were built from the CCD4a- β -carotene equilibrated complexes, replacing the original protein structure by that of CCD4b and adding OH moieties or modifying a double bond in other ligands. In the case of lycopene (aliphatic), the ligand was placed by the best docked pose.

Water boxes were built having the final models about 12,000 water molecules and 9 Na^+ ions. After that, a molecular dynamics simulation was made in three steps: (1) heating from 0 to 300 K in 100 ps; (2) equilibration: 900 ps at T = 300 K and (3) production: 10,000 ps at T = 300 K. The Nosé-Poincaré-Andersen algorithm was used as the trajectory generator together with the NVT ensemble. For all trajectories, potential energy and temperature were registered to ensure equilibrium that was reached near the 900 ps. After EM refinement and equilibration by MD, the other ligands were modeled using the β -carotene structure as reference.

A set of 10 ns molecular dynamics of the complexes of CCD4b with zeaxanthin, β -cryptoxanthin, β -carotene, α -carotene, lycopene and lutein were obtained. Complementarily, as a validation of the simulation protocols, the molecular dynamic of 2BIW-apocarotenoid was done (data not shown). An extension in the CCD4b with zeaxanthin to 20 ns was made.

The MD calculations of the complexes were done at least with a replicate using as starting point a random structure between 2 and 5 ns of the former MD.

2.4. MD analysis

2.4.1. Ligand interactions

In MOE the ligand interactions can be analyzed through a 2D plot, this facility allows explore the active site (in a schematic form) with the 2D chemical structure of the ligand. When the ligand is selected the plot summarizes all the relevant interactions between the atoms of the ligand and the receptor, solvent molecules or metal bonds. There are listed: hydrogen bonds, electrostatic (acidic or basic), hydrophobic and aromatic interactions. When the 3D distance of the

Table 1. Percentages of identity (under diagonal)/similarity (over diagonal) for all protein sequences CCD4a, CCD4b and CCD4c studied compared with the template 2BIW.

Identity/similarity	CCD4a	CCD4b	2BIW(A)		
CCD4a	100.0	62.7	37.1		
CCD4b	48.8	100.0	38.9		
2BIW(A)	22.9	23.1	100.0		

The BLOSUM62 substitution matrix, with gap start =7 and a gap extend =1, was used to perform the initial alignment of CCD4a, CCD4b over 2BIW sequence as the template. The process ran up using an iteration limit =100 and a failure limit =10. Then, the final refinement of the alignment was done using a gap start =1 and a gap extend =0.1. The values in italic over the diagonal correspond to Similarity, and the values below the diagonal correspond to Identity.



Figure 2. Ribbon draws of main chain of the modelled CCD4b protein before and after 10 ns molecular dynamics simulation. Grey fill space draying: apocarotenoid co-crystallized in the 2BIW structure. The oxygen atoms in the apocarotenoid are red colored. The Fe(II) atom of catalytic center is shown as a cyan ball and some details of surrounding are given in sticks.

bonding interaction is lower than cut-off distance of the program its value is displayed in the plot.

2.4.2. Interaction energy measurements

The potential energy U(r) was expressed in partitions Ua, Ub and Uab (all in kcal/mol). Ua corresponds to the potential energy of the receptor and Ub is the potential energy of the molecular fragment selected (the whole ligand for instance). Thus, Uab is the potential energy of the interactions between the atoms related to Ua and Ub.

3. Results and discussion

3.1. Sequence alignment and selection of the best template

After alignment, the sequence % of identity and similarity between 2BIW and CCD4s was measured and is summarized



Figure 3. Ribbon draws of main chain of the modelled CCD4s proteins (red, yellow and green) overlapped with the template 2BIW (pink). Fill space drawings in blue are the Fe++ atom in the catalytic center and the apocarotenoid co-crystallized in the 2BIW structure.

in the Table 1. In the case of CCD4a, the identity with respect to 2BIW is of 22.9% and the similarity is of 37.1%, and the corresponding values for CCD4b are 23.1 and 38.9%, respectively. Despite that these values are relatively low, there is expected good conservation in the secondary and tertiary structure in the CCD superfamily, therefore, it is acceptable to use 2BIW as the homology modeling template. Additionally, considering the high number of freedom degrees in the carotenoids' structure, it is of crucial importance to have a crystallographic reference of the ligand into the binding site. 2BIW structure has this requirement, thus, was the one selected as template.

3.2. Molecular modeling

The homology model of CCD4b, including the apocarotenoid ligand and secondary structure, as well of the coordination sphere of Fe(II) in the active site, are highlighted in Figure 2. The superposition of the four initial CCD4 models (a, b, c and d) over the template 2BIW is shown in the Figure 3. The modeled proteins can be described as a seven bladed β-propeller with seven β -sheets arranged around a central axis, conforming a toroidal structure. Four histidine residues highly conserved are in the surrounding of Fe(II) cation. The tertiary structure contains a series of antiparallel strands that are connected by long loops. The beta strands form a spiralized cylinder as the rigid part of the structure, while the loops together with some short a-helixes conform the flexible part of the structure. The active site is located between these two substructures and the four His coordinated Fe(II) is contained in the beta strands substructure. The Fe(II)



Figure 4. Fe(II) (cyan ball) coordination sphere. Water molecules are annotated with a "W." (a) Coordination sphere of template 2BIW. (b) Coordination sphere of CCD4a model.

Table 2. Scores estimating the interaction energies resulting from the molecular docking procedure for zeaxanthin, β -cryptoxanthin and β -carotene in CCD4a active site.

CCD4a poses	Zeaxanthin	β-Cryptoxanthin	β-Carotene
1	-20.35	-17.55	-16,11
2	-19.67	-17.52	-15,82
3	-19.39	-17.41	-15,63
4	-19.18	-17.33	-15,52
5	-18.84	-17.15	-15,47

coordination spheres of the modeled structures were analyzed to confirm the adequate disposition and in the Figure 4 are shown those of 2BIW and CCD4a.

3.3. Docking and MD of the carotenoids database with CCD4a

The best five docking scores are shown in the (Table 2). All the values shown in this table are in a range of (15–21) units, which is reasonable for so similar ligands. Figure 1SM was included in Supplementary Material, in which an overlay of all docked structures is presented. Despite of it, the score equation is enough sensitive to detect the low differences in polarity and H-bond capability of the three ligands.

The main outcome of this section of the work is that the score ranking can be associated with the hydroxyl number placed in the cyclohexenyl ring of this ligands. Then, the scores and the number of hydroxyl moiety in the molecules could be ordered in the following way: zeaxanthin > β -cryptoxanthin > β -carotene. This result confirms previous estimations related with the importance of the H-bond pattern between the carotenoids and the oxidases (Rodrigo, Alquézar, Alós, Medina, et al., 2013).

As mentioned above in Section 2.4.2, a more accurate form to estimate the interaction energies between ligands and receptor is to calculate the potential energy of both fragments, and then, subtract them of the potential energy of the complex After equilibration time of 2 ns of MD, the average Uab were calculated (2–10 ns) and used as a estimation of interaction energies, the values for CCD4b are presented in Table 3 and 5 (CCD4b and CCD4a comparison) (Vega-Teijido, Paulino Zunini, López, & Rodrigo, 2015). In Table 3 are shown the Uab values for six substrates which were experimentally tested (Rodrigo, Alquézar, Alós, Medina, et al., 2013). The Uab values reflect the H-bond capability of the ligands, being correlative to the number of OH moieties. Lutein and zeaxanthin $> \beta$ -cryptoxanthin $> \alpha(\beta)$ -carotene and lycopene.

The distances between Fe(II) and the C atoms of the ligands show that the position of the two double bonds reported as susceptible of cleaving, 7' = 8' and 9' = 10', are adequate. The $\log P_{o/w}$ values are shown as a reference of the hydrophilic-hydrophobic profile of the ligands.

In Table 5 (last right column) are shown the Uab values for three ligands studied in both CCD4 enzymes, aiming a comparison between them. In both cases, the difference in the Uab values between zeaxanthin and cryptoxathin is *ca.* 16 kcal/mol and between zeaxanthin and β -carotene is *ca.* 40 kcal/mol. In the case of CCD4a the distances between Fe(II) and C atoms (Table 4) are slightly longer than CCD4b, but still in an adequate distance for cleaving.

The 2D plots of the MD complexes of CCD4a with β -carotene, β -cryptoxanthin and zeaxanthin are shown in Figure 5 where can be observed the relevance of the hydrophobic interaction with the binding site, which is depicted as blue clouds. The H-bonds formed are with the water molecules inside de binding site, playing a role of bridge with the enzyme residues.

To give quantitative evidence of these observations, the root mean squares deviation (RMSD) and the hydrogen bonding evolution were analyzed for two selected examples: a duplicated trajectory of 10 ns each of the CCD4a-zeaxanthine complex, and a 20 ns trajectory of CCD4b1- β -carotene complex.

For one hand, the RMSD was plotted and averaged (Figures 2SM and 3SM, Supplementary Material) resulting 0.286 A and 0.300 A for both 10 ns duplicated trajectories for CCD4a-zeaxanthine complex. In the case of CCD4b1- β -carotene complex the average RMSD was 0.380 A. This picture evidenced stable equilibrium situations.

On the other hand, we registered the evolution of hydrogen bonds (Hbond) pattern through the time and evaluated the average: 176 and 178 Hbonds for the CCD4a-zeaxanthine complex 10 ns simulations and 154 for the CCD4b1- β -carotene complex 20 ns simulation. This is another evidence that the interaction of xanthophylls is more intense than in the case of carotenes.

Table 3. Distances (Å) between Fe(II) and the atoms number 15, 15', 10', 9', 8' and 7' in the corresponding studied carotenoid (annotated in the first column).

CCD4b	15	15′	10′	9′	8′	7′	logP _{o/w}	Uab mean 2–10 ns (kcal/mol)	Bioactivity in CCD4b
LUT	11.70	12.48	7.05	6.19	5.70	5.82	8.77	-98.12	Substrate*
ZEAX	10.90	9.68	5.51	5.38	5.97	6.13	8.57	-97.45	Substrate
B-CRYPTO	11.68	11.16	5.24	5.81	5.28	6.22	8.65	-81.05	Substrate
LYC	11.87	10.75	4.84	4.49	5.27	5.47	17.65	-60.81	Low activity
AC	12.84	11.69	7.45	8.17	7.65	8.18	17.28	-63.29	Substrate*
BC	12.91	11.92	8.02	8.53	7.68	8.19	17.44	-60.51	Substrate*

LogPo/w and interaction energies (Uab) are provided in the seventh and eight columns. Bioactivity data (Rodrigo, Alquézar, Alós, Medina, et al., 2013) were indicated in the last column.

LUT = Lutein; ZEAX = zeaxanthin; B-CRYPTO = β -cryptoxanthin; LYC = lycopene; AC = α -carotene; BC = β -carotene. Bioactivity data (Rodrigo, Alquézar, Alós, Medina, et al., 2013) were indicated in the last column. *Present only in immature (green) citrus fruit.

Table 4. Distances between the methalic center Fe(II) in the CCD4a protein and the atoms 15, 15', 10', 9', 8' and 7' of zeaxanthin (ZEAX), measured at the 10 ns seconds of molecular dynamics simulation for the complexes with b-cryptoxanthin (BCRYPTO) and b-carotene (BC).

CCD4a	Dist (Å) Fe-15	Dist (Å) Fe-15'	Dist (Å) Fe-10'	Dist (Å) Fe-9'	Dist (Å) Fe-8'	Dist (Å) Fe-7'	LogPo/w
ZEAX	11.32	10.93	8.11	7.23	7.95	7.69	8.57
BCRYPTO	9.80	8.02	7.85	7.07	7.70	7.48	8.65
BC	12.30	11.62	8.14	6.90	6.67	7.75	17.65

Last column: LogPo/w were indicated.

Figures 4SM and 5SM of Supplementry Material shown the HBond evolution and are another evidence that ligands as β -carotene, with lower polarity and high hydrophobicity and displaying less interaction energy with CCD4a, have less average amount (152) of hydrogen bonds than the more polar and less hydrophobic zeaxanthin (the more interactive xanthophyll) with CCD4b1. Finally, the averaged amount of zeaxanthin Hbonds (175) is in correspondence with the presence of β -cryptoxanthin and zeaxanthin clusters of H-bonds formed with the water molecules inside de binding site, and the absence of such pattern in the case of β -carotene, observed in Figure 5.

3.4. The coordination sphere and the putative influence in reactivity of CCD4 enzymes

All trajectories were analyzed in the last point (10 ns) to compare the stabilized structure of the complexes. The distances between the double bonds 15 = 15', 10' = 9' and 8' = 7'were measured and are presented in Tables 3 and 4 for all complexes with CCD4b and CCD4a, respectively. In Table 5 and Figure 4 are shown all residues making contact in the coordination Fe(II) sphere for three complexes of each CCD4 and for the crystallographic structure 2BIW. As can be seen Fe(II) remains in contact with the four His residues and two aligned water molecule, which is in the direction of the ligand. In the case of CCD4 family, the corresponding Thr residue is farther away than in 2BIW and an Asp residue is closer. Sometimes during the MD, this Asp are in contact with Fe(II) canceling part of the positive charge.

In a real situation, it is necessary to give more information about the oxygen mediated reaction itself. A new Figure incorporated as Supplementary Material (Figure 6SM) was introduced to explain the atomistic details involved in the carotenoids oxidation. The oxidation state of the Fe is +2and that electronic state it was maintained through all molecular dynamics. In the crystal and in the corresponding models, a water molecule take the place of the oxygen. In a real situacion, the Fe+2 activates the oxygen to make feasible than the attack could be done over some of the double bonds. As it could be visualized in Figure 6, that is an overlapping of all docked carotenoids, a distance between the Fe(II) and this replacing water molecule is of 2.54 Å and between the water molecule and the double bond is of 2. 63 Å. Then, we could conclude that there is room to an oxygen replacing this water molecule to attack some double bond position involved in the carotenoid oxidation.

Analysis of the observed flexibility in the coordination of Fe(II) during MD simulations and the use of alternative protonation states of residues could give relevant information in the reactivity of these enzymes. This part of our studies is by the time under development with QM/MM methods and further MD calculations, which will be reported in future publications.

4. Conclusions

Here we present validated molecular models for CCD4a and CCD4b with a series of six different carotenoid ligands. Estimations of the interaction energies are presented by docking and MD calculations. Both molecular models, initially obtained by homology modeling from the 2BIW crystal, are energetically stable after 2 ns of molecular dynamics simulations. Thus, were considered validated and adequate for sampling. The models built confirm the stability of the typical tertiary structure of CCD4 homologs, and can be divided in two parts: one flexible conformed by long loops and short alpha-helices and the other rigid conformed by beta-strands containing a Fe(II) coordinated with four histidine residues.

In the case of the CCD4 models four expected His residues are in the coordination sphere as it is expected. Additionally, it was found an Asp residue is in the second sphere of coordination. However, during the MD calculations, it was observed that residues and water molecules alternate



Figure 5. 2D plots of the MD complexes of CCD4a with β -carotene, β -cryptoxanthin and zeaxanthin obtained with the ligand interaction utility inside MOE program. The residues making contacts were shown as balls in colors corresponding to the hydrophobicity/hydrophobicity (pink = polar residues; green = hydrophobic). Blue clouds represent the solvent accessibility and the dotted line represents the binding pocket. In the case of xantophylls β -cryptoxanthin and zeaxanthin, clusters of H-bonds formed with the water molecules inside de binding site, playing a role of bridge with the enzyme residues, are depicted.

Table 5. Contacting amino acidic residues (CCD4a numeration) detected in a 4.5 A sphere around the metallic catalytic center Fe(II) are annotated in the second to tenth columns.

	His218	His267	Asp268	Met329	His333	Phe332	Glu389	His513	Phe504	HOH	Uab
CCD4b-ZEAX	×	×	×		×			×		2	-97.45
CCD4b-BCRYPTO	×	×	×		×			×		2	-81.05
CCD4b-BC	×	×	×		×			×		2	-60.81
CCD4a-ZEAX	×	×	×		×			×		2	-101.52
CCD4a-BCRYPTO	×	×	×		×			×		2	-76.98
CCD4a-BC	×	×	×		×			×		2	-59.04
2BIW-3ON	×	×			×			×		1	

Number of water molecules in the surrounded are indicated in the eleventh column. For comparison and discussion, Uab interaction energies were annotated in the last column. Zeaxanthin (ZEAX), b-cryptoxanthin (BCRYPTO), b-carotene (BC).



Figure 6. Overlapped ball and stick display of docked carotenoids in CCD4a. Magenta: zeaxanthin; Green: b-carotene; Orange: beta cryptoxanthin. For the best docked zeaxanthin, a distance between the water molecule and the Fe(II) is shown (2.54A).

in the closer positions to Fe(II), suggesting a flexible environment.

The docking results of carotenoids in the CCD4a active site suggest that the binding is more favorable for zeaxanthin than for β -cryptoxanthin and that β -carotene is the less favored. CCD4a MD calculations show that the interaction energies averaged over the trajectories are in the same binding order of zeaxanthin > β -cryptoxanthin > β -carotene.

In the case of CCD4b, it was found an agreement with experimental results for all carotenoids with β -ionone ring, the binding order was: lutein > zeaxanthin > β -cryptoxanthin > α -carotene > lycopene > β -carotene. Therefore, these results support the better behavior of xanthophylls compared with α - and β -carotene, being the polarizability of hydroxylated β -ionone essential for binding to the enzyme.

The initial all-trans configurations were conserved during the MD simulations for most ligands. However, lycopene (the aliphatic carotene) suffered an isomerization from trans to cis of two conjugated double bonds, result that deserves more research. In previous works, the geometrical and conformational space was studied for lycopene and zeaxanthin (Meléndez-Martínez et al., 2014 and unpublished results). An in silico isomerization is an unexpected event during a MD simulation given the relative high energy required for a methyl-substituted double bond rotation and the small free space in the active site when is occupied by the ligand. However, in vivo isomerization could happen, through the catalytic activity of these enzymes, when they promote the changes in the ligand bonds. The distances between the catalytic Fe(II) and the ligands are adequate to a cleavage of the 7' = 8' double bond. In all cases, it can be discarded the possibility of a symmetric 15 = 15' cleavage, which is in agreement with the experimental results. Although, a cleavage of the double bond 9'=10' cannot be ruled out due that the distances are close to those of 7' = 8'.

Acknowledgements

Authors thanks the research financial support of Comisión Sectorial de Investigación Científica (CSIC), Proyecto de Desarrollo de Ciencias Básicas (PEDECIBA) and IBERCAROT CYTED Network. Research on MJR laboratory was supported by MINECO (Spain) grant AGL2015-70218.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Adami, M., De Franceschi, P., Brandi, F., Liverani, A., Giovannini, D., Rosati, C., ... Tartarini, S. (2013). Identifying a carotenoid cleavage dioxygenase (CCD4) gene controlling yellow/white fruit flesh color of peach. Plant Molecular Biology Reporter, 31(5), 1166–1175.
- Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., ... Rosati, C. (2011). Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. BMC Plant Biology, 11(1), 24. doi:10.1186/1471-2229-11-24
- Bruno, M., Beyer, P., & Al-Babili, S. (2015). The potato carotenoid cleavage dioxygenase 4 catalyzes a single cleavage of β-ionone ring-containing carotenes and non-epoxidated xanthophylls. Archives of Biochemistry and Biophysics, 572, 126–133.
- Campbell, R., Ducreux, L. J., Morris, W. L., Morris, J. A., Suttle, J. C., Ramsay, G., ... Taylor, M. A. (2010). The metabolic and developmental roles of carotenoid cleavage dioxygenase 4 from potato. Plant Physiology, 154(2), 656–664.
- Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., ... Morgante, M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. The Plant Journal, 76, 175–187.

FAO. (2016). Retrieved from http://www.fao.org/3/a-i8092e.pdf

- Fukamatsu, Y., Tamura, T., Hihara, S., & Oda, K. (2013). Mutations in the CCD4 carotenoid cleavage dioxygenase gene of yellow-flesh peaches. Bioscience, Biotechnology, and Biochemistry, 77(12), 2514–2516.
- Gasteiger, J., & Marsili, M. (1980). Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. Tetrahedron, 36(22), 3219–3228.
- Gonzalez-Jorge, S., Ha, S.-H., Magallanes-Lundback, M., Gilliland, L. U., Zhou, A., Lipka, A. E., ... Dellapenna, D. (2013). Carotenoid cleavage dioxygenase4 is a negative regulator of β-carotene content in Arabidopsis seeds. The Plant Cell, 25(12), 4812–4826. doi:10.1105/ tpc.113.119677
- Huang, F. C., Molnár, P., & Schwab, W. (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. Journal of Experimental Botany, 60(11), 3011–3022.
- Halgren, T.A. (1999a) MMFF VI. MMFF94s option for energy minimization studies. Journal of Computational Chemistry, 20, 720–729.
- Halgren, T.A. (1999b) MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. Journal of Computational Chemistry, 20, 730–748.
- Kato, M. (2012). Mechanism of carotenoid accumulation in citrus fruit. Journal of the Japanese Society for Horticultural Science, 81(3), 219–233.
- Kloer, D. P., Ruch, S., Al-Babili, S., Beyer, P., & Schulz, G. E. (2005). The structure of a retinal-forming carotenoid oxygenase. Science, 308(5719), 267–269. doi:10.1126/science.1108965
- Lätari, K., Wüst, F., Hübner, M., Schaub, P., Beisel, K. G., Kim, G., ... Welsch, R. (2015). Tissue-specific apocarotenoid glycosylation contributes to carotenoid homeostasis in Arabidopsis leaves. Plant Physiology, 168(4), 1550–1562. doi:10.1104/pp.15.00243
- Rodrigo, M. J., Alquézar, B., Alós, E., Lado, J., & Zacarías, L. (2013). Biochemical bases and molecular regulation of pigmentation in the peel of Citrus fruit. Scientia Horticulturae, 163, 46–62.
- Rodrigo, M. J., Alquézar, B., Alós, E., Medina, V., Carmona, L., Bruno, M., ... Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments.

Journal of Experimental Botany, 64(14), 4461–4478. doi:10.1093/jxb/ ert260

- Ma, J., Li, J., Zhao, J., Zhou, H., Ren, F., Wang, L., ... Han, Y. (2014). Inactivation of a gene encoding carotenoid cleavage dioxygenase (CCD4) leads to carotenoid-based yellow coloration of fruit flesh and leaf midvein in peach. Plant Molecular Biology Reporter, 32(1), 246–257.
- Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., ... Kato, M. (2013). Enzymatic formation of β -citraurin from β -cryptoxanthin and zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. Plant Physiology, 163(2), 682–695. doi:10.1104/ pp.113.223297
- McAndrew, R. P., Sathitsuksanoh, N., Mbughuni, M. M., Heins, R. A., Pereira, J. H., George, A., ... Adams, P. D. (2016). Structure and mechanism of NOV1, a resveratrol-cleaving dioxygenase. Proceedings of the National Academy of Sciences, 113(50), 14324–14329. doi:10.1073/pnas.1608917113
- Meléndez-Martínez, A. J., Paulino, M., Stinco, C. M., Mapelli-Brahm, P., & Wang, X. D. (2014). Study of the time-course of cis/trans (Z/E) isomerization of lycopene, phytoene, and phytofluene from tomato. Journal of Agricultural and Food Chemistry, 62(51), 12399–12406. doi:10.1021/ jf5041965
- MOE The Molecular Operating Environment, Version (2014). Chemical computing group inc. Montreal, Quebec, Canada. http://www.chem-comp.com.
- Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., & Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiology, 142(3), 1193–1201.
- Rubio, A., Rambla, J. L., Santaella, M., Gómez, M. D., Orzaez, D., Granell, A., Gómez-Gómez, L. (2008). Cytosolic and Plastoglobule-targeted

Carotenoid Dioxygenases from Crocus sativus Are Both Involved in Ionone Release. The Journal of Biological Chemistry, 283(36), 24816–24825.

- Silvarrey, M. C., Echeverría, S., Costábile, A., Castillo, E., Paulino, M., & Esteves, A. (2016). Identification of novel CAP superfamily protein members of *Echinococcus granulosus protoscoleces*. Acta Tropica, 158, 59–67.
- Sui, X., Golczak, M., Zhang, J., Kleinberg, K. A., von Lintig, J., Palczewski, K., & Kiser, P. D. (2015). Utilization of dioxygen by carotenoid cleavage oxygenases. Journal of Biological Chemistry, 290(51), 30212–30223. doi:10.1074/jbc.M115.696799
- Vega-Teijido, M., Paulino Zunini, M., López, C., & Rodrigo, M. J. (2015). A theoretical study of CCD4a dioxygenase of citrus, a cleavage enzyme of carotenoids in plants. Revista Processos Químicos, XVIII Simpósio Brasileiro De Química Teórica, AG, 38, 163–165.
- Wang, J., Cieplak, P., & Kollman, P. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. Journal of Computational Chemistry, 21(12), 1049–1074.
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. Journal of Computational Chemistry, 7(2), 230–252.
- Zheng, X., Xie, Z., Zhu, K., Xu, Q., Deng, X., & Pan, Z. (2015). Isolation and characterization of carotenoid cleavage dioxygenase 4 genes from different citrus species. Molecular Genetics and Genomics, 290(4), 1589–1603. doi:10.1007/s00438-015-1016-8
- Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., ... Liu, K. (2015). Disruption of a carotenoid cleavage dioxygenase 4 gene converts flower colour from white to yellow in *Brassica* species. The New Phytologist, 206(4), 1513–1526.





Graphical Abstract





Article



1

2

3

4

5

6 7

8

9

10

11

12

13

CITRUS CAROTENOIDS CLEAVAGE DIOXYGENASES TYPE CCD4 FAMILY AND THEIR INTERACTION WITH MEMBRANES

Jorge Cantero ^{1,2}, Fabio Polticelli ^{3,4,*} and Margot Paulino ^{1,*}

1	Bioinformatics Area, DETEMA Department, Faculty of Chemistry, UdelaR, General Flores 2124, 11600,
	Montevideo, Uruguay

- ² Research Department, Faculty of Agronomic Engineering, Universidad Nacional del Este, Ruta Internacional N°7 Km 17.5 CP:7420, Fax: (+595) 0644 20440, Ciudad del Este, Paraguay
- ³ Department of Sciences, Roma Tre University, 00146 Rome, Italy
- ⁴ National Institute of Nuclear Physics, Roma Tre Section, 00146 Rome, Italy
- * Correspondence: MP; margot @fq.edu.uy; Tel.: +598 99 2 71865; FP; <u>fabio.polticelli@uniroma3</u>; Tel.: +39 06 5733 6362

Abstract: Coloring is one of the most important characteristics in commercial flowers and fruits, 14 generally due to the accumulation of carotenoid pigments. Enzymes of the CCD4 family in citrus 15 intervene in the generation of β -citraurin, an apocarotenoid that gives the reddish-orange color in 16 mandarins. Citrus CCD4s enzymes could be capable of interacting with the thylakoid membrane 17 inside chloroplasts. However, to date this interaction has not been studied in detail. In the present 18 work we present three new complete models of the CCD4 family members (CCD4a, CCD4b and 19 CCD4c), modeled with a lipid membrane. To identify the preference for substrates, typical carote-20 noids were inserted in the active site of the receptors and the protein-ligand interaction energy was 21 evaluated. The results show a clear preference of CCD4s for xanthophylls over aliphatic carotenes. 22 Our findings indicate the ability to penetrate the membrane and maintain a stable interaction 23 through the N-terminal α -helical domain, spanning a contact surface of 2250 to 3250 Å². The orien-24 tation and depth of penetration at the membrane surface suggests that CCD4s have the ability to 25 extract carotenoids directly from the membrane through a tunnel consisting mainly of hydrophobic 26 residues that extends to the catalytic center of the enzyme. 27

Keywords: CCD4 enzymes; Citrus Clementina; protein-membrane interactions

28

1. Introduction

The carotenoids are isoprenoid pigments (terpens) abundant in the membranes of all 30 phototrophic and most of heterotrophic organisms. The empiric formula of β -carotene 31 (C₄₀H₅₆) was determined for the first time in 1907 by Willstatter and Mieg [1] and its chemical structure dilucidated by Paul Karrer in 1931. It was the first molecule declared as provitamin and this research deserved the Nobel Prize in Chemistry in 1937 [2]. The name 34 "carotenoids" derived from its major member, β -carotene [3], a molecule abundant in carrots (*Daucus carota*). 36

The abundance and diversity of carotenoids in plants is due to multiple functions 37 and properties related to their chemical structure. Most of the animals are not capable to 38 synthetize them and for this reason an equilibrated diet must include their intake [3]. Exceptionally, some aphides are able to synthetize carotenoids and it has been demonstrated 40 that this capacity is the result of a horizontal transfer of genes involved in the biosynthesis 41 of carotenoids from bacteria and fungi [4].

1.1 Carotenoids biosynthesis

Carotenoids are localized in the thylakoid membranes and their biosynthesis is car-44 ried out in chloroplasts [1]. The biosynthesis in higher plants is initiated with the covalent 45 binding of two geranyl diphosphate (GGPP) molecules, catalysed by phytoene synthase 46 (PSY – EC 2.5.1.32). After, phytoene is desaturated by phytoene desaturase (PDS – EC 47 1.3.99.31) to generate ζ -carotene. ζ -carotene, through the action of ζ -carotene desaturase 48(ZDS – EC 1.3.5.6) yields a lycopene molecule. Lycopene is then cyclized in two ways: on 49 one hand ε -lycopene cyclase (ε -LYC – EC 5.5.1.18) produces α -carotene and on the other 50 hand β -lycopene cyclase (β -LYC – EC 5.5.1.19) produces β -carotene. Both products can be 51 hydroxylated by specific enzymes and generate xanthophylls. α -cryptoxanthin is the 52 product of the activity of ε -carotene hydroxylase (ε -CHX – EC 1.14.14.158) while β -cryp-53 toxanthin is the product of the activity of β -carotene hydroxylase (β -CHX – EC 1.14.15.24). 54 α -cryptoxanthin can be further hydroxylated by β -CHX to give lutein and the same can 55 happen for β -cryptoxanthin to give zeaxanthin [5] (Figure 1a) 56

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Int. J. Mol. Sci.* 2021, 22, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date Accepted: date Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses /by/4.0/). 29

43


Figure 1. a) Carotenoids biosynthetic pathways b) Structural formulas and numbering of apocarot-59enoid 3-hydroxy-8'-apocarptenol (3ON) and α -carotene, β -carotene, lutein, lycopene, β -cryptoxan-60thin and zeaxanthin carotenoids.61

Zeaxanthin can be reversibly modified and generate epoxicarotenoids such as antheraxanthin and violaxanthin. The set of reactions is known as xanthophylls cycle and the products represent important protective factors against the oxidative stress resulting from the photosynthetic metabolism in plants [6].

All the carotenogenic genes in vascular plants are nuclear [6] and have as in the carotenoids cleavage enzymes genes (CCDs) a region that codes for a N-terminal chloroplast 67

57

71

74

75

85

86

87

88

89

branes through a hydrophobic region that penetrates in the membrane where their substrates are located [8]. The carotenoids accumulate in the chloroplasts of immature fruits, which in the sub-72 sequent steps of the maturation process lose their photosynthetic system and become 73

chromoplasts. The color of fruits is then due to the preexistent carotenoids [9].

targeting peptide (cpTP). Thus, biosynthesis and cleavage reactions occur in the chloro-

plast [6,7]. These enzymes are not transmembrane proteins but they interact with mem-

1.2 Carotenoids reactivity

The presence of conjugate double bonds makes carotenoids chromophores that ab-76 sorb light in the UV-visible spectrum [10]. The polyenic structure of carotenoids gives to 77 these compounds the capacity of being antioxidants by accepting electrons from reactive 78species and neutralizing free radicals [1]. 79

In mammals, the carotenoids are Vitamin A (Vit-A; retinal) and retinoic acid precur-80 sors, which play an important role in nutrition, vision and cellular differentiation pro-81 cesses [1,11]. In plants they are crucial for the synthesis of hormones such as abscisic acid 82 (ABA) [12], a molecule participating in the physiological stress responses of the strigolac-83 tones and responsible for the plants lateral ramification [13]. 84

Some carotenoids could have a role as photosystems antennas and in different types of stress [13]. In fruits and flowers, the presence of carotenoids gives color and smell, sensorial properties appealing to pollinators and organisms that help in the seeds dispersion [14].

1.3 Carotenoids cleavage dioxygenases (CCDs)

The CCDs constitute a big family of non heme Fe⁺² dependent enzymes participating 90 in the oxidation of carotenoids and apocarotenoids, widely distributed in all living organ-91 isms[15]. These enzymes are able to catalyze an oxidative cleavage at the level of carote-92 noids double bonds[16]. 93

The CCDs are usually very specific both in terms of cleavage site and carotenoid 94 specificity. All of them have a highly conserved site formed by four histidines that coor-95 dinate a Fe⁺² atom. The CCD4 family is characterized by a high level of structural conser-96 vation [17]. A vast majority of CCD4s is associated to the thylakoid membranes inside the 97 chloroplasts. The function of the cleavage products is not yet entirely dilucidated [18]. The 98 first CCD whose X-ray three-dimensional structure was solved was the Synechocystis sp 99 apocarotenoids cleavage oxygenase (ACO – EC 1.13.11.75) expressed in Escherichia coli 100 (PDB id: 2BIW). This enzyme cuts the apocarotenoid in the position 15=15' [19]. 101

A very diverse set of CCDs exists in plants [18]. The CCDs are classified by their 102 capacity to produce ABA, a hormone implicated in the oxidative stress response and in 103 the seeds dormancy [20]. The first characterized subfamily was the one involved in the 104ABA biosynthesis, i.e. the 9-cis-epoxicarotenoids dioxygenases (NCEDs) [21,22]. Members 105 of this family oxidize selectively in the position C11-C12 9-cis-violaxantine and 9-cis-ne-106 oxantine, giving xantoxine as product, an ABA precursor[23]. The functions of other sub-107 families (CCD1, 2, 4, 7 y 8) are not yet fully known[18,24]. 108

1.4 CCD4 subfamily

The CCD4 subfamily is distributed in all flowering plants [25]. The functions of sub-110 family members are directly related to the color of fruits and flowers and their scents [26]. 111 They are present in leaves, flowers, roots, and stems [27]. A variable number of genes 112 coding for the CCD4 subfamily members is present in different species, with most of them 113 having at least two CCD4 genes [25,28]. Some CCD4 genes are specifically expressed only 114

121

122

123

stress [27].116Similarly, to other members of the CCD family, members of this subfamily initially117possess a chloroplast targeting peptide (cpTP) in the N-terminal region [27]. They are118housed inside the chloroplasts together with other enzymes of the carotenoids biosynthe-119sis pathways thus giving them easy access to their substrates [22]. Inside the chloroplasts120

in certain organs (Rodrigo et al., 2013) and their expression can be triggered by abiotic

they are among the most abundant proteins, making evident their importance in the ca-

1.5 Functional characterization of CCD4s

rotenoids synthesis routes in plants [29].

Most of CCD4s are capable of cleaving carotenoids at the C9 = C10 and C9' = C10' 124 position producing C13 apocarotenoids. However, in citrus evidence on CCD4b indicates 125 the ability to cleave in one of the positions C7 = C8 or C7' = C8' [27]. 126

1.6 Characteristics of the CCD4 of citrus species

127

A total of 8 CCD genes have been currently identified in citrus: two NCEDs, one 128 CCD1 and five members of the CCD4 subfamily. The CCD4 subfamily members are 129 named a, b1, b2, c and d. 130

The largest number of CCD4 genes are found in citrus. The number of base pairs (bp) 131 that make up the coding region varies from species to species and ranges from 1200 to 132 1800 bp (Zheng et al., 2015) [30]. The vast majority of genes do not have introns. Sequence 133 analysis shows extensive allelic diversity, including SNPs and in-frame shift mutations 134 [30]. The length of the predicted sequence ranges from 400 to 600 amino acids. The degree 135 of amino acid sequence identity varies throughout the members of this subfamily, ranging 136 from 30% to 98% with typical values between 50% and 70% [18]. 137

In oranges and mandarins, CCD4b1 was characterized as responsible for the asymmetric cleavage of β -cryptoxanthin and zeaxanthin to generate β -citraurin (C30) whose accumulation produces the intense orange-reddish color in the pigmentation of mandarin hybrids of the Fortune variety [27].

CCD4a is expressed in all tissues, but it predominates in leaves and flowers, while CCD4b1 is expressed predominantly in the peel and to a lesser extent in flowers [27]. CCD4c is expressed only in flowers and has been proposed as ideal molecular marker for the study of the relationships between citrus species based on their allelic polymorphism [30]. The CCD4b2 and CCD4d are not expressed. These genes have lost important regions coding for conserved residues of the CCD family and evidence indicates that they are pseudogenes [27].

Given the biotechnological importance of the CCD4 dioxygenases and the absence of 149 enough data about their structure and mechanism of action, the present work was focused 150 on the CCD4s from Citrus clementina varieties with emphasis on those that are expressed 151 (CCD4a, CCD4b1 (hereinafter CCD4b) and CCD4c). To this aim in this work a combina-152 tion of different bioinformatics techniques was used to model and validate their three-153 dimensional structures, to build the complexes with putative carotenoid ligands, to eluci-154 date substrate selectivity. Furthermore, a detailed analysis of the residues that participate 155 in enzyme-substrate interactions, as well as in the interactions with biological membranes 156 was carried out using molecular simulation techniques. 157

2. Results

2.1. Molecular Modeling

The threading models of CCD4a, CCD4b and CCD4c are shown in Figure 2, superimposed onto the three-dimensional structure of VP14, a dioxygenase from corn that possesses the α -helix domains that characterize CCDs of plant origin. These are thought to 162

158

have a function in the interaction of the protein with thylakoid membranes, which is why 163 the model that will be proposed obeys to the hypothesis that such α -helical domains in-164teract with biological membranes in CCD4s as well. 165

166



Figure 2. Ribbon representation of the CCD4s models: CCD4a (green), CCD4b (blue), CCD4c (pink) and VP14 (red). α -helices 1 and 3 are indicated by arrows.

The models obtained with PHYRE 2 display the β -propeller structure characteristic 170 of CCDs. This structural domain is formed by seven β -sheets arranged around a central axis. At the center of this structure is located the catalytic center composed of a Fe^{+2} ion coordinated by four histidine residues, highly conserved in the members of this family. 173

The structural alignment between the models and the crystallographic structure of 174 VP14 show very small structural differences, with RMSD values ranging from 0.85 to 1.38 Å (Table 1). 176

Table 1. RMSD values (lower diagonal) and percentage of sequence similarity (upper diagonal, col-178 ored in violet) calculated for the CCD4a, CCD4b and CCD4c models as compared to the crystal 179 structure of VP14. The sequence alignment was obtained with the BLOSUM62 substitution matrix 180 (gap open penalty = 7, gap extension penalty = 1) using the MOE-Align program (Molecular Oper-181 ating Environment [31]) 182

RMSD / Similarity	CCD4a	CCD4b	CCD4c	VP14
CCD4a		62.70	72.00	51.90
CCD4b	1.14		64.30	52.60
CCD4c	1.38	0.85		49.60
VP14 (X-Ray)	0.95	0.69	0.85	

The main differences in the three-dimensional structure observed in the models are 185 found in the loops, which constitute the most mobile regions of the structure as can be 186 seen in Figure 3. 187

171 172

167

168 169

175

177



189

193

Figure 3. Ribbon representation of the models obtained for CCD4a, CCD4b, CCD4c and of the three-190dimensional structure of VP14 colored according to RMSD values in the range 0.0-1.5 Å. Low values191(green), medium values (yellow) and high values (red).192

2.2 Insertion of α -helix domains into the membrane

The α -helix domain of CCD4s is made up of two antiparallel α -helices, $\alpha 1$ and $\alpha 3$. 194 In the Figure 4, the alignment of the sequences of the $\alpha 1$ and $\alpha 3$ helicoidal regions for all 195 models is shown. It is interesting to note that in CCD4b, the helicoidal region starts at the 196 45th amino acid, while in CCD4a and CCD4c starts at the 86th and 78th amino acid, respectively. However, the tertiary structure of the helical domain is highly conserved in the 198 CCD4 subfamily members. 199

- 200 201
- 202
- 203
- 204
- 205

217

218

219 220

a)																							206
Receptor	Init	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	end	
CCD4a	86	А	Е	Р	Т	Ι	Р	Т	I	Ι	L	Ν	А	С	D	-	D	Ι	Ι	N	N	104	
CCD4b	45	Р	Ι	Q	S	L	М	G	Т	Ν	S	S	Y	Ν	Т	К	S	А	Р	S	L	64	
CCD4c	78	т	Т	Т	Ν	v	S	А	V	I	L	Ν	Т	L	Ν	К	-	L	Ι	к	Р	96	
																							207
b)																							208
		_	_		_		_	_	_														209
Receptor	init	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	en	d				210
CCD4a	218	Т	Α	S	Α	А	R	G	А	L	S	А	А	R	L	L	А	23	3				211
CCD4	100		v		м	17	0	C			c	-		Б			м	10	-				212
CCD4b	180	L	v		IVI	v	Q	C	v	А	5	1	A	к	v	L	M	19	5				213
CCD4c	210	L	А	F	L	Т	R	А	G	V	L	А	А	R	V	L	Т	22	5				214
																							215

Figure 4. Sequence alignment of the α 1 and α 3 helicoidal regions of CCD4a, b and c. Colors indicate the degree of polarity of the amino acids: red: negatively charged; blue positively charged; green: hydrophobic; violet: polar. *Init* and *end* indicate the initial and final position of the helical regions in the sequences.

The α 1 domain is formed by 19-20 amino acids and polar amino acids are predomi-221 nant. The α 3 domain is formed by 17 mainly hydrophobic amino acids that interact with 222 the hydrophobic region of the membrane. The polar and charged surface of the helices 223 interacts with the thylakoid membrane through polar interactions with the phospholipid 224 moieties characterized by a zwitterionic structure (as, for example, in the case of phospha-225 tidylcholine). To refine the analysis of the interactions between the proteins and the mem-226 brane, the time evolution of hydrogen bonds through the trajectory has been recorded and 227 plotted (Supplementary Material – Figure S1), and the number of hydrogen bonds aver-228 aged. The average resulted to be 9 for the CCD4a, 7 for CCd4b and 10 for CCD4c. Thus, 229 there is no a significant difference in this respect between the different proteins indicating 230 that the mechanism of the protein-membrane binding is similar. 231



Figure 5. Top) Ribbon representation of CCD4b model inserted in the membrane (dotted grey surface). Bottom) Molecular surface representation colored by hydrophobic (green), negatively charged (red) and positively charged (blue) areas.

Figure 5 shows two representations of the CCD4b protein model inserted in the 237 membrane following the protocol described in the Methods section. A ribbon 238 representation of the model inserted in the membrane allows to see the two $\alpha 1$ and $\alpha 3$ 239 helices (in red) penetrate in the membrane (whose plane is represented by the grey dotted 240 area). Protein solvent accessible surface colored by charge distribution is also shown from 241 two different viewpoints. 242

It can be seen from Figure 5 that the penetration of the protein into the thylakoid 243 membrane is such that the hydrophobic groups of the α -helices interact with the fatty 244 acids inside the membrane. Depending on the protein subtype, the initial phospholipid 245 membrane penetration was from 5.2 to 7.0 Å, with a predicted average $\Delta G_{transfer}$ value of -246 58.1 kJ/mol (Table 2). These values are consistent with those observed for VP14, with a reported penetration depth of 7 Å [22]. Thus, the proposed model indicates that CCD4s are putative peripheral membrane proteins.

A comparison between all the CCD4-membrane models gives evidence of a similar 250 placement inside the membrane. In Table 2 values of the angle, depth and energy of 251 transfer from the solvent to the membrane are reported. In Figure 6 all the CCD4-252 membrane complex models are represented. 253

232 233

236



Figure 6: Models of CCD4a (pink), CCD4b (orange) and CCD4c (cyan) in complex with a membrane model, represented by a plane of grey balls defined by the phospholipids phosphate groups. The red arrow indicates the Fe⁺² catalytic center (cyan sphere).

Table 2. Values of the angle of inclination of the proteins' first principal axis of inertia with respect260to the membrane plane, of the depth of the penetration in the membrane and of the free energy of261transfer from the solvent to the membrane ($\Delta G_{transfer} \ln kJ/mol$).262

Protein	Angle	Depth (Å)	$\Delta G_{transfer}$ (kJ/mol)
CCD4a	32° (±4°)	7.0 (±0.6)	-64.4
CCD4b	39° (±4°)	5.8 (±1.3)	-61.5
CCD4c	28° (±4°)	5.2 (±0.5)	-48.5
Average	33°	6.0	-58.1

It is evident from Figure 6 and from the values reported in Table 2 that all the CCDs are predicted to interact with the membrane in a similar manner.

The penetration of the hydrophobic region formed by the α -helices of the CCD4s 267 places the substrate channel in proximity of the thylakoid membrane, allowing to hypoth-268 esize that the enzyme can have a direct access to the membrane interior. In this regard, it 269 is important to note that carotenoids are among the most abundant constituents of the 270 thylakoid membrane. 271

2.3 Energies analysis and thermodynamic equilibrium.

In Table 3 are reported the averaged values and the corresponding standard deviations of the van der Waals and electrostatic components of the interaction energy, as well as the value of the total interaction energy (Uab) obtained from the analysis of the molecular dynamics trajectories of the complexes formed by CCD4a, b, c with selected carotenoids. 273

263

264

265

266

272

Carotenoids are highly hydrophobic molecules. The enzymes' binding site is a 278 highly hydrophobic linear channel as well, favoring interactions with the ligand that are 279 governed by short-range VDW-type interactions. Polar interactions are also present and 280 the contribution to the interaction energy is greater in xanthophylls as compared to aliphatic carotenes. The xanthophylls in their polar regions (sites characterized by the presence of hydroxyl groups) tend to form hydrogen bonds with water molecules or to interact 283 with other polar molecules of the system. 284

Table 3. Total interaction energy (Uab) and van der Waals and electrostatic components averaged286for all complexes. Averages are evaluated as a function of time for all trajectories. Best interaction287energy values are highlighted in yellow. All values are in kJ/mol. Abbreviations 3ON: 3-hydroxy-2888'-apocarptenol; ACR: α -carotene; BCR: β -carotene; LYC: lycopene; LUT: lutein; RRX: β -cryptoxan-289thin; ZEX: zeaxanthin.290

		Uab(ELECT)	Uab(VDW)	Uab(Total)
Receptor	Ligand	Mean(±SD)	Mean(±SD)	Mean(±SD)
	30N	-116.50 (±23.93)	-274.37 (±16.58)	-390.87 (±23.61)
	ACR	-33.11 (±10.16)	-344.21 (±18.75)	-377.32 (±22.07)
	BCR	-21.03 (±8.75)	-344.23 (±17.66)	-365.26 (±19.51)
CCD4a	LYC	-32.66 (±10.96)	-351.83 (±15.51)	-384.49 (±18.94)
	LUT	-110.95 (±21.48)	-342.63 (±17.54)	-453.58 (±23.83)
	RRX	-77.24 (±16.85)	-333.07 (±17.81)	-410.30 (±21.77)
	ZEX	-135.94 (±25.19)	-316.85 (±18.62)	-452.79 (±26.42)
	30N	-124.99 (±27.10)	-271.27 (±18.53)	-396.27 (±24.13)
	ACR	-33.41 (±10.80)	-351.57 (±16.37)	-384.97 (±21.24)
	BCR	-31.76 (±10.65)	-342.07 (±14.83)	-373.83 (±18.16)
CCD4b	LYC	-32.05 (±10.86)	-367.96 (±15.46)	-400.02 (±19.02)
	LUT	-119.87 (±23.65)	-331.35 (±17.78)	-451.22 (±24.61)
	RRX	-76.22 (±12.89)	-350.70 (±15.71)	-426.93 (±20.45)
	ZEX	-121.58 (±25.71)	-348.27 (±18.43)	-469.85 (±26.72)
	30N	-137.52 (±25.04)	-257.66 (±16.61)	-395.18 (±24.23)
	ACR	-29.88 (±10.07)	-355.33 (±14.78)	-385.21 (±17.95)
	BCR	-32.91 (±11.40)	-346.46 (±16.16)	-379.36 (±21.21)
CCD4c	LYC	-36.42 (±11.65)	-363.98 (±17.23)	-400.40 (±19.95)
	LUT	-137.28 (±28.76)	-318.32 (±18.68)	-455.60 (±26.78)
	RRX	-83.85 (±18.59)	-329.94 (±18.81)	-413.79 (±23.63)
	ZEX	-132.90 (±23.44)	-360.72 (±18.69)	-493.62 (±24.86)

292

In Figure 7 the evolution of the interaction energy of CCD4b with the best carotenoid 293 substrate β -cryptoxanthin (RRX) is shown as a function of the trajectory time. 294



Interaction energy of CCD4b-RRX complex

Figure 7. Evolution of the interaction energy of CCD4b in membrane with cryptoxanthin along the 20 ns molecular dynamics simulation trajectory. Electrostatic (red), van der Waals (green) and total (blue) potential energies are plotted.

time (ns)

2.4 Structural analysis (RMSD and RMSF).

In the Figures 8A, 8B and 8C are reported, as a function of the simulation time, the 302 RMSD values for the complexes formed by the CCD4a, b and c models with the seven 303 carotenoids under study. 304



Figure 8. Values of RMSD as a function of the MD simulation time for the models of the complexes306between CCD4a (a), CCD4b (b) and CCD4c (c) and seven carotenoid putative ligands. 3ON: 3-hy-
droxy-8'-apocarptenol, ACR: α -carotene; BCR: β -carotene; LUT: lutein; LYC: lycopene; RRX: cryp-
toxanthin: ZEX: zeaxanthin.308





To analyze the influence of the membrane on the structural stability of the models, 310 the RMSF values of models were calculated with and without the membrane. The Δ RMSF 311 values are shown in Figure 9. 312



313 314

Figure 9. Left panels: Δ RMSF values for CCD4a, CCD4b and CCD4c (for details see text). The first blue dotted squares from the left indicate the α 1 helices. The second blue dotted squares indicate the α 3 helices. Right panels: three-dimensional structures of the models. More mobile regions, corresponding to Δ RMSF>0, are colored in red. Less mobile regions (Δ RMSF<0) are colored in blue.

The α 1 and α 3 helical domains, having a direct interaction with the membrane, are 319 the regions most affected by the presence of the membrane, evidenced by a decrease of 320 the average fluctuations (negative Δ RMSF values). It is worthwhile to note that the α 1 321 helix is characterized by a more intense blue color, indicating a lower mobility with 322 respect to the α 3 helix. Regarding the regions whose mobility is increased (higher Δ RMSF 323 values and red colored in the Figure 9 Right), these are loops exposed to the stroma liquid 324 matrix of the chloroplast. 325

315 316 317

2.5 Carotenoids distances from the Fe⁺² atom and binding energy values

Binding energies between CCD4a, b and c and all carotenoids under study were evaluated from the molecular dynamics simulations as a time average. The averaged values of the MMPBSA binding free energy $\Delta G_{\text{binding}}$, and the corresponding standard deviation values, estimated for the structures extracted from the equilibrated molecular dynamics trajectories are reported in Table 4.

333

333

Table 4. Values of the distances between the C7, C8, C9, C10, C15 and C15' carbon atoms of each334carotenoid and the Fe^{+2} atom. Lower distance values are highlighted in yellow. In the last column335binding free energy values, evaluated by MMPBSA method are reported in kJ/mol units. Best bind-336ing energy values are highlighted in green. RRX* indicates the results obtained by extending the337MD trajectory of the CCD4b-RRX complex to 1 μ s.338

Receptor	Ligand	C7	C8	С9	C10	C15	C15′	MMPBSA (±SD)
	ACR	9.3	8.7	8.4	8.2	10.6	11.9	-148.50 (±14.88)
	BCR	8.2	7.1	6.3	6.0	8.0	9.2	-162.54 (±17.60)
CCD4a	LUT	9.5	8.5	8.6	8.3	11.2	11.9	-136.63 (±17.78)
	RRX	7.4	7.3	7.3	8.4	10.9	11.2	-133.16 (±18.15)
	ZEX	7.8	7.04	6.5	7.4	10.2	11.1	-124.20 (±17.85)
CCD4b	ACR	11.8	12.3	12.2	11.4	11.7	11.7	-123.18 (±17.91)
	BCR	11.1	11.2	10.7	9.5	9.4	10	-142.60 (±19.08)
	LUT	7.5	7.1	6.5	7.3	9.5	10.6	-119.18 (±29.12)
	RRX	10.1	9.9	8.9	7.8	5.8	6.8	- 150.19 (±16.83)
	RRX*	8.9	8.0	6.8	6.9	7.9	8.6	-164.57 (±15.74)
	ZEX	7.0	6.7	6.2	7.1	10	11	-88.43 (±22.31)
	ACR	12.3	12	11.3	10	7.9	7.3	-106.32 (±19.08)
CCD4c	BCR	10.9	9.9	8.5	8.2	5.8	6.3	-92.58 (±27.12)
	LUT	11.4	12.2	11.9	11.6	11.6	11.3	-105.77 (±28.35)
	RRX	10.8	11.2	10.5	9.8	8.7	9.12	-128.21 (±20.68)
	ZEX	12.2	11.14	9.93	8.8	6.4	6.8	-99.74 (±19.03)

As predicted, the best CCD4b substrate β -cryptoxanthin (RRX) displays the lowest 340 binding free energy. 341

Another observation that emerges from the analysis of Table 4 (distances 342 highlighted in yellow) is that the MD simulations indicate that the anchorage and the 343 catalytic event is mainly centered around a region near the carotenoid C7=C8 and C9=C10 344 double bonds, suggesting that the catalysis occurs in an asymmetric fashion. 345

When the MMPBSA binding energy values are considered together with the346distance values observed, it can be concluded that the tendency of the catalytic site of347oxidases is to position the carotenoids C9=C10 double bond near the Fe⁺² atom, promoting348in this way an asymmetric cleavage at this site.349

With respect to substrate specificity, data indicate that in the case of CCD4a an asymmetric oxidation in C9=C10 of the β -carotene as the best putative substrate would be favored. In fact, this compound displays the best binding energy of -162.54 kJ/mol, 352 highlighting a certain degree of carotenoid specificity for CCD4a. 353

In the case of CCD4c, if we take into account the distances between the carbon atoms 354 of the putative substrates and the Fe⁺² ion in the catalytic site, it could be favored an 355 asymmetric (C9=C10) as well as a symmetric (C15=C15') cleavage. In this case, β - 356 cryptoxanthin is predicted to be the best substrate with a binding energy of -128.21 kJ/mol. 357

A special attention was devoted to the CCD4b enzyme in which the binding energy 358 values point to β -cryptoxanthin as the best ligand. As a clear result was not obtained from 359

the 20 ns MD trajectory with respect to the cleavage site, a 1 µs trajectory was carried out 360 whose results are presented in Table 4 and indicated as RRX*. These results confirm β -361 cryptoxanthin as the best ligand and indicate that CCD4b catalyzes an asymmetric 362 cleavage of β -cryptoxanthin. 363

To analyze in deeper detail how RRX is positioned in the catalytic site of CCD4b and 364 the interactions taking place, the Ligand interaction tool of MOE was used. The results of 365 this analysis, shown in Figure 10, indicate that the main interactions are established with 366 the C9=C10 double bond. 367



Figure 10. (A) 2D representation of the interactions established by β -cryptoxanthin within the 370 CCD4b catalytic site. Only the residues contacting the carotenoid at less than 4.5 Å distance are shown. (B) 3D representation of the interactions in which the catalytic center and the Fe⁺² atom is 372 represented by a cyan sphere.

369

371

368

373

374

3.6 Analysis of membrane – protein interaction

The model of CCD4b inserted in the membrane, at the end of the 20 ns molecular 378 dynamics trajectory, is displayed in Figure 11. The evolution through the simulation of 379 the membrane thickness, the depth of penetration of the protein into the membrane and the relative membrane penetration of protein (RPD) are shown in Figures 12-14. 381



382 383

Figure 11. Three-dimensional view of the CCD4b inserted into the membrane. The phosphate 384 groups (dark yellow balls) were aligned in the XY plane and projected on the Z axis. A first set MP2 385 include all phosphate groups on the membrane side in which the protein was inserted. The second 386 set MP1 includes the phosphates belonging to the other layer of the membrane. is the maximum 387 atomic position observed for the protein along the Z axis for the 20 ns frame (not considering 388 hydrogen atoms). Z1 and Z2 are the mean values of the phosphate Z coordinates belonging to each 389 layer.

390

391

392

376



Figure 12. Membrane thickness for the three complexes throughout the simulations



Figure 13. Depth of penetration of the protein into the membrane measured for the three studied complexes.



Relative protein depth in membrane

Figure 14. Relative membrane penetration of protein (RPD) evaluated through all simulations.

The area involved in the contact observed throughout the molecular dynamic's 402 trajectory between the different proteins and the membrane indicates that the interaction 403 is significant and covers a non-negligible surface that typically ranges from 2250 to 3250 404 $Å^2$ (Supplementary Material – Figure S2). This contact surface is mainly dominated by 405 residues of the α -helical domain, as was observed in the structure of VP14 by Messing and 406 Amzel [22]. This allows us to suggest the same protein-membrane interaction mechanism 407 for Citrus CCD4s, previously pointed out in the studies by Ma et al. [32]. These results 408 indicate that a stable interaction, due to the orientation and penetration on the surface of 409 the thylakoid membrane, would enable CCD4s to extract the membrane-soluble 410 carotenoids through a tunnel formed mainly by hydrophobic residues that extends to the 411 catalytic center of the enzymes, as also hypothesized before by our research group [33]. 412

4. Discussion

Complete models of the members of the CCD4 family, including membrane, have 414been constructed in order to study the possible interaction with thylakoid membranes. 415 The models have been validated through molecular dynamics simulations in water. All of 416 them have displayed structural stability throughout the simulation trajectory. 417

The α 1 and α 3 helical domains included in these models have been found to be 418 essential for the interaction with thylakoid membranes, which promotes a decrease of 419 their movements. These domains are characterized by an abundance of nonpolar amino 420 acids that favor the interaction with the membrane, allowing a penetration of 421 approximately 10 Å that constitutes a relative penetration of approximately 30-35% of the 422 membrane width (see Figure 12, 13 and 14). 423

According to the interaction energy Uab values, a preference of the enzymes is 424 observed for the binding of xanthophylls (ZEX, LUT and RRX). When the binding energy 425 is calculated using the MMPBSA method a similar result is obtained with the difference 426 that also carotenes such as BCR are predicted to be putative substrates. These results 427 confirm what is known from experimental data, i.e. that both xanthophylls and carotenes 428 bind to CCD4 [18]. 429

413

399

Finally, the prediction of the possible oxidative cleavage site indicates a clearly 430 asymmetric tendency at the C7 = C8 and C9 = C10 positions, a result that is consistent with 431 what is already known for other members of the same enzymes family [18,27]. 432

4. Materials and Methods

The Materials and Methods should be described with sufficient details to allow oth-434 ers to replicate and build on the published results. Please note that the publication of your 435 manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission 437 stage any restrictions on the availability of materials or information. New methods and 438 protocols should be described in detail while well-established methods can be briefly described and appropriately cited. 440

4.1 Proteins structural model building

4.1.1 Modelling by threading

For the construction of structural models of the CCD4a, CCD4b and CCD4c 443 proteins, including the α 1 and α 3 domains that interact with the thylakoid membranes of 444 the chloroplast and are present only in proteins of plant origin, the software PHYRE2 445 (Protein Homology / analogY Recognition Engine) was used [34]. PHYRE2 combines the 446 threading/ab initio modeling method with the modeling of parts of the sequence by 447 homology with multiple templates. 448

Some structural elements of a model previously published by our group were used [33], given their established validity. In this sense, after an inspection of the Fe⁺² coordination sphere, the position of Fe^{+2} was defined by superimposing both models, verifying the Fe⁺² coordination sphere with the histidine residues (His257, His307, His373 and His548).

The models obtained for CCD4a, CCD4b and CCD4c were compared with the 454 structure of the Zea mays dioxygenase viviparous 14 (VP14), an enzyme active in the 455 cleavage of 9-cis-epoxycarotenoids (PDB ID: 3NPE). This is the only enzyme of plant 456 origin with known crystallographic data that possesses the $\alpha 1$ and $\alpha 3$ domains of 457 interaction with the membrane. VP14 is an enzyme acting on carotenoids in the 9-cis 458 configuration and cleaving the double bond at position 11 = 12 (11' = 12'), thus 459 participating to the first steps of the biosynthetic pathway of ABA (EC 1.13.11.51). The 460 resolution of the crystal structure is 3.2 Å, with an R-factor = 0.242 and R-free = 0.272. 461

4.1.2 Refinement and molecular dynamics

All calculations were carried out using NAMD2 version 2.12 software optimized to 463 work on CUDA cores [35]. 464

The force field selected for the simulation was the Charmm36 natively compatible 465 with the NAMD2 program[36]. The parameterization of the ligands was carried out 466 through the CGenFF (CHARMM General Force Field) program , compatible with the 467 Charmm36 force field [37]. In this program, the assignment of parameters and charges is 468 done by analogy, studying the connectivity patterns and the chemical environment of 469 each atom. For the Fe⁺² atom in coordination the unbound scheme was used. The Fe⁺²-470 protein interaction was modeled using electrostatic and VDW terms, thus allowing the 471 study of possible changes in the coordination center. The VDW interactions were modeled 472 by a 12-6 Lennard-Jones (LJ) potential following the Charmm36 force field model. The 473 parameters R_{min} and ε were obtained from Li et al. [38] and were optimized for this system. 474

The validation of the models was carried out by means of molecular dynamics 475 simulations, for which a water box with explicit solvent was built defining 15 Å of margin 476 between the outermost atom and the edge of the box in each direction, with the 477 dimensions of the periodic cells of 88 x 100 x 86 Å. NaCl counter ions were used to ensure 478 the overall charge neutrality of the system. The water boxes were built using the TIP3P 479

433

436

439

441 442

449

450

451

452

453

487

499

500

506

507

515

516

517

water models, and the water molecules were considered rigid bodies allowing their 480 rotation and translation along the box. 481

For all simulations, systems under periodic boundary conditions were considered, 482 with a time step of 2 fs. The sampling stages were carried out every 2 ps. The electrostatics 483 of the system was treated using the Particle Mesh Ewald (PME) method with a grid size 484 of 1 Å and a cutoff distance for long-range interactions of 12 Å. 485

Each simulation consisted of four stages:

• Refinement through energy minimization with 2000 steps using conjugated 488 gradients and applying harmonic restraints on the main chain atoms of the protein 489 and on the Fe⁺² coordination sphere to avoid possible deformations in the initial 490 stages. 491

• Heating stage from 0 to 300 K in 100 ps simulation time, keeping the pressure 492 constant at 1 atm and the harmonic restraints on the main chain atoms and the 493 coordination sphere of Fe^{+2} . 494

• Molecular dynamics simulation stage at a constant temperature of 300 K for 1 495 ns, maintaining the harmonic restraints of the previous stages. 496

Molecular dynamics simulation stage at constant temperature (300 K) for 20 ns, 497
 without harmonic restraints.

4.2 Construction of the CCD4X-Ligand Complexes

Given the structural similarity between the CCD4a, CCD4b, and CCD4c structural 501 models obtained by threading with those obtained by homology, the initial position of the ligands in the active site was established by structurally aligning both models and taking 503 the position of the carotenoid modelled in previous works of our research group [33]. All 504 complexes were validated by molecular dynamics simulations. 505

4.3 Molecular dynamics simulations of the complexes

The molecular dynamics simulation of the complexes was performed following the 508 protocol described above in the Refinement and Molecular Dynamics section. The 509 simulation was divided into four stages: energy minimization, heating from 0 to 300 K in 510 100 ps, equilibrium dynamics at 300 K for 1 ns restricting the position of the main chain 511 (main chain) atoms and production. Then, triplicated sampling trajectories (without any 512 type of restriction) of 20 ns were obtained and statistical sampling of the physicochemical 513 parameters was evaluated. 514

4.4 Construction of membrane complexes

4.4.1 α -helical domains and their interaction with membranes

To propose the spatial arrangement of protein structures in lipid bilayers we used the 518 OPM (Orientations of Proteins in Membranes) program [39], in which each protein is 519 considered as a rigid body that can freely interact with a hydrophobic layer of adjustable 520 width. The orientation of the protein is optimized by a rigid body molecular dynamics 521 simulation, always considering the protein as the solute whose relative position with 522 respect to the membrane must be optimized by means of an implicit (non-structural) 523 representation. As a first step, the free energy $\Delta G_{transfer}$ of protein transfer to the membrane 524 from an aqueous solution medium is minimized by a calculation at different penetration 525 levels, moving the protein at a speed of 0.2 Å / step. Finally, the optimal penetration depth, 526 the angle of rotation of the center of mass with respect to the plane formed by the 527 hydrophobic layer, the residues in contact with the membrane and the residues outside it 528 and the optimal value of the $\Delta G_{\text{transfer}}$ are reported. 529

544

545 546

547

548

549

550

551

552

553

554

555

556

557 558

559

565

566

570

571

572

573

574

For the construction of the membrane, we used a pre-equilibrated membrane model 530 of phosphatidylcholine (POPC) which initially covered a surface area of 102 x 140 Å. The 531 initial thickness, measured as the distance between the bilayer phosphates, was 39 Å for 532 the three CCD4a, b and c models. The protein structures were coupled to the membrane 533 layers, respecting the depth and orientation obtained in the previous step and removing 534 the POPC units that overlapped with the protein. 535

The model of the protein-membrane complexes was subjected to molecular dynamics 536 simulations following the protocol described above in the Refinement and Molecular 537 Dynamics section, keeping the surface of the membrane constant (formed by the X-Y 538 plane). The simulation was divided into four stages: energy minimization, heating from 0 539 to 300 K in 100 ps, equilibrium dynamics at 300 K for 1 ns restricting the position of the 540 main chain (main chain) atoms and a production of all the complexes models was 541 obtained with triplicated trajectories of 20 ns. 542

Finally, a one microsecond (1 μ s) trajectory was obtained for the CCD4b- β cryptoxanthin complex to model a more biologically realistic event, considering that this is one of the substrates reported in literature [27].

4.5 Analysis of molecular dynamics simulations

4.5.1 Ligands Interaction energy (Uab)

The ligand interaction with the modelled macromolecular system (protein, solvent and ions) has been evaluated through molecular dynamics simulation, according to the Equation 1:

$$Uab\rangle = \langle U - Ua - Ub\rangle \tag{1}$$

where,

U is the total potential energy of the system,

Ua is the potential energy of receptor, ions and solvent (ligand was excluded), Ub is the potential energy of the ligand,

(

Uab is composed by non-covalent energy partition and is splitted in the van der Waals

(VDW) and the electrostatic (ELECT) components.

4.5.2 Protein-membrane interactions

For the study of the influence of the membrane on the CCD4x domains throughout 560 the molecular dynamics simulations, the residues that participate directly in the 561 interaction, the temporal evolution of the contact surface between the protein and the 562 membrane and the characterization of the non-bonded interactions, such as the pattern of 563 hydrogen bonds, the electrostatic and VDW components, were analyzed. 564

In addition, changes in the membrane structure related to the insertion of these proteins has been studied.

The calculation of the mobility of the residues that make up a protein is a very 567 important measure of its stability; one of the ways to represent it is the mean square 568 fluctuation or RMSF (Root Mean Square Fluctuation). 569

The differences between the root mean square fluctuations (Δ RMSF) of the CCD4x models solvated or inserted in the membrane were calculated following Equation 2:

$$\Delta RMSF_{i} = \langle RMSF_{i} \rangle_{memb} - \langle RMSF_{i} \rangle_{sol}$$
⁽²⁾

where,

 $\langle RMSF_i \rangle_{memb}$ is the value of the root mean square fluctuation of residue i in the membrane interaction model,

 $(RMSF_i)_{sol}$ is the value of the root mean square fluctuation of residue i in the solvent 575 model. 576

A set of in house developed algorithms was used to estimate three parameters in 578 relation to the penetration of the protein inside the membrane throughout the simulation, 579 i.e. the time evolution of membrane thickness, the depth of penetration of the protein 580 inside the membrane and the relative membrane penetration of the protein. 581

To measure the penetration depth of the protein into the membrane, the phosphate groups aligned in the XY plane and projected on the Z axis were grouped in two sets: a first set MP2 is defined by all phosphate groups on the membrane side in which the protein was inserted. The second set MP1 is defined by the phosphate groups belonging to the other layer of the membrane.

The average thickness of the membrane was calculated according to Equation 3:

Thickness_t =
$$\frac{\sum_{i} (Z_{MP1(i)})_{t}}{n_{(i)}} - \frac{\sum_{j} (Z_{MP2(j)})_{t}}{n_{(j)}}$$
 (3)

where,

 $Z_{MP1(i)}$ is the set of values on the Z axis of the cartesian coordinates of the i phosphates that belong to the set MP1 at time t,

 $Z_{MP2(j)}$ is the set of values on the Z axis of the cartesian coordinates of the j phosphates that belong to the set MP2 at time t,

 $n_{(i)}$ is the number of phosphates belonging to the MP1 set,

 $n_{(i)}$ is the number of phosphates belonging to the MP2 set.

The depth of penetration of the protein inside the membrane has been calculated using Equation 4:

$$Depth_{t} = Max(Z_{prot})_{t} - \frac{\sum_{j} (Z_{MP2(j)})_{t}}{n_{(j)}}$$
(4)

where,

 $Max(Z_{prot})_{t}$ is the observed maximum value of the atomic cartesian Z coordinate of the protein (not considering hydrogen atoms).

Given that the thickness of the membrane is variable through the simulation, the relative membrane penetration of the protein (RPD) through all the simulation was 606 evaluated using Equation 5:

$$RPD_t = \frac{Depth_t}{Thickness_t} * 100$$
(5)

609 610

4.5.3 Calculation of binding free energy using the MMBPSA method

The conformational sampling for MMBPSA binding free energy calculation was 611 carried out on the trajectories generated during the equilibrium molecular dynamics. 612 Three energy contributions were calculated following the MMPBSA method. First, in the 613 gas phase, the energy difference between the complex and the sum of the energies of the 614 receptor and the ligand separately was calculated, using NAMD and the Charmm36 force 615 field (MM). Then, the polar contribution to the free energy of solvation was calculated 616 numerically using the Poisson-Boltzmann (PB) equation implemented in the APBS 617 (Adaptive Poisson-Boltzmann Solver) program [40]. Subsequently, the changes 618 (differences) of the solvent accessible surface area (SASA) were measured and from them 619 the nonpolar contribution to the free energy of solvation was estimated, through the linear 620 relationship with SASA. Finally, the binding free energy $\Delta G_{\text{binding}}$ was estimated by the 621 following Equation 6: 622

623

593 594

590

591

592

582

583

584

585

586 587

588 589

- 595 596
- 597 598
- 599 600

601 602

603

604 605

$$\Delta G_{\text{binding}} = \Delta H - T\Delta S = \langle \Delta M M_{\text{gas}} + \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}} - T\Delta S \rangle \tag{6}$$

where,

H is the enthalpy, *T* is the absolute temperature, *S* is the entropy of the system,

 ΔMM_{gas} is the energy difference, calculated by molecular mechanics, between the complex and the sum of the energies of the receptor and the ligand separately, in the gas phase,

 ΔG_{polar} is the polar contribution to the free energy of solvation, $\Delta G_{nonpolar}$ is the non-polar contribution to the free energy of solvation, $T\Delta S$ is the entropic contribution to the free energy.

The group of compounds studied is structurally similar, as they all belong to the 635 structural class "carotenoids". It can be assumed that their contribution to the differences 636 in entropy is similar and that therefore the changes in free energy due to the entropic 637 factors cancel each other out when the differences between the free energies are calculated 638 (expressed as $\Delta\Delta G$). As a consequence, the T ΔS term can be ignored as usual in the 639 MMBPSA methods. 640

5. Conclusions

Novel tridimensional models of CCD4 (a, b and c) from Citrus clementina, including 643 α -helix domains that interact with membranes have shown structural stability throughout 644 molecular dynamics simulations. 645

Evidence is provided that the α 1 and α 3 helices domains included in these models are 646 essential for the interaction with thylakoid membranes and that in their presence the 647 average fluctuation in these domains decreases, stabilizing the structure. The region that 648involves $\alpha 1$ is characterized by the domain of polar amino acids that interact with the 649 polar groups of the membrane, α 3 is characterized by the abundance of apolar amino acids 650 that favor interaction with the hydrophobic nucleus of the membrane. A protein 651 penetration of 10 Å inside the membrane constitutes a relative penetration of 652 approximately 30-35% to the membrane. 653

The averaged values of the MMPBSA binding free energy $\Delta G_{\text{binding}}$, indicates the 654 preference in the union with xanthophylls (ZEX, LUT and RRX), without excluding 655 carotenes such as BCR as possible substrates, in agreement with what is known from 656 experimental data (Ahrazem et al., 2016). 657

Finally, the possible site of oxidative cleavage shows a clearly asymmetric tendency 658 at positions 7 = 8 (7 = 8') and 9 = 10 (9 = 10'), a result consistent with other already known 659 members of the CCD4 type (Ahrazem et al., 2016; Rodrigo et al., 2013). 660

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, 662 Figure S1: title, Table S1: title, Video S1: title., 663

Author Contributions: Conceptualization, F.P. and M.P.; methodology, J.C., F.P. and 665 M.P.; software, J.C.; validation, J.C. and M.P.; formal analysis, J.C.; investigation, J.C., M.P. 666 and F.P.; resources, F.P. and M.P.; writing-original draft preparation, J.C. and M.P.; 667 writing-review and editing, F.P.; visualization, J.C.; supervision, M.P.; project 668 administration, M.P.; funding acquisition, F.P. and M.P. 669

All authors have read and agreed to the published version of the manuscript.

671 672

670

661

664

624 625 626

627 628

629 630 631

632

633 634

641

674
675
676

680 681

682

Funding: This research was funded by the Italian Ministry of University and Research	673
(MIUR), grants "Dipartimenti di Eccellenza" and PRIN (grant n. 2017483NH8).	674

Acknowledgments: J.C acknowledge the fellowship 2011-2016 funding from INBIO (Instituto de Biotecnología Agrícola) – Paraguay. J.C and M.P acknowledge to 677 PEDECIBA Química for their support to QUITEL congresses 2016 and 2018. 678

Conflicts of Interest: The authors declare no conflict of interest.

References

1.	Johnson, E.J. The Role of Carotenoids in Human Health. Nutr. Clin. Care 2002, 5, 56–65, doi:10.1046/j.1523-5408.2002.00004.x.	683 684
2.	Karrer, P. Carotenoids, flavins and vitamin A and B2: Nobel lecture, December 11, 1937. In <i>Nobel Lecture</i> ; Elsevier: Amsterdam, 1966; pp. 433–448.	685 686
3.	Ahmad, T.; Cawood, M.; Iqbal, Q.; Arino, A.; Batool, A.; Tariq, R.M.S.; Azam, M.; Akhtar, S. Phytochemicals in Daucus Carota and Their Health Benefits-Review Article. <i>Foods</i> (<i>Basel, Switzerland</i>) 2019 , <i>8</i> , doi:10.3390/foods8090424.	687 688 689
4.	Zhao, C.; Nabity, P.D. Phylloxerids Share Ancestral Carotenoid Biosynthesis Genes of Fungal Origin with Aphids and Adelgids. <i>PLoS One</i> 2017 , <i>12</i> , 1–14, doi:10.1371/journal.pone.0185484.	690 691
5.	Ikoma, Y.; Matsumoto, H.; Kato, M. Diversity in the Carotenoid Profiles and the Expression of Genes Related to Carotenoid Accumulation among Citrus Genotypes. <i>Breed. Sci.</i> 2016 , <i>66</i> , 139–147, doi:10.1270/jsbbs.66.139.	692 693
6.	Latowski, D.; Kuczyńska, P.; Strzałka, K. Xanthophyll Cycle - a Mechanism Protecting Plants against Oxidative Stress. <i>Redox Rep.</i> 2011 , <i>16</i> , 78–90, doi:10.1179/174329211X13020951739938.	694 695
7.	Shumskaya, M.; Wurtzel, E.T. The Carotenoid Biosynthetic Pathway: Thinking in All Dimensions. <i>Plant Sci.</i> 2013 , 208, 58–63, doi:10.1016/j.plantsci.2013.03.012.	696 697
8.	Wurtzel, E.T. Changing Form and Function through Carotenoids and Synthetic Biology. <i>Plant Physiol.</i> 2019 , <i>179</i> , 830–843, doi:10.1104/pp.18.01122.	698 699
9.	Li, L.; Yuan, H.; Zeng, Y.; Xu, Q. Plastids and Carotenoid Accumulation. <i>Subcell. Biochem.</i> 2016, 79, 273–293, doi:10.1007/978-3-319-39126-7_10.	700 701
10.	Hashimoto, H.; Uragami, C.; Cogdell, R.J. Carotenoids and Photosynthesis. <i>Subcell. Biochem.</i> 2016 , <i>79</i> , 111–139, doi:10.1007/978-3-319-39126-7_4.	702 703
11.	von Lintig, J. Metabolism of Carotenoids and Retinoids Related to Vision. J. Biol. Chem. 2012, 287, 1627–1634, doi:10.1074/jbc.R111.303990.	704 705
12.	Parry, A.D.; Babiano, M.J.; Horgan, R. The Role of Cis-Carotenoids in Abscisic Acid Biosynthesis. <i>Planta</i> 1990 , <i>182</i> , 118–128, doi:10.1007/BF00239993.	706 707
13.	Jia, KP.; Baz, L.; Al-Babili, S. From Carotenoids to Strigolactones. J. Exp. Bot. 2018, 69, 2189–2204, doi:10.1093/jxb/erx476.	708 709
14.	Cazzonelli, C.I. Carotenoids in Nature: Insights from Plants and Beyond. <i>Funct. Plant Biol.</i> 2011 , <i>38</i> , 833–847, doi:10.1071/FP11192.	710 711

15. Bai, S.; Tuan, P.A.; Tatsuki, M.; Yaegaki, H.; Ohmiya, A.; Yamamizo, C.; Moriguchi, T. Knockdown of Carotenoid 712

Cleavage Dioxygenase 4 (CCD4) via Virus-Induced Gene Silencing Confers Yellow Coloration in Peach Fruit: 713 Evaluation of Gene Function Related to Fruit Traits. *Plant Mol. Biol. Report.* **2016**, *34*, 257–264, doi:10.1007/s11105-015-0920-8. 715

- Harrison, P.J.; Bugg, T.D.H. Enzymology of the Carotenoid Cleavage Dioxygenases: Reaction Mechanisms, 716 Inhibition and Biochemical Roles. *Arch. Biochem. Biophys.* 2014, 544, 105–111, doi:10.1016/j.abb.2013.10.005.
- Kloer, D.P.; Schulz, G.E. Structural and Biological Aspects of Carotenoid Cleavage. *Cell. Mol. Life Sci.* 2006, 63, 718
 2291–2303, doi:10.1007/s00018-006-6176-6.
- Ahrazem, O.; Rubio-Moraga, A.; Berman, J.; Capell, T.; Christou, P.; Zhu, C.; Gómez-Gómez, L. The Carotenoid 720 Cleavage Dioxygenase CCD2 Catalysing the Synthesis of Crocetin in Spring Crocuses and Saffron Is a Plastidial 721 Enzyme. *New Phytol.* 2016, 209, 650–663, doi:10.1111/nph.13609. 722
- Kloer, D.P.; Ruch, S.; Al-Babili, S.; Beyer, P.; Schulz, G.E. The Structure of a Retinal-Forming Carotenoid 723 Oxygenase. *Science* 2005, *308*, 267–9, doi:10.1126/science.1108965.
- 20. Rodrigo, M.J.; Alquézar, B.; Alós, E.; Lado, J.; Zacarías, L. Biochemical Bases and Molecular Regulation of 725 the Peel of Citrus Fruit. Sci. Hortic. (Amsterdam). 2013, 163, Pigmentation in 46-62, 726 doi:https://doi.org/10.1016/j.scienta.2013.08.014. 727
- Tan, B.C.; Schwartz, S.H.; Zeevaart, J.A.; McCarty, D.R. Genetic Control of Abscisic Acid Biosynthesis in Maize.
 Proc. Natl. Acad. Sci. U. S. A. 1997, 94, 12235–12240, doi:10.1073/pnas.94.22.12235.
- Messing, S.A.J.; Mario Amzel, L.; Gabelli, S.B.; Echeverria, I.; Vogel, J.T.; Guan, J.C.; Tan, B.C.; Klee, H.J.; McCarty, 730
 D.R. Structural Insights into Maize Viviparous14, a Key Enzyme in the Biosynthesis of the Phytohormone 731
 Abscisic Acid. *Plant Cell* 2010, 22, 2970–2980, doi:10.1105/tpc.110.074815. 732
- Schwartz, S.H.; Tan, B.C.; Gage, D.A.; Zeevaart, J.A.; McCarty, D.R. Specific Oxidative Cleavage of Carotenoids 733 by VP14 of Maize. *Science* 1997, 276, 1872–1874, doi:10.1126/science.276.5320.1872.
- Ohmiya, A. Carotenoid Cleavage Dioxygenases and Their Apocarotenoid Products in Plants. *Plant Biotechnol.* 735
 2009, 26, 351–358, doi:10.5511/plantbiotechnology.26.351.
 736
- Ahrazem, O.; Trapero, A.; Gómez, M.D.; Rubio-Moraga, A.; Gómez-Gómez, L. Genomic Analysis and Gene Structure of the Plant Carotenoid Dioxygenase 4 Family: A Deeper Study in Crocus Sativus and Its Allies.
 Genomics 2010, *96*, 239–250, doi:https://doi.org/10.1016/j.ygeno.2010.07.003.
- Huang, F.-C.; Molnár, P.; Schwab, W. Cloning and Functional Characterization of Carotenoid Cleavage 740 Dioxygenase 4 Genes. J. Exp. Bot. 2009, 60, 3011–3022, doi:10.1093/jxb/erp137.
- 27. Rodrigo, M.J.; Alquézar, B.; Alís, E.; Medina, V.; Carmona, L.; Bruno, M.; Al-Babili, S.; Zacarías, L. A Novel
 742 Carotenoid Cleavage Activity Involved in the Biosynthesis of Citrus Fruit-Specific Apocarotenoid Pigments. J.
 743 *Exp. Bot.* 2013, *64*, 4461–4478, doi:10.1093/jxb/ert260.
 744
- Vallabhaneni, R.; Bradbury, L.M.T.; Wurtzel, E.T. The Carotenoid Dioxygenase Gene Family in Maize, Sorghum, 745 and Rice. Arch. Biochem. Biophys. 2010, 504, 104–111, doi:10.1016/j.abb.2010.07.019.
 746
- Ytterberg, A.J.; Peltier, J.-B.; van Wijk, K.J. Protein Profiling of Plastoglobules in Chloroplasts and Chromoplasts. 747 A Surprising Site for Differential Accumulation of Metabolic Enzymes. *Plant Physiol.* 2006, 140, 984–997, 748 doi:10.1104/pp.105.076083. 749
- Zheng, X.; Xie, Z.; Zhu, K.; Xu, Q.; Deng, X.; Pan, Z. Isolation and Characterization of Carotenoid Cleavage 750 Dioxygenase 4 Genes from Different Citrus Species. *Mol. Genet. Genomics* 2015, 290, 1589–1603, 751 doi:10.1007/s00438-015-1016-8.

31.	Chemical Computing Group Inc Molecular Operating Environment (MOE) 2014.	753
32.	Ma, G.; Zhang, L.; Matsuta, A.; Matsutani, K.; Yamawaki, K.; Yahata, M.; Wahyudi, A.; Motohashi, R.; Kato, M.	754
	Enzymatic Formation of Beta-Citraurin from Beta-Cryptoxanthin and Zeaxanthin by Carotenoid Cleavage	755
	Dioxygenase4 in the Flavedo of Citrus Fruit. Plant Physiol. 2013, 163, 682-695, doi:10.1104/pp.113.223297.	756
33.	Vega-Teijido, M.; Cantero, J.; Rodrigo, M.J.; López, C.; Paulino Zunini, M. An In Silico Study of the Citrus	757
	Dioxygenases CCD4 Family Substrates. J. Biomol. Struct. Dyn. 2018, 0, 1–28, doi:10.1080/07391102.2018.1477619.	758
34.	Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 Web Portal for Protein Modeling,	759
	Prediction and Analysis. Nat. Protoc. 2016, 10, 845–858, doi:10.1038/nprot.2015-053.	760
35.	Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J. V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.;	761
	Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. J. Chem. Phys. 2020,	762
	<i>153</i> , doi:10.1063/5.0014475.	763
36.	Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D. Optimization of the Additive	764
	CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ , ψ and Side-Chain X1	765
	and X2 Dihedral Angles. J. Chem. Theory Comput. 2012 , 8, 3257–3273, doi:10.1021/ct300400x.	766
37.	Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes,	767
	P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with	768
	the CHARMM All-Atom Additive Biological Force Fields. J. Comput. Chem. 2010, 31, 671–690,	769
20		770
38.	Li, P.; Roberts, B.P.; Chakravorty, D.K.; Merz, K.M. Rational Design of Particle Mesh Ewald Compatible Lennard-	771
	doi:10.1021/ct400146w	772
20	Lomize MA: Receptore ID: Ice H: Mechang HI: Lomize AI OPM Database and PDM Web Server	773
59.	Resources for Positioning of Proteins in Membranes Nucleic Acids Res 2012 40 370–376 doi:10.1093/nar/gkr703	774
40	Lurras E - Engel D - Star V - Manson V - Brandi L - Felberg J E - Braakes D H - Wilson J - Chan J - Liles V -	775
40.	et al Improvements to the APBS Biomolecular Solvation Software Suite Protein Sci 2018 27 112–128	770 777
	doi:10.1002/pro.3280.	778
	· 1	779

Supplemetary Material



Figure S1. Temporal evolution of hydrogen bonds between proteins and membranes. CCD4a (red), CCD4b (green) and CCD4c (blue). Averaged H bond amounts are nine for CCD4^a, seven for CCD4b and ten for CCD4c.

Hydrogen bonds between proteins and membrane



Figure S2. Temporal evolution of contact surface area ($Å^2$) in the complexes CCD4a (red), CCD4b (green) and CCD4c (blue). CCD4a has a mean of 2797.15 (±153.89), CCD4b : 2674.08 (±163.75) and CCD4c: 2885.87 (±118.13).



Anexo 3

