

Lorena Tremiño Agulló

Bases estructurales de la señalización y regulación por nitrógeno y procesos asociados

Dirigida por: Vicente Rubio Zamora



INSTITUTO DE BIOMEDICINA DE VALENCIA CSIC



UNIVERSITAT POLITÈCNICA DE VALÈNCIA



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

BASES ESTRUCTURALES DE LA SEÑALIZACIÓN Y REGULACIÓN POR NITRÓGENO Y PROCESOS ASOCIADOS

Memoria presentada por
Lorena Tremiño Agulló
Para optar al grado de Doctora

Director:
Dr. Vicente Rubio Zamora

*Unidad de Enzimopatología Estructural
Departamento de Genómica y Proteómica
Instituto de Biomedicina de Valencia (IBV)
Consejo Superior de Investigaciones Científicas (CSIC)*

Valencia, a 16 de Noviembre de 2018

Para la realización de esta Tesis, Lorena Tremiño Agulló ha disfrutado de una Beca de Formación de Personal Investigador (FPI) (BES-2012-058304) otorgado por el Ministerio de Economía, Industria y Competitividad.

El trabajo se ha llevado a cabo en el grupo 739 del CIBERER-Instituto de Salud Carlos III (IP, V. Rubio) y se ha enmarcado dentro de los proyectos:

- “Luz estructural sobre señalización y regulación por nitrógeno y sobre biosíntesis de arginina/urea, sus errores congénitos, y su conexión con biología del envejecimiento”, (BFU2011-30407) del Ministerio de Economía y Competitividad del Gobierno de España (Investigador principal, V. Rubio).
- “Una mirada molecular al control de la detoxificación de amonio y a sus patologías y errores congénitos, y a la señalización por amonio. En busca del papel de la proteína CutA”, (BFU2014-58229) del Ministerio de Economía y Competitividad del Gobierno de España (Investigador principal, V. Rubio).
- "BioMeder. Genes, proteínas y rutas de señalización en enfermedades raras" (PrometeoII/2014/029) de la Conselleria d'Educació de la Generalitat Valenciana (investigadores, P. Sanz, A. Marina y V. Rubio).

**ACEPTACIÓN DE LOS COAUTORES PARA QUE LA DOCTORANDA
PRESENTE LOS SIGUIENTES ARTÍCULOS COMO TESIS; Y RENUNCIA
EXPRESA A PRESENTARLOS COMO PARTE DE OTRA TESIS DOCTORAL**

Los/las coautores/coautoras de los artículos científicos citados al pie, Drs. Alicia Forcada Nadal, Javier Espinosa Manzano, Asunción Contreras de Vera y Vicente Rubio Zamora, aceptan que la doctoranda D^a Lorena Tremiño Agulló, con DNI 74370568E, presente los artículos listados a continuación como parte de la Tesis Doctoral con título “**BASES ESTRUCTURALES DE LA SEÑALIZACIÓN Y REGULACIÓN POR NITRÓGENO Y PROCESOS ASOCIADOS**”, y renuncian a presentarlos como parte de otra tesis doctoral.

Artículos científicos:

[1] Tremiño L, Forcada-Nadal A, Contreras A, Rubio V. (2017) Studies on cyanobacterial protein PipY shed light on structure, potential functions, and vitamin B₆-dependent epilepsy. *FEBS Lett.* 591, 3431-3442.

[2] Tremiño L, Forcada-Nadal A, Rubio V. (2018) Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations. *Hum Mutat.* 39, 1002-1013.

[3] Tremiño L, Espinosa J, Contreras A, Rubio V. Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria. *Manuscript to be submitted.*

Valencia, 8 de noviembre de 2018.



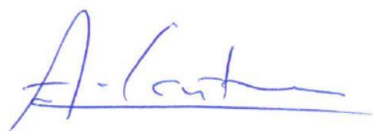
Dra. Alicia Forcada Nadal



Dr. Vicente Rubio Zamora



Dr. Javier Espinosa Manzano



Dra. Asunción Contreras de Vera

AGRADECIMIENTOS

Tras unos años de intenso trabajo y formación, tras varios viajes a congresos y sincrotrones, tras haber vivido multitud de experiencias científicas interesantes... llega el momento de demostrar lo que he estado haciendo y todo lo que he aprendido durante esta etapa. Como es lógico, esta tesis recoge la parte académica de mi período como estudiante predoctoral, pero, además, este apartado de *Agradecimientos* refleja la parte social/emocional que me ha permitido crecer como persona y convertirme en quien soy a día de hoy.

En primer lugar, quiero agradecer a mi director de tesis la oportunidad que me dio al concederme la beca FPI asociada a proyecto que le habían asignado. Vicente, gracias por dejarme formar parte de tu grupo, por permitirme usar todo el material e instalaciones del laboratorio y por haberme enseñado tanto.

Gracias a mi profesora de genética Asunción Contreras de la Universidad de Alicante con la que he tenido la oportunidad de colaborar y desarrollar gran parte del trabajo de esta tesis. Por supuesto, gracias también a todas las personas de su laboratorio por acogerme, especialmente a Javier Espinosa que siempre ha estado para resolver mis dudas.

Gracias también a todos los compañeros del laboratorio. ¡Os deseo lo mejor, que tengáis mucha suerte y que continuéis disfrutando de lo que hacéis!

- a los que estuvieron antes que yo y he tenido la oportunidad de conocer personalmente: Lali, Santi, Alberto, Leonor, Fer, Enea, Jose y Mariano. Gracias por vuestros consejos y por vuestras figuras, que, aunque no sepáis, muchas veces me han servido de inspiración para hacer las mías.

- a los que han coincidido temporalmente conmigo: Carmen, Irene, Phu, Juanma, Sergio, Carles y Silvia. Gracias por seguir estando en todo momento al otro lado del teléfono o por correo.

- a los que aún quedan: Belén, Clara, Nadine y Alicia. Gracias por todo lo que me habéis enseñado y por todos los momentos juntas.

A mis cuatro estudiantes de prácticas externas o trabajos de fin de grado con los que he aprendido muchísimo: Martín, Diego, Laura y Nacho. Gracias por transmitirme vuestra ilusión.

Gracias también a todos los compañeros del IBV:

- personal de seguridad y conserjería: gracias por esos buenos días por la mañana que suelen venir acompañados de una sonrisa
- limpieza: gracias por mantener el IBV en perfectas condiciones.
- administración: especialmente a Mari Luz y Julio, gracias por estar siempre dispuestos a ayudarme a resolver todos los trámites burocráticos.
- informática: Manolo y Kiko, ¡mil gracias por todo! por todos los programas instalados, por conseguir que pueda hacer animaciones de proteínas en full-HD, por vuestras recomendaciones para la compra de un portátil personal... Algún día descubriré cómo sois capaces de hacer funcionar la impresora que “no va” con sólo mirarla.
- biblioteca: Amparo, la dulzura y elegancia que te caracterizan encajan perfectamente con tu eficacia (eres la única persona que conozco capaz de encontrar artículos imposibles), ¡muchas gracias por todo!
- Ada, gracias por estar para todo!
- Carmen, gracias por transmitirme tu energía en los entrenamientos por el río.
- Gracias en general a todos los compañeros que siempre estáis dispuestos a ayudar con vuestros consejos y a prestar el material necesario para los experimentos.

Y cómo no, quiero dar las gracias también a mi querida Comisión De Fiestas (CDF) del IBV: Carles, Carla, Jordi, Raquel, Edu, Carol y Alicia. Desde que llegué me tratasteis como a una más, y hasta me nombrasteis nada más y nada menos que “Presidenta”, qué honor! Gracias por todo amigos: cañas/vinitos, paellas, fiestas, juegos, cine...

Durante estos años también he tenido la suerte de poder formar parte del Comité Organizador de Conbiopreval, algo que empezó siendo un Congreso de Biomedicina para estudiantes predoctorales de Valencia y que actualmente ya es un evento anual de carácter nacional, y altamente recomendado para todos los doctorandos del campo. Gracias equipo!

A mis amigas de la carrera, las biólogas: Sandra, M^a José y Nuria. Gracias por escucharme, comprenderme y apoyarme durante todo este tiempo. Es increíble que por mucho tiempo que pasemos sin vernos, cuando nos juntamos es como si no hubiese pasado ni un día desde la última vez.

A todos los amigos y familia de Elche y de Puebla de Don Fadrique (Granada), gracias por la curiosidad que habéis mostrado por mi trabajo y por todo vuestro apoyo. En especial quiero dar las gracias a todos los amigos y familiares que habéis venido a visitarme, con vosotros Valencia me gusta más! María y Naiara, gracias por estar siempre al otro lado del teléfono con vuestras palabras de ánimo y comprensión, gracias también por todas esas cenas y momentos inolvidables que hemos vivido y nos quedan por vivir.

A mis padres. Gracias por ser mi mejor ejemplo, por enseñarme todo de la vida, por darme todo lo que tenéis y por conseguir que me esfuerce tanto para conseguir lo que quiero, como bien me habéis dicho tantas veces “quien algo quiere, algo le cuesta”. Y esta tesis es prueba de ello, ha habido momentos muy buenos, pero otros algo duros, y con vuestro apoyo parece que al fin lo he conseguido.

A Rosa, mi hermana, compañera de risas desde pequeñas. Siempre has estado a mi lado cuando te he necesitado y me has ensañado a ver la parte práctica de las cosas. Gracias por todo!

Y por supuesto gracias a ti, Miguel Ángel. Por mucho que escribiese no tendría espacio suficiente ni encontraría las palabras apropiadas para agradecértelo todo.

Resumen

Enmarcada en la línea de investigación de nuestro laboratorio sobre señalización por nitrógeno principalmente en la cianobacteria *Synechococcus elongatus* PCC7942, con esfuerzos centrados en las proteínas PII y PipX y su red de señalización, esta Tesis amplía el espectro de moléculas investigadas en relación con dicha red.

Estudia y caracteriza el miembro no canónico de la superfamilia de la proteína PII denominado CutA, generalmente anotado como de protección frente a metales divalentes, altamente conservado en todos los dominios de la vida (incluidos animales y el ser humano). En ella examinamos la posible protección frente a metales provista por CutA en dos bacterias muy distantes, *Escherichia coli* y *Synechococcus elongatus*, usando *knockouts* para ambas del gen que codifica para CutA. Ni los estudios de complementación en *E. coli* del gen silvestre ni los observacionales de sensibilidad a metales en *S. elongatus* han dado soporte a la función anotada para CutA, a pesar de que demostramos mediante seguimiento turbidimétrico que el Cu^{2+} hace agregar a la proteína CutA pura de *S. elongatus* (producida recombinantemente) y por tanto se une a ella, aunque con una afinidad baja por comparación con las concentraciones tóxicas de este metal para dicha cianobacteria.

Buscando profundizar en el conocimiento de CutA, hemos determinado a muy alta resolución mediante difracción de rayos X la estructura de esta proteína de *S. elongatus*, sin evidenciar complejo alguno con cobre, pero demostrando que los tres bolsillos intersubunidades en el homotrímero de CutA son capaces de transportar moléculas orgánicas (en nuestro caso Bis-Tris). Estos resultados apoyan una posible función de CutA basada en la unión a estos bolsillos de biomoléculas neutras o positivamente cargadas y capaces de formar varios puentes de hidrógeno con las paredes de potencial negativo y fuerte carácter polar de estos bolsillos.

También hemos estudiado la proteína PipY de *S. elongatus*, identificada recientemente como pareja funcional de la antes mencionada PipX, determinando sus propiedades espectroscópicas, unión de piridoxal fosfato (PLP) y resolviendo su estructura mediante difracción de rayos X. Probamos que PipY es monomérica y que tiene PLP unido. Su estructura no apoya que sea un enzima, siendo aparentemente apropiada para ejercer una posible función en la homeostasis de PLP. Dado que muy recientemente se ha descrito una epilepsia genética humana dependiente de vitamina B₆ debida a mutaciones en el gen humano ortólogo de *pipY*, *PROSC* (ahora llamado *PLPBP*; codifica la proteína PLPHP), usamos inicialmente PipY de *S. elongatus* y luego *PROSC* humana como banco de pruebas de la patogenicidad de las mutaciones que se han asociado a esta epilepsia, utilizando para ello mutagénesis dirigida y producción de las formas silvestre y mutantes de estas proteínas, comparando sus propiedades. Estos estudios han demostrado la patogenicidad y establecido mecanismos para la misma para cada una de las mutaciones de cambio de sentido de *PROSC* descritas hasta ahora en esta epilepsia. Nuestros estudios han representado un importante avance en la comprensión de las proteínas de tipo PipY y de la epilepsia asociada a la forma humana de las mismas.

Summary

In the context of research of our laboratory on nitrogen signaling mainly in the cyanobacterium *Synechococcus elongatus* PCC7942, with efforts focused on the PII and PipX proteins and their signaling network, this Thesis extends the spectrum of molecules investigated in relation to such network.

It studies and characterizes the non-canonical member of the PII protein superfamily named CutA, a highly conserved protein in all domains of life (including animals and humans) which is generally annotated as protecting against divalent metals. We examine the possible protection provided by CutA against metals, using knockouts for the CutA-encoding gene of two phylogenetically distant bacteria, *Escherichia coli* and *Synechococcus elongatus* PCC7942. Neither complementation studies in *E. coli* by the wild-type gene, nor observational studies of sensitivity to metals in the *S. elongatus* knockout have supported the function annotated for CutA, although we show by turbidimetric monitoring that Cu^{2+} causes aggregation of pure *S. elongatus* CutA (produced recombinantly) and therefore binds to it, although with a low affinity by comparison with the concentrations of this metal that are toxic for this cyanobacterium.

Aiming at getting further insight into CutA, we have determined at very high resolution, by X-ray diffraction, the structure of this protein of *S. elongatus*, failing to observe Cu^{2+} bound in this structure, but showing that the three pockets formed at intersubunit boundaries in the CutA homotrimer are capable of transporting organic molecules (in our case Bis-Tris) without inducing conformational changes in the protein. This finding supports a possible function of CutA based on the binding to these pockets of neutral or positively charged biomolecules capable of forming several hydrogen bonds with the pocket walls, which are endowed with negative potential and have a strong polar character.

We have also studied the PipY protein of *S. elongatus*, recently identified as a functional partner of the aforementioned PipX, determining its spectroscopic properties, binding of pyridoxal phosphate (PLP) and solving its structure by X-ray diffraction. We prove that PipY is monomeric and has PLP attached. Its structure does not favor an enzymatic role of PipY, being more appropriate for exerting a possible function in the homeostasis of PLP. Given the very recent description of a human vitamin B₆-dependent genetic epilepsy associated to mutations in the human orthologue of the *pipY* gene, *PROSC* (now called *PLPBP*, encoding the PROSC protein, now named PLPHP), we used first *S. elongatus* PipY and afterwards and more extensively human PROSC to test by site-directed mutagenesis the pathogenicity of the mutations that have been associated with this epilepsy. These studies have demonstrated the pathogenicity and established mechanisms for this pathogenicity for each of the missense mutations reported thus far in patients with PROSC-associated epilepsy. Our studies represent an important advance in the understanding of PipY-like proteins and of epilepsy associated with the human form thereof.

Resum

Emmarcada en la línia d'investigació del nostre laboratori de senyalització per nitrogen principalment en el cianobacteri *Synechococcus elongatus* PCC7942, amb esforços centrats en les proteïnes PII i PipX i la seua xarxa de senyalització, esta Tesi amplia l'espectre de molècules investigades en relació amb la dita xarxa.

Estudia i caracteritza el membre no canònic de la superfamília de la proteïna PII denominat CutA, generalment anotat com de protecció a metalls divalents, altament conservat en tots els dominis de la vida (inclosos animals i l'ésser humà). En ella examinem la possible protecció front a metalls proveïda per CutA en dos bacteris molt distants, *Escherichia coli* i *Synechococcus elongatus*, usant *knockouts* del gen que codifica CutA per a ambdues. Ni els estudis de complementació en *E. coli* del gen silvestre ni els observacionals de sensibilitat a metalls en *S. elongatus* han donat suport a la funció anotada per CutA, tot i que vam demostrar mitjançant seguiment turbidimètric que el Cu^{2+} fa agregar a la proteïna CutA pura de *S. elongatus* (produïda de forma recombinant) i per tant s'uneix a ella, encara que amb una afinitat baixa per comparació amb les concentracions tòxiques d'aquest metall per a aquest cianobacteri.

Buscant aprofundir en el coneixement de CutA, hem determinat a molt alta resolució mitjançant difracció de raigs X l'estructura d'aquesta proteïna de *S. elongatus*, sense evidenciar cap complex amb coure, però demostrant que les tres cavitats formades entre les subunitats del homotrimer de CutA són capaços de transportar molècules orgàniques (en el nostre cas Bis-Tris). Aquests resultats donen suport a una possible funció de CutA basada en la unió a aquestes cavitats de biomolècules neutres o positivament carregades i capaços de formar diversos ponts d'hidrogen amb les parets de potencial negatiu i fort caràcter polar d'aquestes cavitats.

També hem estudiat la proteïna PipY de *S. elongatus*, identificada recentment com a parella funcional de l'abans esmentada PipX, determinant les seves propietats espectroscòpiques, unió de piridoxal fosfat (PLP) i resolent la seva estructura mitjançant difracció de raigs X. Vam provar que PipY és monomèrica i que té PLP unit. La seva estructura no recolza que sigui un enzim, sent aparentment apropiada per a exercir una possible funció en l'homeòstasi de PLP. Atès que molt recentment s'ha descrit una epilèpsia genètica humana dependent de vitamina B₆ deguda a mutacions en el gen humà ortòleg de *pipY*, *PROSC* (ara anomenat *PLPBP*; codifica la proteïna PLPHP), fem servir inicialment PipY de *S. elongatus* i després *PROSC* humana com banc de proves de la patogenicitat de les mutacions que s'han associat a aquesta epilèpsia, utilitzant mutagènesi dirigida i produint les formes silvestre i mutants d'aquestes proteïnes, comparant les seves propietats. Aquests estudis han demostrat la patogenicitat i establert mecanismes per a la mateixa per a cadascuna de les mutacions de canvi de sentit de *PROSC* descrites fins ara en aquesta epilèpsia. Els nostres estudis han representat un important avanç en la comprensió de les proteïnes de tipus PipY i de l'epilèpsia associada a la forma humana de les mateixes.

ABREVIATURAS NO ESTÁNDAR

2-OG: 2-oxoglutarato o α -cetoglutarato

AAK: aminoácido quinasa

ACT: ion acetato

APP: proteína precursora de la sustancia amiloide

BACE1: β -secretasa 1

BSA: albúmina sérica bovina

BTB: Bis-Tris o 2-[bis-(2-hydroxietil)-amino]-2-hidroximetil-propano-1,3-diol

CD: dicroísmo circular

CSD: 3-sulfinoalanina

CPSI: carbamil fosfato sintetasa I

DCS: D-cicloserina

dNTPs: trifosfatos de desoxirribonucleósidos

DSC: calorimetría diferencial de barrido

DTT: ditioneitol o 1,4-ditio-D-treitol

EDO: etilenglicol (etanodiol)

EDTA: ácido etilendiaminotetraacético

FPLC: cromatografía líquida de proteínas de alta velocidad

GABA: ácido γ -aminobutírico

GOGAT: glutamina/2-oxoglutarato aminotransferasa, o glutamato sintetasa

GOL: glicerol

GS: glutamina sintetasa

IPTG: isopropil- β -D-tiogalactopiranosido

MES: ácido 2-(N-morfolino)-etanosulfónico

NAGK: N-acetilglutamato quinasa

NtcA: regulador global de nitrógeno de cianobacterias

ODC: ornitina descarboxilasa

P5C: Δ^1 -pirrolin-5 carboxilato

P6C: L- Δ^1 -piperideina-6-carboxilato

PAGE: electroforesis en gel de poliacrilamida

PCR: reacción en cadena de la polimerasa

PDB: banco de datos de proteínas (*Protein Data Bank*:
<https://www.rcsb.org/pdb/home/home.do>)

PEG: polietilén glicol

P11-NG: nuevo grupo de proteínas P11

PipX: proteína de interacción con P11

PL: piridoxal

PLP: piridoxal 5'-fosfato

PLPBP: gen codificante de la proteína PLPHP

PLPHP: proteína de homeostasis de PLP

PM: piridoxamina

PMP: piridoxamina 5'-fosfato

PN: piridoxina

PNP: piridoxina 5'-fosfato

PNPO: piridox(am)ina 5'-fosfato oxidasa

PROSC: nombre previo de proteína PLPHP

r.m.s.d.: desviación cuadrática media

SDS: dodecilsulfato sódico

Tris: Tris-aminometano

ÍNDICE

Sección 1. Introducción general y objetivos -----	1
A. La temática de esta tesis -----	3
B. CutA -----	10
B.1. Regulación del metabolismo de nitrógeno y proteína PII -----	10
B.2. De PII a la superfamilia PII -----	13
B.3. Identificación inicial del gen codificante de la proteína CutA -----	17
B.4. La proteína CutA -----	19
B.5. CutA y metales -----	25
C. El gen <i>COG0325</i> , sus productos en cianobacterias y en el ser humano, y su implicación en patología humana -----	30
C.1. El marco de referencia para nuestros estudios con la proteína PipY -----	30
C.2. PipY, miembro cianobacteriano de la familia de proteínas COG0325 -----	31
C.3. Por qué nos hemos interesado en la proteína COG0325 humana -----	36
C.4. Epilepsia dependiente de vitamina B ₆ y mutaciones en el gen <i>PLPBP</i> -----	37
D. Estructura de esta Memoria de Tesis-----	45
E. Objetivos -----	47
Sección 2. Publicaciones -----	49
Capítulo 1: Studies on cyanobacterial protein PipY shed light on structure, potential functions and vitamin B ₆ -dependent epilepsy -----	51
Capítulo 2: Insight into vitamin B ₆ -dependent epilepsy due to PLPBP (previously PROSC) missense mutations -----	85
Capítulo 3: Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria -----	117

Sección 3. Discusión general -----	167
A. Proteínas de la familia COG0325: PipY y PLPHP -----	169
A.1. Antigüedad y distribución de la familia COG0325 -----	169
A.2. Principales diferencias entre PipY y PLPHP -----	171
A.3. Dificultades para la cristalización de PLPHP: diferencias estructurales entre PipY y PLPHP -----	172
A.4. Epilepsia dependiente de vitamina B ₆ provocada por mutaciones en <i>PLPBP</i> -----	175
B. CutA, miembro no canónico de la superfamilia PII -----	176
B.1. Anotación incorrecta: CutA no parece implicada en los mecanismos de tolerancia a cobre -----	176
B.2. La proteína CutA une cobre -----	178
B.3. Los bolsillos intersubunidades como elementos con potencialidad funcional -----	180
B.4. Consideraciones adicionales sobre el papel de la estructura de CutA en la elevada estabilidad térmica de la proteína -----	185
C. Nota final -----	188
 Sección 4. Conclusiones -----	 191
 Sección 5. Referencias correspondientes a las Secciones 1 y 3 -----	 197

SECCIÓN 1

Introducción general y objetivos

Sección 1. INTRODUCCIÓN GENERAL Y OBJETIVOS

A. La temática de esta tesis.

Este trabajo de tesis se ocupa de dos proteínas, los productos respectivos de los genes *COG1324* y *COG0325*, CutA y PipY/PLPHP (previamente PROSC), los cuales, de acuerdo con la base de datos EGGNOG (<http://eggnogdb.embl.de/#/app/results>), son de distribución universal (“*All organisms*”), anotándolos, respectivamente, como “***Inorganic ion transport and metabolism. Tolerance protein***” y “***Function unknown. Alanine racemase domain protein***”. Aún a riesgo de duplicaciones, explicaré en primer lugar por qué me he ocupado de estas dos proteínas y qué las conecta en el contexto de esta tesis.

El interés clásico del grupo de mi director, V. Rubio, se centra en la síntesis de arginina y urea y en sus patologías. Sin embargo, los intentos de su grupo para determinar la estructura de una proteína clave del ciclo de la urea, la enzima carbamil fosfato sintetasa I (CPSI), les llevaron a determinar la estructura de una enzima más simple, la carbamato quinasa (Marina et al., 1998), enzima que se había propuesto (equivocadamente, ver de Cima et al., 2015) como análogo estructural más simple de la CPS. La estructura de la carbamato quinasa desveló un nuevo plegamiento, el *fold* aminoácido quinasa (AAK) (Pfam: PF00696, InterPro: IPR001048) en cuya exploración y caracterización de nuevos miembros el grupo se volcó en la primera década del siglo XXI.

En ese proceso determinaron estructuras de las N-acetil-L-glutamato quinatas (NAGK) no retroinhibibles (Ramón-Maiques et al., 2002) y retroinhibibles por arginina (Ramón-Maiques et al., 2006). En 2004 Asunción

Contreras (Alicante) y Karl Forchhammer (Giessen, y más tarde Tübingen) y sus grupos descubrieron independientemente que la NAGK retroinhibible por arginina es en cianobacterias y plantas una diana de PII (Burillo et al., 2004; Heinrich et al., 2004). PII es (Forcada-Nadal et al., 2018) una proteína señalizadora muy antigua, de muy amplia difusión, presente en todos los dominios de la vida (aunque no está presente en animales), siendo descrita inicialmente, en los años 60 del siglo XX, como sensor clave de riqueza de nitrógeno y controladora de la modificación covalente de la glutamina sintetasa y por tanto de la regulación de esta importante enzima de acumulación de nitrógeno en microorganismos y plantas (revisado en Stadtman, 2001). Para tratar de entender cómo PII controla a sus dianas, el laboratorio del Dr. Rubio, incitado por los de Contreras y Forchhammer, determinó la estructura del complejo de PII con NAGK, ambas de la cianobacteria *Synechococcus elongatus* PCC7942 (en adelante *S. elongatus*) aclarando cómo es ese complejo (Fig. 1A) y por qué la unión de PII activa a la NAGK de organismos fotosintéticos (Llácer et al., 2007), y explicando el impacto de esa función en la acumulación de nitrógeno como arginina de los organismos fotosintéticos (Llácer et al., 2008).

La estructura del complejo PII-NAGK fue la segunda determinada de un complejo de PII con una de sus dianas, pocos meses tras la determinación estructural del primer complejo de PII, con el canal de amonio AmtB de *Escherichia coli* (Gruswitz et al., 2007; Conroy et al., 2007). Tres años después el grupo de V. Rubio, también en colaboración con los de Contreras y Forchhammer, arrojó la primera luz estructural sobre cómo puede controlar PII la actividad génica (Llácer et al., 2010). Para ello usó también proteínas de *S. elongatus*, determinando la estructura del complejo de PII con una proteína adaptadora de la que se ignoraba todo, PipX (Fig. 1B), proteína que podía interactuar bien con PII, bien con el regulador transcripcional general de

nitrógeno, NtcA, activando a este último (Espinosa et al., 2006). También determinó las estructuras de NtcA solo y en complejo con PipX (Llácer et al., 2010) y, más recientemente (aún no publicado) en complejo con el DNA y con PipX, así como de PipX aislada (no asociada a otras proteínas) (Forcada-Nadal et al., 2017).

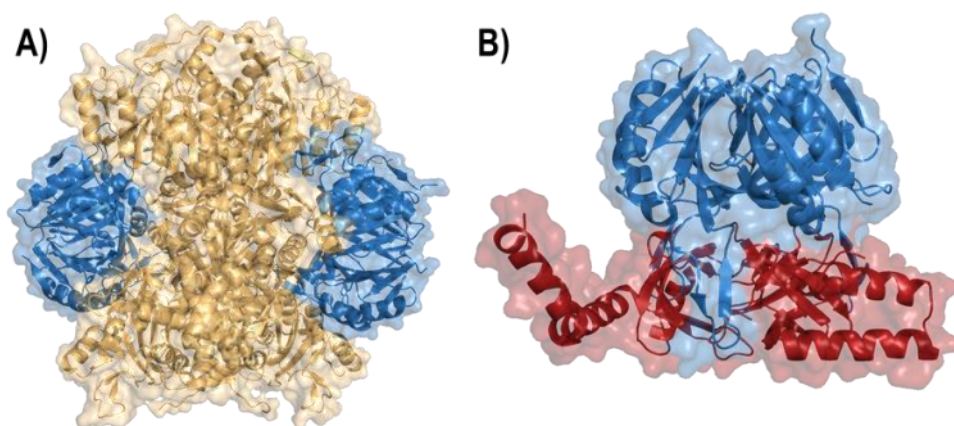


Figura 1. Representación en superficie semitransparente y cintas de los complejos determinados por nuestro laboratorio entre PII de *S. elongatus* y sus proteínas diana: A) PII-NAGK, azul y dorado, respectivamente. PDB:2V5H (Llácer et al., 2007) y B) PII-PipX, azul y granate, respectivamente, PDB: 2XG8 (Llácer et al., 2010).

Estos hitos estructurales han situado a nuestro grupo en la vanguardia de los estudios de señalización por nitrógeno, llevándole a establecer una línea de investigación estable dedicada a PII, PipX y a los diferentes elementos relacionados con los mismos. Esta línea, en la que se encuadra esta tesis, con ya tres tesis doctorales defendidas (J.L. Llácer, 2011; C. Palanca, 2016; A. Forcada-Nadal, 2018), incluye como logros adicionales la determinación de las primeras estructuras de una PII halófila (en colaboración con el grupo de la Dra. M^a José Bonete, Universidad de Alicante; Palanca et al., 2014), de la forma uridililada de GlnB (uno de los dos parálogos de PII de *E. coli*) (Palanca y Rubio, 2017), y del

regulador transcripcional AmtR, diana de PII en *Corynebacterium glutamicum* (Palanca y Rubio, 2016). También se ha conseguido la caracterización funcional (en colaboración con K. Forchhammer) de la influencia de PipX en la unión de NtcA a DNA (Forcada-Nadal et al., 2014); y, en colaboración con A. Contreras, la demostración y modelización grosera de un complejo ternario de PII, PipX y PlmA (Labella et al., 2016; soy coautora de este trabajo, que no se ha incluido en esta Tesis Doctoral), siendo PlmA un regulador transcripcional de cianobacterias; y la demostración (esta vez como colaboradores de A. Contreras) de que PipX establece interacciones funcionales con la proteína de unión a piridoxal fosfato (PLP) PipY (Labella et al., 2017), proteína objeto de uno de los capítulos de esta tesis doctoral.

Así, mi trabajo se encuadra en la línea del grupo que se centra en el estudio de PII y de su red de señalización con PipX, resumida para *S. elongatus* en la Fig. 2, y en la que se ilustran dos nuevas dimensiones simbolizadas por las dos flechas sólidas verticales que emergen del plano de la red a nivel de las proteínas PII y PipX.

Una de esas flechas señala a la proteína CutA (Fig. 2), el producto del gen *COG1324*. Se trata de una proteína que los estudios filogenéticos finos y las comparaciones estructurales han establecido como miembro de la superfamilia de PII (Kinch y Grishin, 2002), si bien las comparaciones de secuencia no detectan identidad significativa con PII clásica (ver más abajo en esta Introducción). Dada esta falta de reconocimiento por identidad de secuencia, y puesto que a CutA le falta un elemento clave para las interacciones de PII con sus dianas, el largo y flexible lazo T (ver secciones posteriores), a esta proteína se la considera un miembro no canónico de la superfamilia de PII (Forchhammer y Lüddecke, 2016).

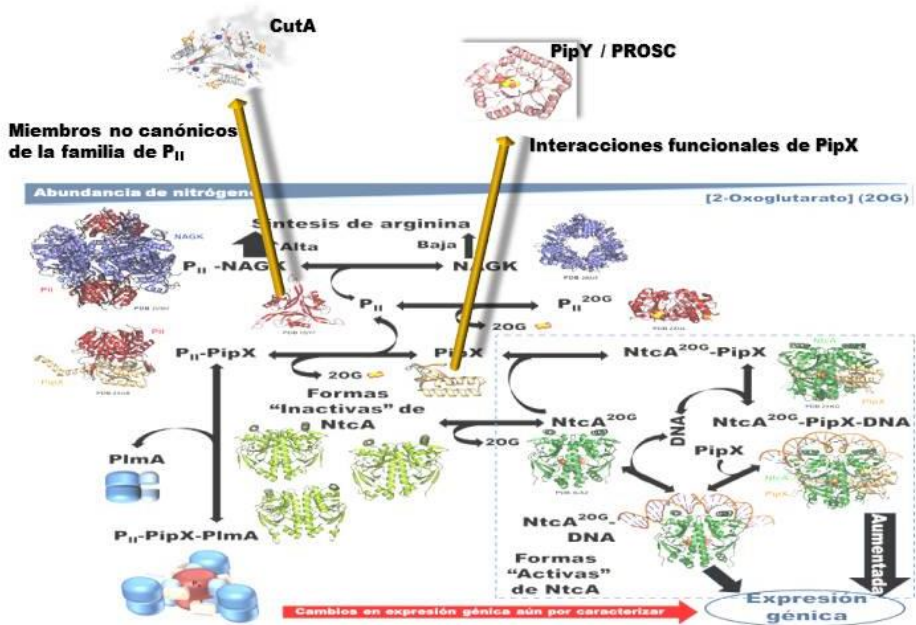


Figura 2. Esquema de la red de señalización mediada por PII-PipX en *S. elongatus* tal como la presenta en su tesis doctoral A. Forcada Nadal. Sobre este esquema se han superpuesto, saliendo hacia arriba desde el plano de la red, dos flechas. Una de ellas simboliza la pertenencia de PII a su superfamilia y su relación, por tanto, con la proteína no canónica de esta superfamilia llamada CutA. La otra flecha simboliza la existencia de interacciones funcionales de PipX con la proteína PipY.

Cuando empecé en el laboratorio, éste trataba de extender sus estudios a los miembros no canónicos de la familia de PII, concentrándose en CutA por su presencia universal en todos los filos, incluyendo los mamíferos y el ser humano. La pregunta básica sobre dicha proteína era cuál era su función, pregunta que venía asociada a la de por qué CutA tiene difusión universal desde bacterias al ser humano. Ya he mencionado la anotación de la misma en la base de datos EggNOG como ***"Inorganic ion transport and metabolism. Tolerance protein"***. Es más explícita en este contexto la base de datos UniProtKB (<https://www.uniprot.org>), que denomina a las dos formas de CutA

bacteriana usadas en nuestros estudios (de *E. coli* y *S. elongatus*; UniProtKB P69488 y Q31KX8, respectivamente) “Divalent-cation tolerance protein CutA” y “Periplasmic divalent cation tolerance protein”, respectivamente, con las siguientes anotaciones funcionales: “Involved in resistance toward heavy metals” y “Response to metal ions”, respectivamente. Sin embargo, como se detalla luego, las evidencias a favor de esa función de resistencia a metales nos parecieron poco robustas, así que desde el principio de mi trabajo un objetivo central fue someter a prueba la posible implicación de CutA en la resistencia a metales en el organismo en el que se había propuesto tal función, *E. coli*, aunque adicionalmente tratamos de extender las evidencias a *S. elongatus*, que siendo una cianobacteria está lejana filogenéticamente de *E. coli*. Además, como bien estructuralmente bien mediante evidencias indirectas (ver más adelante) se habían publicado propuestas sobre la ubicación de los sitios de unión de Cu^{2+} en CutA de varias especies (incluida la humana) tratamos de someter a prueba la universalidad de estos sitios determinando la estructura de CutA de una estirpe bacteriana (*S. elongatus*) distinta a las investigadas previamente por otros. Igualmente, espoleados por un provocador artículo firmado por el premio Nobel James Watson (Laskowski et al., 2003) sobre la posible definición de las funciones de CutA por su capacidad de carga (en el sentido de cargamento, no de carga eléctrica) de sus tres bolsillos intersubunidades, hemos explotado la alta resolución de nuestras estructuras para demostrar que de hecho dichos bolsillos pueden cargarse con compuestos orgánicos, aunque el detectado por nosotros es un tampón orgánico no natural. Añadiré también en cuanto al interés de CutA, su excepcional termoestabilidad, que para la proteína del mesófilo *E. coli* tiene una temperatura de desnaturalización (establecida mediante calorimetría diferencial de barrido, DSC) de aproximadamente 90°C (Matsuura et al., 2010).

La segunda flecha sólida que emerge fuera del plano desde PipX en la Fig. 2 señala a la proteína PipY de *S. elongatus* (PLPHP en el ser humano, antes llamada PROSC), el producto del gen *COG0325* señalado en EggNOG como de función desconocida, y para la que los estudios (en colaboración con nuestro grupo) del equipo de la Dra. Contreras (Labella et al., 2017) indicaron la existencia de interacción funcional en *S. elongatus* con PipX. Dicha interacción se manifestó en las influencias recíprocas detectadas mediante modificación genética de *S. elongatus*, pero sin que se encontrara una interacción física directa entre ambas proteínas. Mi investigación sobre la función y estructura de la proteína PipY de *S. elongatus* constituye el Capítulo 1 de la Sección 2 de este trabajo de tesis. Además, en línea con los intereses de nuestro grupo en patología genético-molecular humana, la descripción de una forma de epilepsia genética dependiente de vitamina B₆ asociada a mutaciones en el gen *COG0325* humano (Darin et al., 2016) propició nuestro uso inmediato de la proteína PipY de *S. elongatus* como sistema modelo para estudiar el efecto patogénico de las mutaciones de cambio de sentido asociadas a la enfermedad. Seguidamente extendimos y profundizamos estas investigaciones (Capítulo 2 de la Sección 2 de esta Tesis) al estudio de la proteína correspondiente humana, entonces denominada PROSC, y hoy denominada PLPHP como acrónimo del inglés *pyridoxal phosphate homeostasis protein*, por la inferencia de que su función tiene que ver con la homeostasis para PLP, si bien tal homeostasis no ha sido concretada molecularmente.

En resumen, y cierro aquí este anteproyecto, me he ocupado de dos proteínas que tienen en común el estar relacionadas con el sistema regulador de PII-PipX: CutA, producto de un lejano gen parálogo del gen codificante de PII; y PipY, interactuante funcional de PipX. Para ambas proteínas la función que desempeña cada una dista de haber sido aclarada en términos

moleculares. De ambas se conocen sus estructuras, y las dos tienen ortólogos humanos, de los que el de PipY se asocia a una patología neurológica genética. Estos son los componentes con los que he elaborado la presente tesis.

B. CutA

B.1. Regulación del metabolismo de nitrógeno y proteína PII.

El nitrógeno es un elemento esencial para los seres vivos (Alberts et al., 2016), del que, para su utilización biológica, hay que convertir la forma abundante, pero inerte, de la atmósfera, N_2 (Lee et al., 2014), así como el nitrato y nitrito del suelo, en amonio (NH_4^+), utilizable por todos los organismos (Burris, 1991; Hoffman et al., 2014; Nelson y Cox, 2005) (Fig. 3). Bacterias y plantas usan generalmente el NH_4^+ mediante el sistema GS/GOGAT, con dos reacciones consecutivas que involucran a la glutamina, el glutamato y el compuesto intermediario del ciclo de Krebs 2-oxoglutarato (2-OG) (Fig. 3), produciendo una molécula de glutamato por amonio utilizado, glutamato que proveerá el nitrógeno para la síntesis de otros aminoácidos (Merrick y Edwards, 1995; Reitzer, 2003; Nelson y Cox, 2005) (Fig. 3).

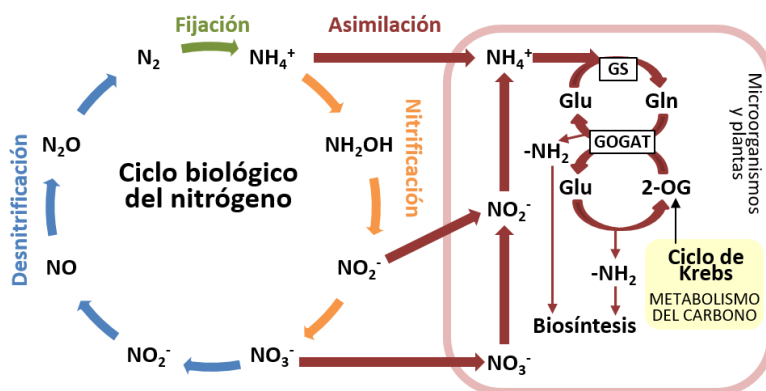


Figura 3. Principales formas de nitrógeno utilizadas por los seres vivos y resumen del ciclo y de la incorporación biológica de nitrógeno (Richardson, 2001; Ohashi et al., 2011).

El metabolismo y el uso de nitrógeno están fuertemente regulados. Un punto clave de esta regulación es la enzima glutamina sintetasa (GS), para la que se ha descrito una compleja cascada regulatoria que incluye elementos de

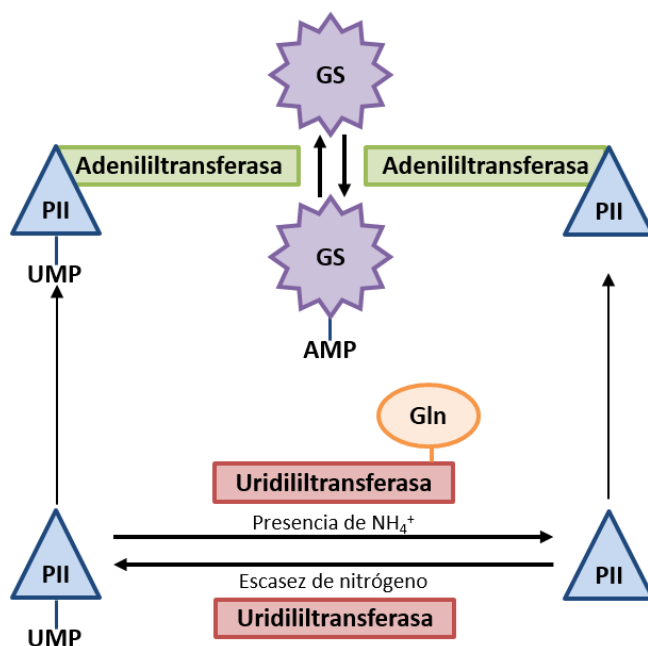


Figura 4. Esquema muy simplificado de la regulación de la actividad GS de *E. coli* por la proteína PII.

control *feed-back* y otros de modificación covalente (Fig. 4). En el estudio de esa cascada se descubrió la importante proteína reguladora PII (revisado en Stadtman, 2001), proteína que controla una enzima bifuncional (ATasa; Jiang et al., 2007) que adenilila/desadenilila la GS, y la cual, a su vez, está regulada por uridililación/desuridililación de su residuo Tyr51 por una uridililtransferasa bifuncional (Fig 4).

Hoy sabemos (revisado en Forchhammer y Lüddecke, 2016) que las proteínas PII constituyen una amplia familia que incluye ortólogos y parálogos que han sido clasificados en formas canónicas y no canónicas según identidades de secuencia, y según la presencia o ausencia del largo y flexible lazo T que en la proteína PII cumple misiones de interacción y transducción de señales. Las

formas canónicas se distribuyen muy ampliamente en los tres dominios de la vida, aunque sin estar presentes en el reino animal. Son importantes sensores del estado de abundancia de carbono, nitrógeno y energía, siendo elementos clave en la señalización por nitrógeno. Este papel regulador está mediado por la unión a PII de los efectores alostéricos 2-OG y ATP y, excepto en plantas, por la unión de ADP (Forchhammer y Lüddecke, 2016), y en algunos organismos (Palanca et al., 2014) incluso por la unión de AMP. Además, en bastantes bacterias las proteínas PII se regulan también por modificación postraduccional de sus flexibles lazos T como ya hemos señalado para PII de *E. coli*, que es uridililable o adenililable en proteobacterias y actinobacterias (Merrick, 2015) o fosforilable en Ser49 en algunas cianobacterias (Forchhammer y Tandeau de Marsac, 1995).

Las acciones de PII se ejercen a través de interacciones con proteínas diana que pueden ser canales de membrana (Thomas et al., 2000; Conroy et al., 2007; Gruswitz et al., 2007), enzimas (Jiang et al., 2007; Burillo et al., 2004; Heinrich et al., 2004), o sistemas reguladores de la expresión génica, como el sistema mediado por la proteína PipX, co-activador del factor de transcripción NtcA en cianobacterias (Espinosa et al., 2006). La proteína PII actúa a la vez como sensor del estado de riqueza en nitrógeno/carbono/energía mediante la unión de sus efectores alostéricos (Forchhammer y Lüddecke, 2016), y como transductor de señales, funciones ejercidas porque los efectores unidos, dependiendo de cuales sean, favorecen una conformación u otra del lazo T y así facilitan o impiden la unión con una u otra proteína diana (Llácer et al., 2007; Llácer et al., 2010). No revisaré aquí por brevedad cómo ejerce su función la modificación covalente de PII.

B.2. De PII a la superfamilia PII

Desde que se descubrió que *E. coli* tiene dos genes parálogos para proteínas PII (*glnB* y *glnK*) fue patente que las proteínas PII constituían una familia de al menos estos dos elementos (Xu et al., 1998; Thomas et al., 2000). Con la secuenciación de genomas se definieron tres subtipos de proteínas PII (Arcondéguy et al., 2001): *i*) GlnB, que controla a la GS, con cuyo gen generalmente forma un operón, coexpresándose con él, (Arcondéguy et al., 2001), *ii*) GlnK, cuyo gen se asocia al del canal de amonio AmtB al que controla (Radchenko et al., 2014), y *iii*) Nifl, cuyo gen se asocia al operón multicistrónico de la nitrogenasa en arqueas y bacterias fijadoras de N₂ (Arcondéguy et al., 2001). Esta familia se amplió más recientemente (Sant'Anna et al., 2009) con proteínas con identidad de secuencia reconocible y significativa con los tipos anteriores, aunque no encuadrables en los mismos. Para estos genes y proteínas se generó el nuevo grupo PII-NG (Fig. 5). Los genes para el grupo PII-NG con frecuencia se localizan junto a genes para transportadores de metales pesados (Sant'Anna et al., 2009).

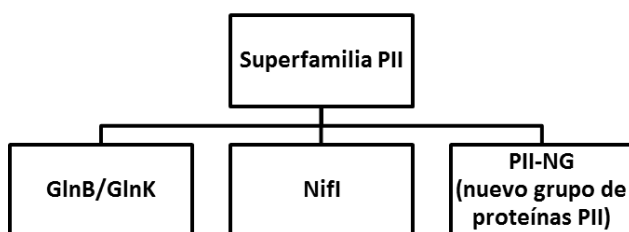


Figura 5. Clasificación de la superfamilia PII propuesta por Sant'Anna et al. (2009)

La familia ampliada de PII forma parte de una superfamilia más extensa (Kinch y Grishin, 2002) construida tras un rastreo centrado en las proteínas canónicas de tipo PII que hizo uso transitivo de Blast-P. La superfamilia se compone de 7 grupos, de los que un grupo central al que denominaron grupo II

incluye secuencias de la familia clásica PII. Las búsquedas encadenadas mediante Blast-P relacionaron los grupos IV y Va con la familia PII clásica vía similitudes con los grupos intermedios III y V, respectivamente (Fig. 6) (Kinch y Grishin, 2002).

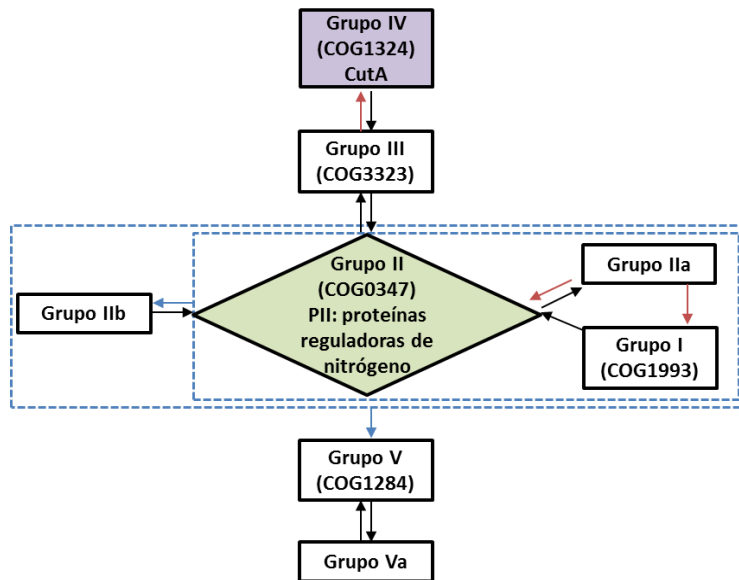


Figura 6. Superfamilia PII, de acuerdo con Kinch y Grishin (2002). Las flechas negras indican que la detección del grupo se ha producido por Blast usando en la búsqueda las proteínas del grupo del que procede la flecha. Las flechas azules indican que la detección del grupo se ha producido por Blast y alineamiento de los grupos incluidos en el recuadro azul de línea discontinua. Las flechas granate indican que el grupo se ha detectado por Blast de secuencias individuales del grupo de procedencia e la flecha. Los identificadores COG se refieren a grupos de genes ortólogos de diferentes organismos que pueden encontrarse en la base de datos EggNOG (<http://eggnogdb.embl.de/#/app/results>). Modificado de (Kinch y Grishin 2002) con autorización de Wiley.

Es de particular interés que para la identificación del grupo IV se utilizaron tanto criterios de similitud de secuencia (que dieron puntuaciones débiles de similitud con el grupo III y solo para la mitad C-terminal de las proteínas de los grupos III y IV) como la predicción de *fold* estructural, que arrojó para uno de los miembros del grupo IV la máxima nota para el *fold* ferredoxina de la

proteína *spliceosomal* U1A [archivo del Protein Data Bank (PDB) 1URN, cadena A]. Dado que esta proteína exhibe con pequeñas diferencias el mismo plegamiento que PII [patrón de estructura secundaria de tipo ferredoxina, $(\beta\alpha\beta)_2$], este hallazgo dio mayor robustez a la adscripción del grupo IV a la superfamilia de PII. Este grupo IV es de particular interés para nosotros porque lo constituyen las proteínas CutA, objeto del Capítulo 3 de la Sección 2 de esta Tesis.

En el detallado trabajo de Kinch y Grishin (2002) se concluyó que los diferentes miembros de la superfamilia deberían ser homotrímeros con arquitecturas similares a PII, cuya estructura se había determinado ya por entonces (Cheah et al., 1994). También se predijo que en base a la comparación de secuencias era posible que mantuvieran la cavidad utilizada por PII para la unión de un nucleótido de adenina y de 2-oxoglutarato/ATP. Ambas conclusiones fueron corroboradas para la proteína CutA cuando se determinaron las estructuras de CutA de *E.coli* y rata (Arnesano et al., 2003).

Los estudios de conservación de secuencia de Kinch y Grishin (2002) también aportaron evidencias en contra de que el ATP fuera un ligando de las proteínas de la superfamilia que no pertenecían al grupo II. En resumen, se concluía que el sitio de unión de ligando parecía estar conservado en los trímeros de la superfamilia, pero que la especificidad del sitio debería variar. También señalaron en el caso de los miembros de los grupos III y IV que el sitio potencial de unión de ligando exhibía residuos cargados e invariantes en esos grupos. Igualmente, identificaron para estos grupos residuos conservados en el canal central del trímero, lo que les llevó a plantear la posibilidad de que las proteínas del grupo IV tuvieran alguna función de canal (Kinch y Grishin, 2002).

Más recientemente, Forchhammer y Lüddecke (2016) han sometido a análisis la familia de proteínas PII mediante comparación de las estructuras de

proteínas de esta familia depositadas en la base de datos PDB. En su estudio, la superfamilia comprende dos grupos, uno (Fig. 7, Proteínas PII) de proteínas que presentan tanto identidad estructural y carácter homotrimérico como identidad de secuencia significativa con proteínas PII genuinas (sin aplicación de un procedimiento transitivo a través de parálogos intermedios como hicieron Kinch y Grishin) y otro (Fig. 7, Proteínas con plegamiento tipo PII) que incluye proteínas con similitud estructural pero sin identidad significativa de secuencia con PII clásica. El primer grupo incluye proteínas de los subtipos GlnB, GlnK y Nifl. El segundo incluye las proteínas de tipo CutA que, como ya se ha dicho, son objeto de una parte de este trabajo de tesis. Este grupo también incluye proteínas como PstA o DarA (Gundlach et al., 2015), que unen AMP cíclico, así como SbtB (Forchhammer y Lüddecke, 2016), un posible componente regulador de un transportador de bicarbonato dependiente de sodio (Shibata et al., 2002).

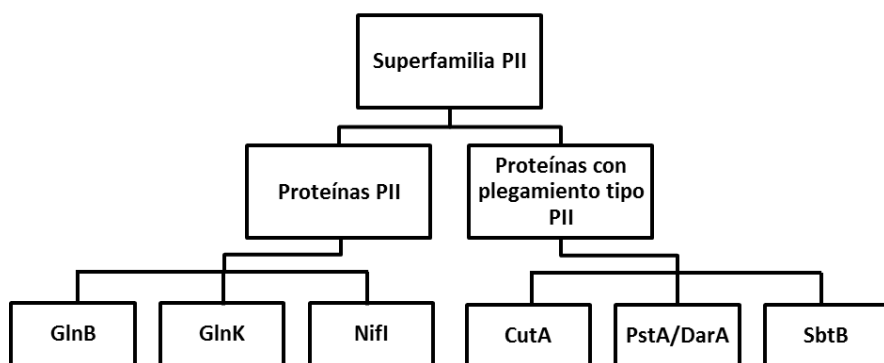


Figura 7. Clasificación de los miembros de la superfamilia PII propuesta por Forchhammer y Lüddecke (2016).

Todas estas proteínas comparten el ser homotriméricas de una subunidad de 100-170 aminoácidos [con excepciones representadas por proteínas del grupo III de Kinch y Grishin (2002), que son proteínas multidominio uno de cuyos

dominios se corresponde con una proteína de esta superfamilia], en las que cada polipéptido presenta un plegamiento tipo ferredoxina ($(\beta\alpha\beta)_2$), ensamblándose el trímero por continuidad de sus hojas beta, que conforma una hoja β molecular cerrada de carácter tronco-cónico, con un poro-túnel central, mientras que las hélices α recubren externamente esa hoja β cerrada (Forchhammer y Lüddecke, 2016).

En resumen, tanto la comparación cuidadosa de secuencias aumentada con la predicción estructural (Kinch y Grishin, 2002) como el análisis primariamente estructural (Forchhammer y Ludeckke, 2016) confluyen en establecer que CutA es un miembro de la superfamilia PII. A continuación, pasamos a introducir los conocimientos existentes sobre la proteína CutA, que en dos de sus formas bacterianas es objeto de parte de esta memoria de Tesis.

B.3. Identificación inicial del gen codificante de la proteína CutA.

El gen que codifica CutA en *E. coli*, *cutA1*, fue identificado originalmente en esta bacteria durante el estudio del locus *cutA*, cuya mutación se había asociado con un aumento de sensibilidad por parte de la bacteria a los efectos tóxicos del cobre (Cu^{2+}) y de los metales pesados, por lo que se propuso que este locus de 3.2 kb localizado en el minuto 94 del genoma de *E. coli* (Fong et al., 1995) podría estar implicado en los mecanismos de transporte de cobre. De hecho, su nombre fue asignado por las siglas en inglés de Cu²⁺ *transport* (Rouch et al., 1989). Se identificaron en este *locus* tres pautas abiertas de lectura consideradas significativas, a las que se denominó *cutA1*, *cutA2* y *cutA3* (Fig. 8). Estudios de complementación para tolerancia a cobre del mutante *cutA*⁻ (cepa carente del locus *cutA*) usando plásmidos que contenían fragmentos de ADN de la zona deletada fueron capaces de mejorar la tolerancia al cobre solo cuando contenían tanto *cutA1* como *cutA2* (Fig. 8, señalados con un signo +)

(Fong et al., 1995). La introducción exclusiva de *cutA1* no aumentó la tolerancia, y en ningún caso se introdujeron *cutA2* o *cutA3* solos, por lo que, aunque estos estudios demostraron claramente la implicación de *cutA2* en la tolerancia al cobre, no aclararon el papel de *cutA1*, a pesar de lo cual se asumió que *cutA1* también estaba implicado en la tolerancia a este metal.

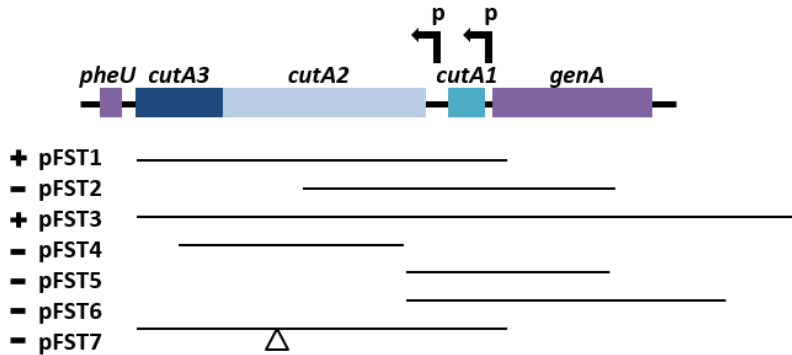


Figura 8. Esquema de Fong et al. (1995) de las pautas de lectura en el locus *cutA* de *E. coli*. Nótese que, aunque en este esquema no se representa, *cutA1* solapa con *cutA2*. Se alinean debajo los fragmentos de DNA utilizados en dicho trabajo en estudios de complementación indicándose sus nombres a la izquierda y el resultado obtenido en la complementación con un signo positivo (+) o negativo (-). El triángulo en pFST7 representa una interrupción del gen con un cassette de resistencia a antibiótico. Reproducido con autorización de Wiley.

De forma simultánea, otro grupo (Crooke y Cole, 1995) secuenció el mismo locus en un estudio de complementación de un mutante de *E. coli* que presentaba afectada la biogénesis de citocromos de tipo C y que es incapaz de usar nitrito como aceptor de electrones durante el crecimiento anaeróbico. Así se identificó como responsable la proteína DipZ de *E. coli*, con un dominio típico de disulfuro isomerasa, enzima que cataliza la formación y ruptura de puentes disulfuro entre residuos de cisteína durante el plegamiento de una proteína. Ese mismo año, otro estudio independiente basado en la búsqueda de genes implicados en la formación de puentes disulfuro identificó el gen *dsbD*,

codificante de la proteína del mismo nombre, que se comprobó actuaba como agente reductor en el periplasma (Missiakas et al., 1995). En ese último trabajo se detectó que la secuencia de *dsbD* era idéntica a las de *dipZ* y *cutA2*, lo que permitió extender la función de *dipZ* a la biosíntesis de otras proteínas adicionales a citocromos de tipo C. Poco después se comprobó la disminución de la tolerancia al cobre de la cepa con la mutación en *dipZ* (Metheringham et al., 1996). Esta implicación de *dsbD* (sinónimos, *dipZ* o *cutA2*) en los mecanismos de tolerancia al cobre se justificó por la importancia de los grupos sulfhidrilo en la resistencia a metales (Brown et al., 1991; Metheringham et al., 1996).

Como ya se ha mencionado, actualmente la proteína codificada por *cutA1*, generalmente denominada CutA, aparece frecuentemente anotada en las bases de datos como proteína de tolerancia a cationes divalentes o como proteína implicada en la resistencia a metales. Sin embargo, como ya se ha indicado y como se insistirá más adelante, la implicación de esta proteína en tal tolerancia o resistencia está aún pendiente de ser demostrada rigurosamente, siendo esta una de las cuestiones abordadas aquí.

B.4. La proteína CutA.

A juzgar por la presencia de su gen codificante, la proteína CutA se encuentra ampliamente distribuida en la naturaleza, estando presente en bacterias, arqueas, plantas, hongos y animales (Perrier et al., 2000). Esta proteína es el miembro no canónico de la superfamilia PII que ha sido más estudiado, a pesar de lo cual su función dista de haber sido esclarecida.

Tabla 1: Estructuras de CutA resueltas y depositadas en el Protein DataBank (PDB)

Organismo	PDB	Año	Resol. (Å)	Ligandos ¹	Notas ²	Ref. ³
<i>Pyrococcus horikoshii</i>	1UKU	2003	1.45	Cu ²⁺	-	1
	1J2V	2003	2.00	-	SeMet	1
	1UMJ	2003	1.60	Guanidina	-	2
	2E66	2006	2.00	Cl ⁻ , Na ⁺	D60A	-
	4NYP	2013	2.00	Na ⁺ , SO ₄ ⁻	CSD	-
	4NYO	2013	1.80	Cl ⁻ , MES, Na ⁺ , SO ₄ ⁻	CSD	3
<i>Archaeoglobus fulgidus</i>	1P1L	2003	2.00	-	-	-
<i>Escherichia coli</i>	1NAQ	2002	1.70	p-hidroximercuri-benzoato	-	4
	3AA9	2009	2.30	-	E61V	5
	3AA8	2009	2.30	-	2 muts.	5
	3AH6	2010	2.40	-	-	5
	3X3U	2015	2.09	-	-	-
	4Y65	2015	1.70	-	3 muts.	-
	4Y6I	2015	1.70	-	4 muts.	-
<i>Thermotoga maritima</i>	1KR4	2002	1.40	-	SeMet	6
	1O5J	2003	1.95	-	SeMet	-
	1VHF	2003	1.54	-	SeMet	7
<i>Thermus thermophilus</i>	1NZA	2003	1.70	Cl ⁻ , Na ⁺ , SO ₄ ⁻ , GOL	-	3
	1V6H	2003	1.90	SO ₄ ⁻	-	3
	4ZK7	2015	3.40	Corismato mutasa	13 muts.	8
<i>Xylella fastidiosa</i>	2NUH	2006	1.39	Imidazol	-	-
<i>Yersinia pestis</i>	3GSD	2009	2.05	Na ⁺ , HEPES, EDO, PEG	-	-
<i>Shewanella sp.</i>	3AHP	2010	2.70	-	-	9
<i>Salmonella enterica</i>	3OPK	2010	1.90	Na ⁺ , Mg ²⁺ , ACT	-	-
<i>Ehrlichia chaffeensis</i>	4IYQ	2013	2.55	Ca ²⁺	-	-
<i>Cryptosporidium parvum</i>	4E98	2012	2.00	Cl ⁻	-	10
<i>Oryza sativa</i>	2ZOM	2008	3.02	SO ₄ ⁻ , GOL	-	-
<i>Rattus norvegicus</i>	1OSC	2003	2.15	-	-	4
<i>Homo sapiens</i>	1XK8	2004	2.70	Na ⁺	-	-
	2ZFH	2006	2.05	-	-	11

Los colores azul, verde y rojo indican si el organismo es Arquea, Bacteria o Eucariota, respectivamente.

¹ MES: ácido 2-(N-morfolino)-etanosulfónico, GOL: glicerol, ACT: ion acetato, EDO: etilenglicol.

² CSD: 3-sulfinoalanina. Muts: mutaciones.

³ 1: Tanaka et al., 2004a. 2: Tanaka et al., 2004b. 3: Bagautdinov 2014. 4: Arnesano et al., 2003. 5: Matsuura et al., 2010. 6: Savchenko et al., 2004. 7: Badger et al., 2005. 8: Bale et al., 2015. 9: Sato et al., 2011. 10: Buchko et al., 2015. 11: Bagautdinov et al., 2008.

El gen *cutA* codifica en bacterias un polipéptido de 112 residuos con una masa molecular de unos 12 kDa. Se han realizado numerosos estudios cristalográficos con esta proteína de diferentes especies (Tabla 1) concluyéndose que forma un homotrímero (Fig. 9) con idéntica topología relativa de las subunidades que en PII canónica. Como en PII, su subunidad (Fig. 9A) presenta un plegamiento tipo ferredoxina ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$), con las hebras β dispuestas de forma antiparalela y cuyas hebras $\beta_2\beta_3$ son muy largas y están conectadas por un lazo, equivalente al lazo T de PII pero mucho más corto (Fig. 9). Al dominio ferredoxina le siguen una hebra y una hélice extra ($\beta_5\alpha_3$) hacia el C-término del polipéptido (Arnesano et al., 2003; Savchenko et al., 2004; Tanaka et al., 2004a). Tres subunidades de esta proteína se unen para producir el homotrímero característico de las proteínas de esta superfamilia al que le faltan los prominentes lazos T de PII (Fig. 9). Como ya se ha mencionado, y aunque no se muestran detalladamente aquí, entre las subunidades se forman cavidades que son accesibles desde el solvente (Arnesano et al., 2003).

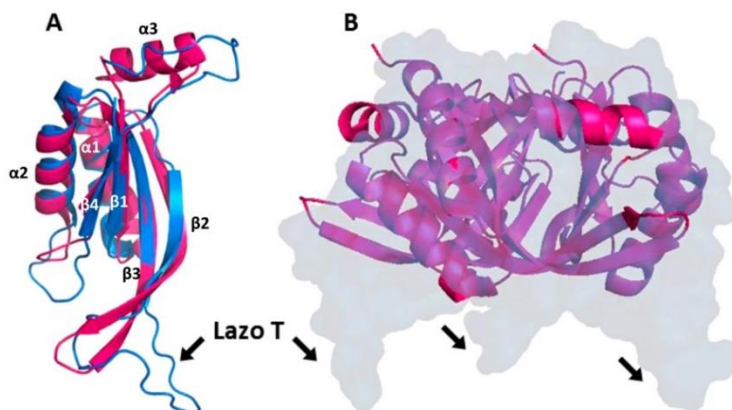


Figura 9. Comparación de las estructura de CutA (PDB: 3AA8) y de GlnB (PII; PDB: 2PII) de *E. coli*. (A) Subunidad de CutA (en magenta) superpuesta a la de GlnB (azul), ambas en representación de cintas. Se señalan los elementos principales de su estructura secundaria. (B) Representación en cintas de CutA superpuesta sobre la superficie (transparente) de PII. Las flechas negras señalan los lazos T de PII.

Una característica particular de CutA es su muy elevada estabilidad térmica, con temperaturas de desnaturalización (irreversible) determinadas por calorimetría diferencial de barrido (DSC), bajo idénticas condiciones, de 90°C para un mesófilo como *E. coli* y de 150°C para el hipertermófilo *Pyrococcus horikoshii* (Matsuura et al., 2010; Tanaka et al., 2006). Estas temperaturas son mucho más elevadas que las óptimas para la vida de estos organismos [37°C para *E. coli*; aproximadamente 98°C para *P. horikoshi* (González et al., 1998)]. Igualmente, las temperaturas de desnaturalización de CutA de *Thermus thermophilus*, del arroz y de humanos se han estimado en aprox. 113, 97 y 95 °C, respectivamente (Bagautdinov et al., 2015; Sawano et al., 2008). Incluso en la bacteria psicrótrófica (es decir, que se reproduce en ambiente frío) *Shewanella* sp. SIB1, CutA no se despliega completamente hasta alcanzar una temperatura de 90°C (Sato et al., 2011). Se ha responsabilizado de esa gran termoestabilidad al gran número de pares iónicos intrasubunidad (Sawano et al., 2008; Matsuura et al., 2012) y a su carácter trimérico (Tanaka et al., 2006; Sawano et al., 2008; Matsuura et al., 2010; Sato et al., 2011; Matsuura et al., 2012), habiéndose incrementado mucho por mutagénesis la termoestabilidad de la forma de *E. coli* (Matsuura et al., 2018). Su gran termoestabilidad ha dado alcance biotecnológico a CutA en el ámbito de los biomateriales (Ito et al., 2010; Guan et al., 2013) o para el diseño de nanoestructuras proteicas autoensamblables (King et al., 2014), incluida la generación de un complejo artificial entre CutA y la enzima corismato dismutasa que ha permitido la cristalización y resolución estructural cristalográfica de dicho complejo, hecho más soluble por mutaciones en CutA (Bale et al., 2015).

Poco sabemos con seguridad sobre la función de CutA en bacterias aparte de lo ya apuntado sobre atribución de funciones no caracterizadas de tolerancia a metales (Burkhead et al., 2003; Savchenko et al., 2004; Tanaka et

al., 2004a; Sato et al., 2011; Bagautdinov, 2014; Buchko et al., 2015). La similitud con PII y su carácter trimérico han sugerido (Arnesano et al., 2003) posibles funciones señalizadoras. Se ha apuntado a favor de esta posibilidad, tanto desde una perspectiva bioinformática (Kinch y Grishin, 2002) como estructural (Arnesano et al., 2003; Laskowski et al., 2003) que las cavidades interdominios de CutA podrían permitir su ocupación con biomoléculas que pudieran actuar como efectores alostéricos (Arnesano et al., 2003). Incluso se ha aventurado a partir de la observación de estas cavidades que las biomoléculas candidatas podrían contener algún anillo, como sería el caso del NADP, ATP o gluconato (Laskowski et al., 2003).

En mamíferos se ha postulado que CutA podría formar un complejo con la acetilcolinesterasa cerebral, pues se observó la presencia de CutA en preparaciones de esta enzima (Navaratnam et al., 2000; Perrier et al., 2000) y también se constató que ambas se expresan en forma paralela en distintas áreas del sistema nervioso (Navaratnam et al., 2000). Se pensó que CutA ancla a la acetilcolinesterasa a la membrana, pues el RNA antisentido para el mensajero de CutA era capaz de abolir la expresión de la acetilcolinesterasa en células de neuroblastoma (Perrier et al., 2000). Sin embargo, tanto la interacción directa como la participación en el anclaje han sido excluidos posteriormente (Perrier et al., 2002).

Se ha relacionado en estudios *in vitro* CutA con proteínas implicadas en enfermedad de Alzheimer. Así, de las tres formas de CutA producidas en ratón y humanos dependiendo del comienzo de la traducción, la más pesada (H; se inicia a -63 residuos respecto a las formas bacterianas) se asocia a membrana en Golgi y trans-Golgi, y se excreta (Zhao et al., 2012), e interacciona por su porción N-terminal de asociación a membrana con la porción N-terminal también de asociación a la membrana de BACE1, proteasa de gran relevancia

en enfermedad de Alzheimer (Zhao et al., 2012). La transfección de células de neuroblastoma con el gen codificante de H-CutA (Hou et al., 2015) aumentó los niveles de Cu^{2+} celulares en forma dosis-dependiente en relación a la cantidad transfectada, mientras que otros experimentos (Hou et al., 2015) revelaron una compleja relación entre H-CutA, niveles de la proteína precursora del amiloide (APP) y de su procesado a βAP , y niveles de Cu^{2+} , sin que este metal influyera en la interacción H-CutA-BACE1. Aun así, no hay un dictamen definitivo sobre el potencial papel de CutA en la enfermedad de Alzheimer.

También se ha señalado (Liang et al., 2009a; Liang et al., 2009b) un papel de H-CutA en la secreción de proteínas, y en particular de la acetilcolinesterasa, quizá actuando como chaperona intra-Golgi. En un trabajo (escrito en chino del que solo he tenido acceso a su resumen en inglés) (Zhang et al., 2015) CutA fue detectada en un ensayo de doble híbrido que permitió identificar proteínas de interacción con PER1, la proteína implicada en los ritmos circadianos en humanos, por lo que no se puede descartar que CutA, como PII, interactúe con otras proteínas para llevar a cabo su función biológica.

Otros datos sobre la función biológica de CutA, igualmente indirectos, han establecido que CutA podría estar involucrada en el desarrollo humano (revisado en Yang et al., 2009), a nivel de crecimiento y diferenciación de células troncales embrionarias, en la supervivencia y maduración de los ovocitos, en el reconocimiento del gameto en el aparato genital femenino y en la activación de la secreción y desarrollo de la mama, considerándose a *cutA* uno de los 35 genes clave responsables de la activación de la secreción de leche (Naylor et al., 2005), aunque el mecanismo de ese papel no se ha establecido. En el campo de la patología no se han descrito afecciones debidas a mutaciones exclusivas en el gen *cutA* humano, si bien es uno de los genes perdidos en una microdelección del cromosoma 6p21.3 (Zollino et al., 2011; Writzl y Knegt, 2013)

en la que no ha sido posible establecer una relación clara entre el fenotipo observado y CutA debido a la diversidad de genes implicados, algunos de los cuales habían sido asociados previamente a patologías de carácter genético, como el gen HSET, relacionado con el síndrome de inmovilidad de cilios (Janitz et al., 1999).

B.5. CutA y metales.

La anotación de CutA relacionada con Cu^{2+} y otros cationes divalentes se basó, como se ha dicho, en los resultados de experimentos de genética molecular realizados con *E. coli* (Fong et al., 1995). Sin embargo, no se han encontrado en las proteínas CutA secuencias consenso de unión a cobre del tipo usual CXXC (Arnesano et al. 2003), y los resultados experimentales de unión de Cu^{2+} a la proteína pura de diferentes especies han dado resultados contradictorios o no generalizables. Por ejemplo, a pesar del gran número de estructuras publicadas o depositadas de CutA de diferentes especies (ver Tabla 1), solo en un caso, para CutA de *Pyrococcus horikoshi* (PhCutA), se ha detectado en la estructura la presencia de densidad electrónica atribuible a Cu^{2+} (Tanaka et al., 2004a), habiéndose publicado intentos sin éxito para la obtención de complejos cristalinos con Cu^{2+} (y en algunos casos con otros cationes divalentes) para CutA de *E. coli* (EcCutA) o de rata (RnCutA) (Arnesano et al., 2003), humana (HsCutA) (Bagautdinov et al., 2008), de *Thermus thermophilus* (Bagautdinov, 2014) e incluso en otro intento de cristalización con Cu^{2+} publicado con PhCutA (Bagautdinov, 2014).

En la estructura de PhCutA que contiene Cu(II) unido (Tanaka et al., 2004a), este catión se localizó entre dos subunidades de distintos trímeros de PhCutA, exactamente en un eje de simetría binaria del cristal, produciendo redes continuas infinitas de moléculas de CutA (Fig. 10) en las que el metal se

encuentra compartido por dos subunidades de trímeros vecinos, formando un complejo con el δO del Asp48 (residuo conservado o reemplazado conservadoramente por glutamato en las proteínas CutA) y el O de la Lys49 de las dos subunidades con las que interacciona.

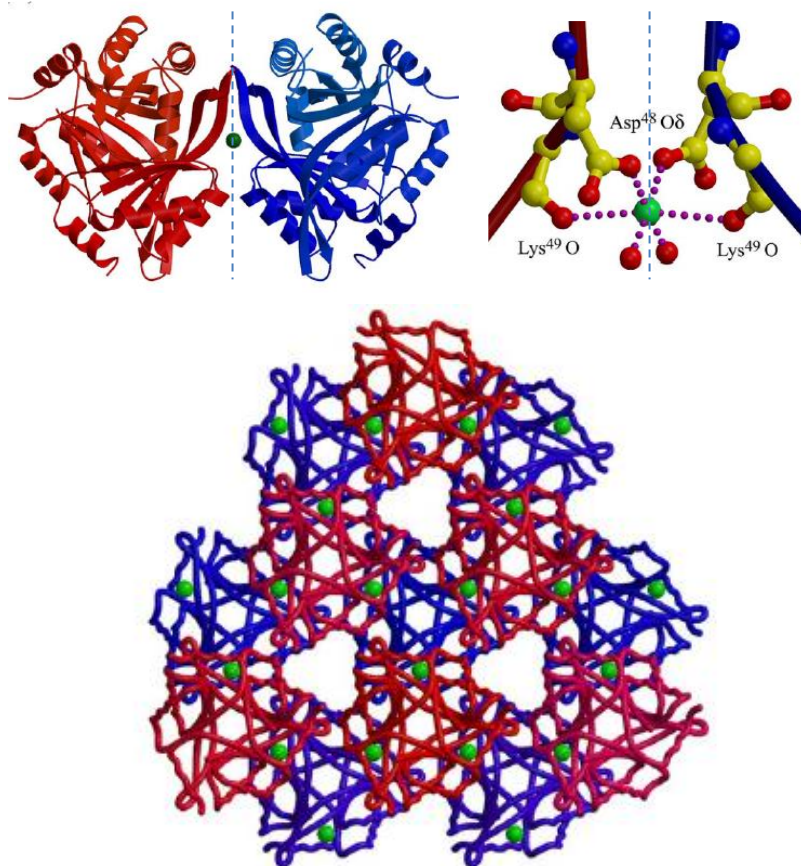


Figura 10. Reproducción de tres paneles de la Fig. 3 de Tanaka et al., 2004a, para mostrar la unión de Cu(II) en la estructura cristalina de CutA de *P. horikoshii* (PDB:1UKU). *Arriba izquierda*, dos trímeros de CutA con un ion Cu(II) en el eje de simetría binario cristalográfico que los relaciona (línea vertical discontinua). *Arriba derecha*, detalle del complejo de Asp48 y Lys49 de ambos trímeros, así como de dos moléculas de agua que participan en la esfera de interacciones del complejo. *Abajo*, estructura del multímero inducido por la unión del Cu^{2+} (verde). Se representan dos capas de trímeros (primera capa, azul; segunda capa, rojo). Reproducido con autorización de Wiley.

Las observaciones de que el Cu^{2+} , a una concentración de 4 mM, produjo precipitación de PhCutA (reversible al eliminar el Cu^{2+}), y de que la mutación Asp48Ala disminuyó fuertemente esta agregación (Tanaka et al., 2004a) avalaron la importancia del Asp48 en la unión de Cu^{2+} y en la formación de las redes que subyacen a la precipitación de la proteína (Fig. 10, panel inferior). Precisamente, en la estructura cristalina de CutA de *T. thermophilus* (Bagautdinov, 2014) el aspartato correspondiente al Asp48 de PhCutA interacciona con un ion Na^+ presente en la ubicación de un eje cristalográfico binario en los cristales de CutA de este organismo, llevando a la formación de dodecámeros organizados como tetraedros de cuatro trímeros de CutA.

En la estructura cristalina de EcCutA obtenida en presencia de p-hidroximercuribenzoato (HMB) (Arnesano et al., 2003) se identificaron tres átomos de Hg (del HMB o derivados de él) unidos a cada subunidad, cada uno asociado al átomo de S de uno de los tres residuos de cisteína de la subunidad (Cys 16, 39 y 79). Esto sugiere la posibilidad de que en EcCutA el Cu^{2+} pudiera unirse a esos sitios. La unión de Cu^{2+} a EcCutA se documentó espectroscópicamente (Arnesano et al., 2003), con una K_d del orden de 0.1 mM. Investigaciones de la naturaleza de esta unión mediante espectroscopia de absorción de rayos X (EXAFS) del complejo 1:1 CutA(subunidad):Cu fueron compatibles con un sitio para Cu^{2+} formado por Cys16, His83, His84 y o bien un agua fija o bien el $\gamma\text{-COO}^-$ del Glu90 de la subunidad vecina. Este sitio no puede generalizarse a proteínas CutA de otras especies puesto que de estos cuatro últimos residuos solo la His84 es invariante. Así mismo, solo la Cys39 de EcCutA está conservada en CutA de otras especies, formando parte del bolsillo entre cada dos subunidades, lo que, por cierto, determina que este bolsillo contenga HMB en el cristal de EcCutA. Desgraciadamente, no fue posible estudiar las

interacciones Cu^{2+} -EcCutA en mayor detalle usando RMN, debido a la precipitación de EcCutA por el Cu^{2+} en el tubo de RMN (Bertini et al., 2008).

En realidad la primera evidencia directa de que CutA puede unir Cu^{2+} se obtuvo con la proteína de *Arabidopsis thaliana* (Burkhead et al., 2003), al observar precipitación de la misma al añadirle Cu^{2+} 0.1 mM, una observación que se ha hecho también con proteínas CutA de otras especies, como se ha señalado anteriormente para las formas de *E. coli* (Arnesano et al., 2003) y de *P. horikoshi* (Tanaka et al., 2004a), y que se ha registrado también para las formas de rata (Arnesano et al., 2003) y de *Cryptosporidium parvum* (Buchko et al., 2015).

Volviendo a CutA de *A. thaliana*, otra evidencia de que une Cu(II) fue la observación (Burkhead et al., 2003) de que la proteína purificada se une selectivamente a columnas de Cu^{2+} , pero no de Ni, Zn, Mn o Cd, eluyéndose de las columnas de Cu^{2+} al añadir EDTA. Para titular cuánto Cu^{2+} puede unir la CutA de *A. thaliana* se incubó la misma con concentraciones variables de Cu^{2+} , eliminándose luego el metal libre mediante ultrafiltraciones repetidas en medio sin cobre, determinando entonces el Cu(II) residual tras la precipitación de la proteína con ácido tricloroacético, estimado en un máximo de un ion Cu(II) por cadena peptídica, infiriendo una constante de disociación de 50 μM para Cu(II), mientras que para Cu(I) no hubo unión sustancial.

La Fig. 11 resume en forma gráfica los distintos sitios identificados o propuestos para la unión de Cu(II) en proteínas CutA, localizados sobre la estructura de la proteína de *E. coli*. De ellos, el sitio con mayores posibilidades de ser generalizable podría ser la Cys39, por ser invariante y por encontrarse en el bolsillo intersubunidades. Sin embargo, nada de lo expuesto hasta ahora prueba la existencia de un sitio específico y funcionalmente significativo para Cu^{2+} en la molécula de CutA, ya que no se ha establecido claramente cuál es el

papel de CutA en cuanto a la homeostasis de cobre. Por ejemplo, el *knockout* de *cutA* en *A. thaliana* (Burkhead et al., 2003) no causó fenotipo alguno en cuanto a su crecimiento en ausencia o en presencia de CuSO_4 30 μM . Como se verá en el Capítulo 3 de la Sección 2 de esta Tesis, tampoco el *knockout* de *E. coli* para *cutA1* presenta un fenotipo de sensibilidad a Cu^{2+} que sea revertido por la transformación con *cutA1*. Hay que recordar que los metales pesados, incluyendo el cobre, causan alteraciones del plegamiento y agregación de proteínas (Tamás et al., 2014), y en particular el cobre puede usar como ligandos átomos de O, N y S de las proteínas, por lo que su unión a las mismas no tiene por qué ser específica o tener sentido biológico-funcional (Lemire et al., 2013).

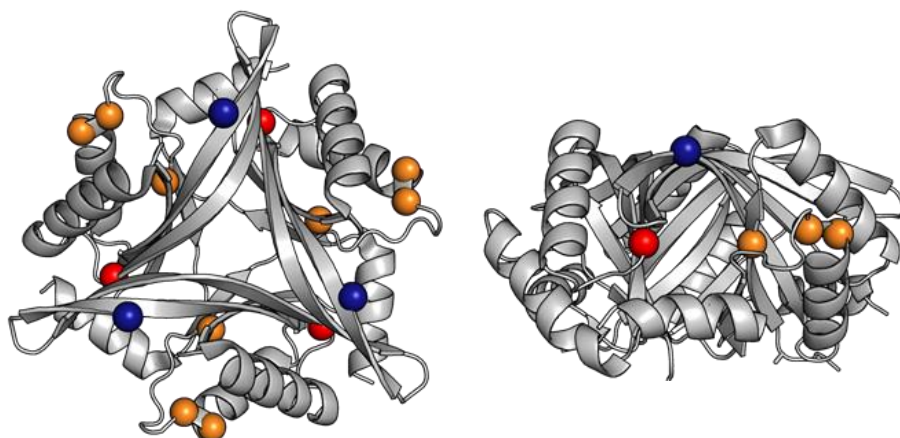


Figura 11. Localización en la estructura de CutA de *E. coli* de los residuos del sitio propuesto para Cu^{2+} a partir de experimentos de EXAFS (Cys16, His83 e His84; los tres en beige-anaranjado), de la Cys39, de carácter invariante (rojo) y del residuo correspondiente al Asp48 de PhCutA (azul).

C. El gen *COG0325*, sus productos en cianobacterias y en el ser humano, y su implicación en patología humana

C.1. El marco de referencia para nuestros estudios con la proteína PipY

Ya se ha comentado con anterioridad el papel central que desempeña la proteína PII en la regulación del flujo de nitrógeno en muchos organismos, y en particular en las cianobacterias (Fig. 2). Una de ellas, *S. elongatus*, es usada en nuestro laboratorio como sistema modelo en lo referente al sistema de PII, aunque nuestro interés en la misma se basa también en la importancia de las cianobacterias tanto en los ámbitos medioambiental y evolutivo (Margulis y Schwartz, 1982; Giordano et al. 2005) como porque poseen enorme potencial como generadoras de biomasa/fijadoras de CO₂ (Beck et al., 2012; Giordano et al., 2005) y como agentes explotables en biotecnología (Abed et al., 2009), sin olvidar sus importantes riesgos para la salud (cianotoxinas; www.who.int/water_sanitation_health/.../toxycyanchap3.pdf; Bláha et al., 2009; Watson et al. 2017).

Recordaremos la existencia en cianobacterias (Forcada-Nadal et al., 2018) (Fig. 2)nde dos dianas de PII privilegiadas y restringidas filogenéticamente: la enzima NAGK y la proteína adaptadora PipX, y la existencia de una red de regulación más amplia que incluye dos factores de transcripción, NtcA y PlmA, este último identificado recientemente (en estudios en que participé yo pero que no se reflejan en este trabajo de tesis Labella et al., 2016;), como tampoco se reflejan otros estudios aún no publicados sobre interacciones de PipX con otra proteína - todavía información confidencial - que fue objeto de un trabajo fin de grado de la UPV del que fui co-tutora).

Además de las interacciones físicas directas de PipX con otras proteínas, muy recientemente se ha demostrado la primera interacción funcional de PipX

sin interacción física detectada, con PipY de *S. elongatus* (Labella et al., 2017; Cantos et al., 2018). A PipY, a su ortólogo humano PLPHP y a las consecuencias patológicas de las mutaciones de este último se dedican los Capítulos 1 y 2 de la Sección 2 de esta Tesis, así como el resto de esta Introducción.

C.2. PipY, miembro cianobacteriano de la familia de proteínas COG0325.

Se sospechó una interacción PipX-PipY cuando se detectó (Labella et al., 2017) la constante asociación en cianobacterias entre el gen *pipX* y el que denominamos *pipY*, adyacente en el cromosoma bacteriano (Fig. 12).

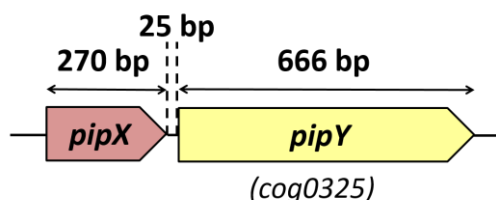


Figura 12. Esquema del operón PipX-PipY de *S. elongatus* PCC7942.

La sintenia de ambos genes y su monótona localización mutua, junto con el hecho de que formaran un operón en *S. elongatus*, donde menos del 40% de los genes forman operones di o policistrónicos, llevaron a suponer la existencia de una interacción entre ambos (Labella et al., 2017). Aunque los estudios exhaustivos de doble híbrido no documentaron interacción física entre las proteínas PipX y PipY, codificadas por ambos genes, sí se evidenció interacción funcional entre los mismos, documentada experimentalmente en las influencias mutuas de la delección de uno u otro de ambos genes sobre: 1) los niveles de expresión del otro a nivel de mRNA y de proteína; 2) los cambios observados en la sensibilidad a los antibióticos D-cicloserina (DCS) y β -cloro-D-alanina (BCDA); y 3) los patrones de transcriptómica global que por razones de brevedad no podemos detallar aquí (Labella et al., 2017). Estudios más recientes (Cantos et al., 2018) han revelado la complejidad de la relación entre

ambos genes en cuanto a expresión génica, implicando mecanismos genéticos en *cis* de prevención de efectos polares dentro del operón.

La inclusión de *pipY* en la red relacional de *pipX* nos hizo ponerlo bajo el foco. La proteína que codifica, PipY, y sus ortólogas de otras especies, responden al identificador COG0325 de EggNOG, con una amplia distribución

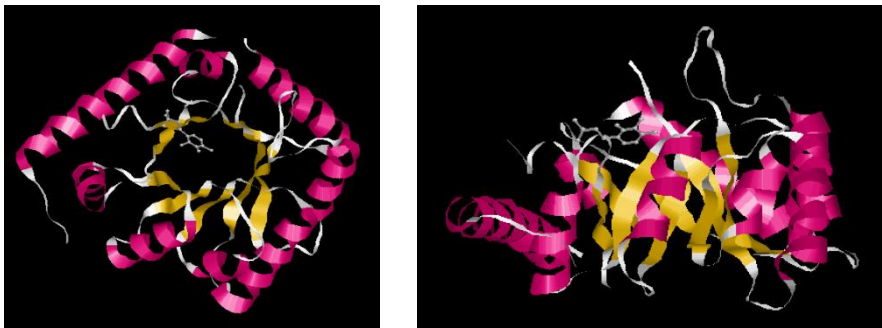


Figura 13. Estructura de la proteína de levadura YBL036, ortóloga de PipY de *S. elongatus* (Eswaramoorthy et al., 2003), vista a lo largo del eje del barril (izquierda) con los extremos C-terminales de las hebras beta paralelas (en amarillo) apuntando hacia el lector, o (derecha) con el eje del barril vertical y los extremos C-terminales apuntando arriba. Las hélices alfa se representan en color magenta, y el PLP (en bolas y palitos) en gris. Nótese que la figura de la derecha corresponde a una mayor ampliación que la de la izquierda. Obtenido del archivo del PDB 1B54.

en el árbol de la vida (Prunetti et al., 2016; y ver la ya citada base de datos EggNOG). Los genes *COG0325* codifican cadenas peptídicas de unos 250 residuos, aunque la longitud del producto predicho puede variar hasta en 25 aminoácidos (<http://www.uniprot.org>). En 2003 se publicó la estructura de *COG0325* de levadura (Eswaramoorthy et al., 2003) (Fig. 13), habiéndose depositado en el PDB posteriormente más estructuras de proteínas *COG0325* de otros organismos, sin publicación asociada (Tabla 2).

Las estructuras de proteínas *COG0325* revelan un solo dominio, plegado como un barril TIM modificado, con piridoxal fosfato (PLP) unido. El PLP es el derivado fosforilado y reducido de la vitamina B₆, o piridoxina. Es el coenzima

de más de 150 enzimas distintos (<http://bioinformatics.unipr.it/cgi-bin/bioinformatics/B6db/home.pl>; Coburn, 2015; Percudani y Peracchi 2009) entre los que se incluyen transaminasas, decarboxilasas y racemasas de aminoácidos. En el mecanismo de las reacciones de transaminación se genera la forma aminada (piridoxamina). Así, piridoxina, piridoxal y piridoxamina (y el producto metabólico ácido piridóxico) y sus formas fosforiladas se consideran los diferentes vitámeros B₆ (Fig. 14) (Bowling, 2011; Coburn, 2015). Posiblemente por la reactividad de su grupo aldehído, el piridoxal y PLP son tóxicos (Prunetti et al., 2016), toxicidad que también se da para piridoxina a elevada concentración para las bacterias (Labella et al., 2017).

Tabla 2: Estructuras de COG0325 resueltas y depositadas en el Protein DataBank (PDB)

Organismo	PDB	Año	Resol. (Å)	Ligandos	Notas ¹	Ref. ²
<i>Escherichia coli</i>	1W8G	2004	2.00	PLP		
	3SY1	2011	1.47	Ácido isocítrico Ácido acético Ácido 2-(N-morfolino)- etanosulfónico	10 muts.	
<i>Bifidobacterium adolescentis</i>	3CPG	2008	1.71	Ácido acético	MSE	
<i>Agrobacterium tumefaciens</i>	3R79	2011	1.90	PLP Ión acetato Ión praseodimio	MSE	
<i>Saccharomyces cerevisiae</i>	1CT5	1999	2.00	PLP	MSE	1
	1B54	1999	2.10	PLP		1

¹MSE: seleno-metionina. Muts: mutaciones.

²1: Eswaramoorthy et al. 2003

Las enzimas que usan PLP se han clasificado en cinco tipos estructurales diferentes (Eliot y Kirsch, 2004; Percudani y Peracchi, 2009):

- Plegamiento Tipo I- familia aspartato aminotransferasa
- Plegamiento Tipo II- familia triptófano sintasa

- Plegamiento Tipo III- familia alanina racemasa
- Plegamiento Tipo IV- familia D-aminoácido aminotransferasa
- Plegamiento tipo V- familia glicógeno fosforilasa

Las proteínas COG0325 presentan el característico plegamiento descrito por las enzimas de tipo III, un barril TIM [descrito originalmente para la triosa fosfato isomerasa (Banner et al., 1975), de cuyo nombre en inglés deriva el acrónimo TIM (Wierenga, 2001)] modificado en el que hay una hélice extra N-terminal precediendo a las ocho repeticiones β/α (Fig. 13). Sin embargo, a las proteínas COG0325 les falta el dominio extra presente en dichas enzimas de tipo III, que en las mismas contribuyen importantes residuos a los procesos de unión de sustratos y de catálisis (Espaillat et al., 2014; y ver Capítulos 1 y 2 de la Sección 2 de esta Tesis).

Aunque el elevado grado de conservación de las proteínas COG0325 y su amplia distribución sugieren que podrían tener una importante función, su gen no es esencial para la vida al menos para *E. coli*, *Bacillus subtilis* y *S. elongatus* (Ito et al., 2013; Ito et al., 2016; Prunetti et al., 2016; Labella et al., 2016) y hoy sabemos que también para el ser humano (Darin et al., 2016; Plecko et al., 2017). No obstante, la inactivación del gen *COG0325* de *E. coli* y *S. elongatus* afecta al metabolismo de aminoácidos y cetoácidos en forma pleotrópica, aumentando además la toxicidad de la piridoxina en estas bacterias (Ito et al., 2013, Prunetti et al., 2016; Labella et al., 2017), habiéndose descrito también que en *E. coli* produce acumulación de ácido oftálmico (Ito et al., 2016). Aunque inicialmente se describió (sin dar detalles) actividad D-alanina racemasa de la proteína COG0325 de levadura (Eswaramoorthy et al., 2003), estudios extensivos y detallados de Ito et al. (2013) con proteína COG0325 de varias fuentes (incluidas la de levadura y la humana) y otros de Prunetti et al. (2016) con la de *E. coli* no encontraron actividad enzimática con ninguna de

ellas. Quizá por exclusión y por la observación de efectos pleotrópicos, la idea más favorecida actualmente es la de que las proteínas COG0325 desempeñan un papel importante en la homeostasis para la vitamina B₆ (Prunetti et al., 2016), papel que apoyan también los resultados de los experimentos genéticos de inactivación del gen *pipY* en *S. elongatus* (Labella et al., 2017) y los resultados bioquímicos en pacientes con mutaciones de pérdida de función en el gen correspondiente humano (Darin et al., 2016; Plecko et al., 2017). Así, la proteína COG0325 estaría involucrada en la entrega de PLP a apoenzimas que deben utilizar PLP para ser capaces de desarrollar sus actividades enzimáticas.

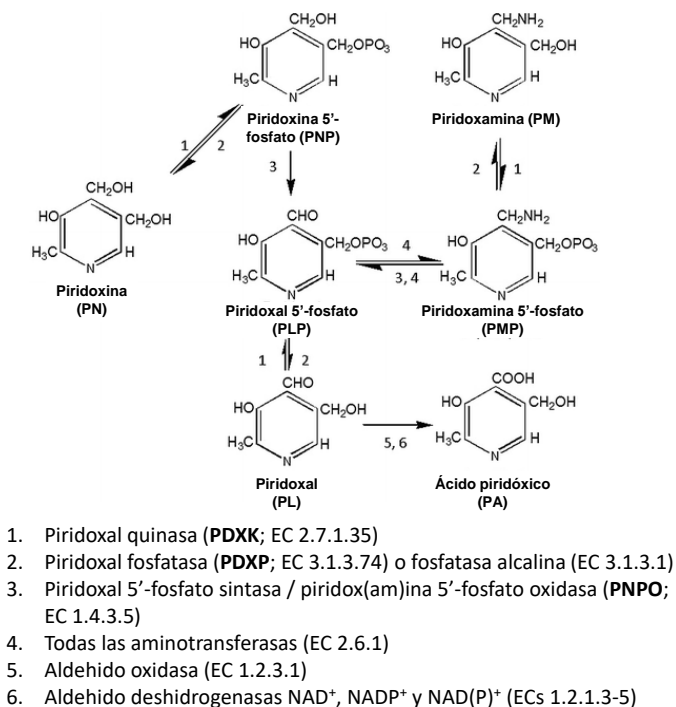


Figura 14. Vitámeros de la vitamina B₆, sus abreviaturas, y enzimas que los interconvierten. El PLP es la forma usada como cofactor enzimático. La piridoxina es la forma farmacéutica de la vitamina. Tomado de Coburn 2015, con licencia de Springer.

Los estudios de inactivación del gen *pipY* en *S. elongatus* (Labella et al., 2017) revelaron nuevos rasgos fenotípicos relacionados con el PLP, como aumento específico de la sensibilidad de la cianobacteria a los antibióticos D-cicloserina y β -cloro-D-alanina, los cuales tienen como dianas PLP-holoenzimas (Labella et al., 2017), aunque ni para estos ni para los demás efectos se aclaró el mecanismo molecular que subyace a la función. La voluntad de caracterizar la molécula de PipY de *S. elongatus* como continuación de dichos experimentos funcionales *in vivo* nos llevó a ocuparnos inicialmente de PipY de *S. elongatus*, aunque nuestro interés se incrementó fuertemente cuando surgieron interacciones claras con patología humana.

C.3. Por qué nos hemos interesado en la proteína COG0325 humana.

Nuestro interés en la proteína COG0325 humana se derivó de la publicación a finales de 2016 de algunos pacientes con epilepsia genética dependiente de vitamina B₆ en los que se identificaron mutaciones en ambos alelos del gen *PROSC* humano (Darin et al., 2016), el ortólogo de *pipY*, en algunos casos siendo las mutaciones claramente de pérdida de función (truncaciones por cambios de fase o codones de terminación prematuros). Por entonces, llevados por los estudios en el sistema PII-PipX, habíamos obtenido la proteína PipY de *S. elongatus* producida recombinantemente y en alto grado de pureza, y estábamos caracterizando sus propiedades moleculares, habiendo determinado ya su estructura cristalina a excelente resolución. La implicación del grupo en enfermedades raras de causa genética y naturaleza metabólica nos inclinó a intentar usar PipY como proteína modelo de *PROSC* para tratar de establecer el mecanismo de patogenicidad de las mutaciones de cambio de sentido (*missense*) encontradas en los pacientes.

A mediados de 2017 apareció un segundo trabajo (Plecko et al., 2017) que identificaba más mutaciones en *PROSC* como causa de epilepsia genética humana dependiente de vitamina B₆. El trabajo era en buena medida de análisis genético focalizado en el gen *PROSC* de pacientes antiguos con epilepsia genética dependiente de vitamina B₆ de naturaleza no filiada. Ello nos llevó a intentar producir la proteína humana en cantidades susceptibles de estudio cristalográfico, y de abordar su cristalización. Aunque todavía no hemos tenido éxito en conseguir cristales de la misma, ha sido muy útil producir la proteína humana, que también hemos caracterizado bioquímicamente, para, usando mutagénesis dirigida, introducir todas las mutaciones de cambio de sentido descritas en pacientes hasta hoy para investigar su capacidad de causar enfermedad, algo que no habría sido posible utilizando PipY de *S. elongatus*, dadas las diferencias de secuencia que presenta con su forma ortóloga humana.

C.4. Epilepsia dependiente de vitamina B₆ y mutaciones en el gen *PLPBP*

Antes de terminar esta Introducción me ocuparé brevemente de las epilepsias genéticas dependientes de vitamina B₆. La epilepsia es una enfermedad frecuente que en una fracción importante de los casos deriva de alteraciones adquiridas como son traumatismos craneales, tumores, ictus, cirugía, infecciones etc. Sin embargo, en la mayoría de casos es criptogenética, reconociéndose a estas formas cada vez más una fuerte carga genética, en general poligénica, siendo más raras las formas hereditarias monogénicas puras en las que la epilepsia es la manifestación principal (Myers y Mefford, 2015). De esas epilepsias genéticas, las debidas a mutaciones en genes que codifican canales iónicos que se expresan en el sistema nervioso central son las más usuales (Berkovic y Scheffer, 2001; Myers y Mefford, 2015). Sin embargo, un

cierto número de epilepsias genéticas se deben a fallos enzimáticos (errores congénitos del metabolismo) y se consideran epilepsias metabólicas (Tabla 3), siendo algunas de ellas tratables de forma específica (Campistol, 2016; Sharma y Prasad, 2017). Este último es el caso de las epilepsias genéticas que responden a vitamina B₆, bien a piridoxina, bien a piridoxal fosfato (resaltadas en rojo en la Tabla 3) (Clayton, 2006).

Tabla 3. Epilepsias asociadas a errores congénitos del metabolismo. Se resaltan en rojo las epilepsias dependientes de vitamina B₆. Adaptado de Sharma y Prasad 2017 (Licencia Creative Commons Attribution CC BY).

Periodo neonatal-primera infancia	Infancia tardía-niñez	Adolescencia-edad adulta
Epilepsia dependiente de piridoxina Deficiencia de PNPO	Defectos de la síntesis de creatina	LNC juvenil
Convulsiones sensibles a ácido fólico	LNC infantil tardía	Enfermedades de Lafora y de Unverricht-Lundborg
Deficiencia de biotinidasa	Enfermedades mitocondriales	MELAS (encefalomiopatía mitocondrial, acidosis láctica, episodios tipo ictus)
Deficiencia de holocarboxilato sintetasa	Acidurias orgánicas	MERRF (epilepsia mioclónica con fibras rojas rasgadas)
Deficiencia de GLUT-1 (transportador-1 de glucosa)	Sialidosis	Enfermedades lisosomales: gangliosidosis de inicio tardío, Niemann-Pick tipo C, Gaucher tipo III
Defectos en la biosíntesis de serina	Gangliosidosis	Deficiencia de GLUT-1
Deficiencia de cofactor de molibdeno y de sulfito oxidasa	Deficiencia de GLUT-1	Porfiria
Enfermedad de Menkes	Alteraciones congénitas de la glicosilación	Enfermedad de Wilson
Hiperglicinemia no-cetósica	Defectos del metabolismo de purinas	
Acidemias orgánicas	Epilepsia uridina-dependiente por mutaciones en CAD	
Defectos del ciclo de la urea	Defectos de la metilación y del metabolismo de folato	
Enfermedades peroxisomales	Defectos en neurotransmisores	
Alteraciones congénitas de la glicosilación	Alteraciones congénitas de la autofagia	
Lipofuscinosis neuronal ceroida (LNC) congénita e infantil temprana		
Enfermedades mitocondriales		

Las epilepsias genéticas dependientes de vitamina B₆ son trastornos caracterizados por convulsiones que no responden o responden insatisfactoriamente a fármacos antiepilépticos convencionales, pero cuyos

síntomas mejoran de forma importante o completa ante la administración de vitamina B₆ (Falsaperla y Corsello, 2017) o de su forma fosforilada PLP (Wang y Kuo, 2007; Plecko, 2013; Sharma y Prasad, 2017). Es bien sabido que la avitaminosis B₆ produce alteraciones electroencefalográficas y convulsiones epilépticas (Leklem, 1990). No es fácil desarrollar tal avitaminosis, dada la abundancia de los distintos vitámeros de la vitamina B₆ en los alimentos, más el aporte de las bacterias intestinales (Wang y Kuo, 2007). El ejemplo más patente de dicha avitaminosis lo proporcionó en los años 50 del pasado siglo un grupo de lactantes que recibieron como única alimentación una fórmula deficitaria en vitamina B₆ y que desarrollaron una forma de epilepsia que mejoró inmediatamente con la administración de esta vitamina (Coursin, 1954). Por esas mismas fechas se describió el primer caso de epilepsia genética dependiente de vitamina B₆ en una niña con convulsiones intratables con fármacos antiepilépticos comunes, que pudieron controlarse con piridoxina (Hunt et al., 1954). Dado que el GABA es un neurotransmisor inhibitorio de gran importancia en epilepsia cuya síntesis requiere la descarboxilación del glutamato por una descarboxilasa que usa PLP, es común atribuir a una deficiencia cerebral de GABA la epilepsia asociada al déficit de vitamina B₆, ya sea nutricional (raro, normalmente solo observada en alcohólicos) o genética (Coburn, 2015). De hecho, se han descrito bajos niveles de GABA en el líquido cefalorraquídeo de niños con epilepsia dependiente de piridoxina (Kurlemann et al., 1992). Sin embargo, el mecanismo proconvulsivo podría ser más complejo, ya que numerosos neurotransmisores además del GABA requieren para su producción o su metabolización PLP-enzimas y el mismo GABA es catabolizado por transaminación, reacción que también requiere PLP (Tabla 4).

Hasta finales de 2016 se conocían cuatro tipos de epilepsia genética sensible a vitamina B₆ (Plecko, 2013) todas ellas con herencia recesiva (Tabla 5). Dos de

ellas (PDE y HP II) se deben a déficits secundarios de vitamina B₆ por acumulación de compuestos tóxicos para el piridoxal, con el que dan reacciones espontáneas de condensación, inactivándolo (Farrant et al., 2001). Como son de hecho hipovitaminosis B₆ no por falta de ingestión sino por inactivación excesiva de esa vitamina, la administración de piridoxina a grandes dosis evita en ambas la sintomatología (Plecko, 2013; Falsaperla y Corsello, 2017).

Tabla 4. PLP-enzimas involucradas en la síntesis o el metabolismo de neurotransmisores

Glutamato descarboxilasa	Conversión de glutamato en GABA
GABA transaminasa	Catabolismo de GABA, regenerando glutamato
Descarboxilasa de aminoácidos aromáticos	Síntesis de serotonina, dopamina y norepinefrina
Transaminasa de aminoácidos ramificados a α -cetoglutarato	Síntesis de glutamato
Componente glicina deshidrogenasa (descarboxilante) del complejo de escisión de glicina	Degradación del aminoácido neurotransmisor inhibidor glicina
Racemasa de L-serina	Formación del neurotransmisor D-serina
Descarboxilasa de histidina	Formación de histamina

Tabla basada en una parte de la Tabla 1 de Clayton 2006. Reproducido con autorización de Springer.

La epilepsia dependiente de piridoxina (PDE) se debe al déficit de α -aminoadípico-semialdehído deshidrogenasa (AASD), también conocida como antiqitina (Mills et al., 2006), déficit que resulta en un bloqueo de la ruta del ácido piperídico del catabolismo de la lisina con acumulación de dicho ácido cíclico (biomarcador de la enfermedad; Plecko 2013) así como, aún más, de acumulación de α -aminoadípico-semialdehído, que se cicla espontáneamente a L- Δ^1 -piperideín-6-carboxilato (P6C). Lo mismo sucede para un carbono menos

en la mucho más rara hiperprolinemia tipo II (HP II), en la que se acumula glutamato semialdehído, que cicla a Δ^1 -pirrolin-5-carboxilato (P5C), idéntico al P6C excepto por tener un carbono menos en su anillo. Para el P5C se sabe desde 2001 (Farrant et al., 2001) que es un veneno del PLP, pues lo inactiva al reaccionar con él en una condensación de tipo Knoevenagel. Por tanto, este mecanismo parece ser también el que causa la deficiencia de vitamina B₆ en el déficit de antiqutina (Mills et al., 2006).

Tabla 5. Epilepsias dependientes de vitamina B₆

Enfermedad	Abreviatura #OMIM	Mecanismo	Gen / proteína	Vitámero B ₆ eficaz
Epilepsia dependiente de piridoxina	PDE #266100	Inactivación de piridoxal por reacción con L- Δ^1 -piperideín-6-carboxilato (P6C) acumulado en déficit de degradación de lisina	<i>ALDH7A1</i> / α -Amino-adípico semialdehído deshidrogenasa (AASD, antiqutina)	Piridoxina
Hiperprolinemia tipo II	HP II #239510	Inactivación del piridoxal por reacción con Δ^1 -pirrolin-5-carboxilato (P5C) acumulado en déficit de degradación de prolina	<i>ALDH4A1</i> / Δ^1 -Pyrrolin-5-carboxilato deshidrogenasa	Piridoxina
Hipofosfatasa congénita	HPP congénita #241500 #146300	Disminución de PLP intracelular por posible baja disponibilidad para el cerebro	<i>ALPL</i> / Fosfatasa alcalina tisular no específica (TNSALP)	Piridoxina
Epilepsia dependiente de PLP	Deficiencia de PNPO #610090	Mala transformación de vitámeros a PLP	<i>PNPO</i> / Oxidasa de piridox(am)ina 5'fosfato	PLP
Deficiencia de PLPHP	PLPHP o EPVB6D #617290	Deficiencia en la homeostasis de vitamina B ₆	<i>PLPBP</i> (o <i>PROSC</i>) / PLPHP	Piridoxina (PLP en déficit total)

En cambio, en la hipofosfatasa congénita el fallo primario es una deficiencia en fosfatasa alcalina tisular no específica que se sabe puede responder a la administración de piridoxina (Rasmussen, 1983). La observación de niveles sanguíneos elevados de PLP en esta enfermedad casa con el esquema de la Fig. 15, en el que el cerebro recibe piridoxal derivado de la hidrólisis de PLP sanguíneo mediada por esta fosfatasa, muy abundante en el plexo coroideo (Coburn, 2015). En los pacientes de este proceso la epilepsia controlable por administración de piridoxina es una manifestación temprana y clínicamente muy visible, pero no la más grave (dada su tratabilidad con piridoxina) de un proceso de hipofosfatasa generalizada que suele acabar pronto con la vida del paciente (Plecko, 2013).

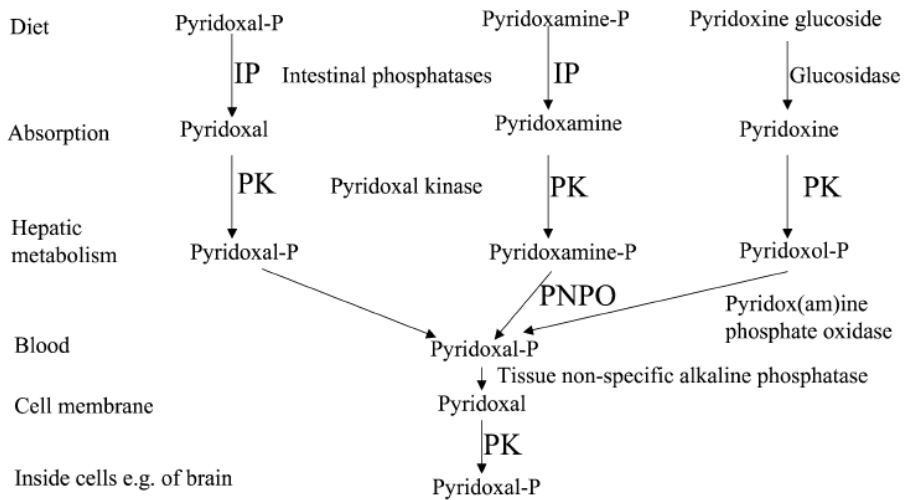


Figura 15. Esquema simplificado del metabolismo de los vitámeros de la vitamina B₆. Tomado (con licencia de Springer) de Clayton 2006.

Solo una de las cuatro formas conocidas antes de finales de 2016 de epilepsia dependiente de vitamina B₆ requiere obligatoriamente utilizar PLP en el tratamiento (Plecko, 2013), ya que no responde a la administración de

piridoxina. Se trata del déficit de PNPO (Mills et al., 2005), la enzima que permite la generación de PLP a partir de los vitámeros fosforilados de piridoxamina y piridoxina (Figs. 14 y 15).

En la Tabla 5 hemos incluido también la nueva forma de epilepsia dependiente de vitamina B₆ debida a mutaciones en el gen *PLPBP* (antes *PROSC*) identificada inicialmente (Darin et al., 2016) en tres niños de una familia siria y en cuatro pacientes de otras familias no relacionadas (dos de origen indio, una alemana y otra italiana), concluyendo que la herencia de la enfermedad era autosómica recesiva. Realizaron algunos estudios de complementación en *E. coli knockout* en el gen *COG0325*, usando transformación con plásmidos que codificaban formas mutantes, obteniendo resultados compatibles con las observaciones clínicas y atribución de la patología a pérdida de función, excepto para dos mutante de cambio de sentido en que los resultados parecían contradictorios (ver Introducciones de los capítulos 1 y 2 de la Sección 2 de esta Tesis). En todo caso sus resultados parecían estar de acuerdo con que la función de *PROSC* era de homeostasis para PLP, como se había propuesto ya para las proteínas *COG0325* bacterianas (Ito et al., 2013; Ito et al., 2016; Prunetti et al., 2016; Labella et al., 2016). Tan solo cuatro meses después se publicaba (Plecko et al., 2017) la confirmación de esta patología con el estudio de cuatro pacientes más, aportando datos analíticos (ver capítulo 2 de la Sección 2) que favorecían una deficiencia de múltiples enzimas PLP-dependientes, por tanto dando apoyo a un papel de la proteína deficiente en la transferencia de PLP a esos enzimas. Incluyendo los pacientes de uno y otro trabajo, se han descrito once mutaciones (seis puntuales de cambio de sentido, bien en homocigosis o heterocigosis compuesta, tres sin sentido y dos de *splicing*), siempre con presentaciones consistentes con herencia recesiva, y con cierta variabilidad de los cuadros

clínicos, que incluían deterioro mental cuando la epilepsia era de larga duración y no tratada con piridoxina (Darin et al., 2016; Plecko et al., 2017). Un dato a resaltar no bien entendido con esta epilepsia, y que requiere ampliación de la experiencia clínica con esta enfermedad con un grupo mayor de casos aún por describir, es el hecho de que la piridoxina fuera más o menos efectiva dependiendo del paciente, haciendo precisa en algún paciente la administración de PLP para obtener una buena respuesta (Darin et al., 20116). Dejamos para el capítulo 2 de la Sección 2 y para la Sección 3 (Discusión General) las reflexiones sobre las razones de este comportamiento, que colocan a esta epilepsia a mitad de camino entre las dependientes de piridoxina (mayoría de los casos de deficiencia de PLPHP) y las dependientes de PLP.

Esta nueva enfermedad plantea varios problemas. El primero es la falta de información sobre las propiedades y características de la proteína humana (antes denominada PROSC y ahora PLPHP). En segundo lugar, el hecho de que no hay lugar específico en el esquema de la Fig. 15 para PLPHP, lo que refleja que la función de esta proteína no ha sido aclarada en términos moleculares. En tercer lugar, el papel/mecanismo patogénico de las mutaciones puntuales de cambio de sentido, que debería ser abordado en forma directa con la proteína pura, y que plantea la dificultad adicional de que la proteína no tiene una actividad bien establecida. Es la existencia de estos interrogantes, y nuestro deseo de hacer patente que conocimientos previos del grupo sobre la biología de cianobacterias podían tener valor e impacto biomédicos, lo que nos impulsó a abordar las investigaciones que están relacionados con esta forma de epilepsia y que se resumen en los capítulos 1 y 2 de la Sección 2 de esta memoria de Tesis.

Finalmente, señalar que, incluso en un esquema muy simplificado como el de la Fig. 15, existen pasos para los que no se han descrito patologías (por

ejemplo los catalizados por la piridoxal quinasa, o los de transporte intestinal o transmembranal, no aclarados; Whittaker, 2016). Por ello, quizá existan todavía epilepsias genéticas susceptibles de terapia con PLP (el caso de déficit de quinasa) o incluso intratables con cualquiera de los vitámeros conocidos, como en el caso de que existan transportadores obligatorios y no redundantes que puedan fallar o tener actividad disminuida por mutaciones en sus genes (Whittaker, 2016). Incluso cabe plantearse si en poblaciones como la de la India, esencialmente vegetariana, o en la comunidad vegana occidental, cada vez más amplia, déficits de la glucosidasa que libera la piridoxina de su mayoritario glucósido vegetal (Fig. 15) podría tener impacto patológico, siendo causa de epilepsia.

D. Estructura de esta Memoria de Tesis

Esta Tesis responde a la forma conocida como compendio de publicaciones. Incluye tres manuscritos, dos ya publicados y el tercero en proceso de publicación. Los dos ya publicados son objeto de los capítulos 1 y 2 de la Sección 2 de esta Tesis. El Capítulo 1 se ocupa de la proteína PipY de *S. elongatus*, que se caracteriza en el grado más completo posible (dentro de nuestras posibilidades), y que se usa como modelo para la epilepsia dependiente de vitamina B₆, de descripción más reciente y de la que me acabo de ocupar en el epígrafe anterior. El Capítulo 2 se ocupa de la proteína PROSC/PLPHP humana, describiendo su producción, purificación, propiedades y estabilidad, y caracterizando con la misma los efectos de todas las mutaciones puntuales de cambio de sentido descritas en los pacientes de la epilepsia que acabo de mencionar. El Capítulo 3 se ocupa de la proteína CutA, investigando su función y su estructura, esta última para la CutA de *S. elongatus*.

Quizá parezca extraño que la Introducción General se ocupe primero de la proteína CutA y luego de PipY y de la epilepsia humana dependiente de vitamina B₆, y que luego en la Sección 2 el orden sea el inverso. El orden de la Introducción ha sido el que nos ha parecido óptimo para introducir la temática y para explicar nuestra peripezia personal. El orden de la Sección 2, sin embargo, responde al orden temporal de publicación, y además no es inconsistente con el desarrollo de la Introducción General (Sección 1) ya que dicha introducción acaba con PipY y con la epilepsia dependiente de vitamina B₆, y por tanto, el empezar con capítulos sobre estas últimas temáticas hará que el lector las aborde en un continuo sin fracturas, y sin necesidad de recordatorio del tema.

BIBLIOGRAFÍA

Para evitar repeticiones, las referencias para esta sección se dan conjuntamente con las de la Discusión General, al final de esta Tesis.

E. Objetivos

Los objetivos de esta tesis se encuadran en el ámbito de la señalización por nitrógeno. Parten de la red de señalización centrada en PII y PipX, y se refieren a tres proteínas: 1) CutA, por su pertenencia a la superfamilia de proteínas PII; 2) PipY de *S.elongatus* PCC7942, por la ubicación de su gen junto al de *pipX* en este organismo y en otras cianobacterias y por los avances sobre el conocimiento de sus funciones *in vivo*, en que ha estado implicado el grupo pero que no han aclarado el mecanismo molecular de dichas funciones; y 3) la proteína PROSC/PLPHP humana por ser la proteína ortóloga a PipY en humanos, de función molecular también por aclarar, y cuyas mutaciones clínicas causan epilepsia dependiente de vitamina B₆. Los objetivos relativos a PipY y PROSC/PLPHP se han visto fuertemente influenciados por la descripción en el último periodo de la tesis de la patología humana genética relacionada con PROSC/PLPHP. En esos contextos, los objetivos específicos de esta tesis son (dados en el orden de los capítulos de la Sección 2):

- Producir de forma pura y recombinante y caracterizar molecularmente la proteína PipY de *S.elongatus* como miembro representativo de la familia de proteínas COG0325 y resolver su estructura cristalina en su forma libre y con PLP unido para tratar de obtener datos conducentes a interpretar su función.
- Utilizar la proteína PipY como modelo de la proteína PLPHP humana para tratar de explicar experimentalmente e *in silico* los efectos producidos por las mutaciones clínicas causantes de epilepsia dependiente de vitamina B₆.
- Producir de forma pura y recombinante y caracterizar la proteína PLPHP humana en su forma silvestre y en sus formas mutantes clínicas descritas, para tratar de contribuir a elucidar el mecanismo molecular de la patología asociada a la misma. También nos fijamos como objetivo cristalizarla, lo que no ha

tenido éxito por ahora, por lo que la descripción de los intentos de cristalización no se incluyen en esta memoria de Tesis.

- Determinar si existe o no implicación de *cutA* bacteriana en los mecanismos de tolerancia a metales, generalizando los resultados a dos especies bacterianas muy separadas evolutivamente, en concreto *Escherichia coli* y *Synechococcus elongatus*.

- Caracterizar la proteína CutA de *S. elongatus* producida de forma recombinante, determinar si une cobre *in vitro* y resolver su estructura cristalina por cristalografía de rayos-X para tratar de establecer el sitio de unión a cobre y contribuir a esclarecer su función.

SECCIÓN 2

Publicaciones

Capítulo 1: Studies on cyanobacterial protein PipY shed light on structure, potential functions and vitamin B₆-dependent epilepsy.

Autores: Lorena Tremiño, Alicia Forcada-Nadal, Asunción Contreras y Vicente Rubio*

Referencia: *FEBS Lett.* (2017) 591,3431-3442

Capítulo 2: Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations.

Autores: Lorena Tremiño, Alicia Forcada-Nadal y Vicente Rubio*

Referencia: *Hum. Mutat.* (2018) 39, 1002-1013

Capítulo 3: Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria

Autores: Lorena Tremiño, Javier Espinosa, Asunción Contreras y Vicente Rubio*

Manuscrito

Capítulo 1: Studies on cyanobacterial protein PipY shed light on structure, potential functions and vitamin B₆-dependent epilepsy.

Autores: Lorena Tremiño, Alicia Forcada-Nadal, Asunción Contreras y Vicente Rubio*

Referencia: *FEBS Lett.* (2017) 591,3431-3442

Studies on cyanobacterial protein PipY shed light on structure, potential functions and vitamin B6-dependent epilepsy

Lorena Tremiño¹, Alicia Forcada-Nadal^{1,2}, Asunción Contreras² and Vicente Rubio^{1*}

¹ Instituto de Biomedicina de Valencia (IBV-CSIC), and CIBER de Enfermedades Raras (CIBERER-ISCIH), Valencia, Spain. ² Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain.

*Corresponding author, rubio@ibv.csic.es

Received 2 July 2017; accepted 1 September 2017; available online 26 October 2017.

Abstract. The *Synechococcus elongatus* COG0325 gene, *pipY*, functionally interacts with the nitrogen regulatory gene *pipX*. As a first step towards molecular understanding of these interactions, PipY was characterized. This 221-residue protein is monomeric and hosts pyridoxal phosphate (PLP), binding it with limited affinity and losing it upon D-cycloserine incubation. PipY crystal structures with and without PLP revealed a single-domain monomer folded as the TIM barrel of type III-fold PLP enzymes, with PLP highly exposed, fitting a PipY role in PLP homeostasis. The mobile PLP phosphate-anchoring C-terminal helix might act as trigger for PLP exchange. Exploiting the universality of COG0325 functions, we used PipY in site-directed mutagenesis studies to shed light on disease causation by epilepsy-associated mutations in the human COG0325 gene, *PROSC*.

Keywords. *Synechococcus elongatus* PCC7942, COG0325, pyridoxal phosphate proteins.

Abbreviations. CD, circular dichroism; DCS, D-cycloserine; DTT, dithiothreitol; ODC, ornithine decarboxylase; PDB, Protein Data Bank; PEG, polyethylene glycol; PLP, pyridoxal 5'-phosphate; r.m.s.d., root-mean-square deviation.

INTRODUCTION

Cyanobacteria, the chloroplasts ancestors [1], are very ancient, widespread and abundant [2] photosynthetic bacteria that play key roles in global oxygen, CO₂ and nitrogen fluxes [3]. Nitrogen incorporation is highly regulated in cyanobacteria [4]. The small protein PipX [5,6] is a key player, as highlighted by transcriptomic studies in *Synechococcus elongatus* PCC7942 (from now on *S. elongatus*) [7]. PipX regulation, mediated by the carbon/energy/nitrogen signalling protein PII and the gene expression regulator NtcA [5,8,9], also involves newly discovered partners as PlmA [10].

Very recently, the synteny in cyanobacteria of the mutually adjacent *pipX* and *pipY* genes led to the identification in *S.elongatus* of functional interactions between these two genes [11]. *pipY* is a member of the highly widespread [12] and intriguing *COG0325* gene family [11] that includes *Escherichia coli yggS*, a gene that when inactivated causes a phenotype of pyridoxine toxicity and pleiotropic changes related to amino acid-related metabolism [12-14]. Inactivation of *pipY* in *S.elongatus* [11] also enhanced pyridoxine toxicity and led to increased sensitivity to antibiotics that target pyridoxal phosphate (PLP) such as D-cycloserine (DCS) [15]; these two traits were influenced by various amino acids [11]. Pleiotropic changes were also observed in null mutants of the *COG0325* genes *ylmE* of *Bacillus subtilis* [16] and *PROSC* of humans [17,18]. The deficiency of the human gene was reported to cause vitamin B6-dependent epilepsy [17,18]. The pleiotropy of the changes associated with *COG0325* mutations is best explained if the *COG0325* products, which are PLP proteins [19], rather than being enzymes play key roles in PLP homeostasis. These roles remain uncharacterized molecularly.

Striving towards delineation of the roles of PipY, we decided to characterize physically and structurally PipY from *S. elongatus*. Such characterization

appeared pertinent, given the extensive functional data collected *in vivo* for its corresponding gene [11]. Furthermore, there was only one published analysis of the structure of another COG0325 protein, YBL036C from yeast [19], an organism that is phylogenetically distant from cyanobacteria. The YBL036C structure had been determined as a part of a structural genomics effort, and the biological *in vivo* role of YBL036C had not been specifically explored. In addition, the structural knowledge of YBL036C did not include the corresponding structure without PLP. However, the changes associated with the presence or absence of PLP may be highly relevant for a role of COG0325 proteins in PLP delivery to other targets. Although crystal structures of three bacterial COG0325 proteins have been deposited in the Protein Data Bank (PDB) (Supplementary Table S1), none of these structures has been analysed, justifying the need for thorough structural analysis of a bacterial COG0325 protein in the absence and the presence of PLP. We carry out here such structural analysis on *S. elongatus* PipY. The detailed structural information gained in the present studies by X-ray crystallography of PLP-containing and PLP-free PipY makes of PipY a true structural paradigm for COG0325 proteins. In addition, we have characterized some PipY traits in solution, including its mass, its spectroscopic properties and both its ability to bind PLP and its sensitivity to the PLP-targeting antibiotic DCS. The toxicity of this antibiotic for *S. elongatus* was previously shown to be increased by *pipY* inactivation [11].

We also get insight here on the effects of two PROSC mutations, Pro87Leu and Arg241Gln, reported in patients with vitamin B6-dependent epilepsy [17,18]. Puzzlingly, the clinical severity inferred for these mutations did not parallel the ability of the corresponding mutant *PROSC* gene forms to complement an *E.coli yggS*⁻ mutant for its phenotype of increased sensitivity to pyridoxine toxicity [17]. To try to explain these puzzling observations, we have introduced into PipY the mutations corresponding to Pro87Leu and

Arg241Gln of PROSC, studying relevant characteristics of the mutant PipY forms. This approach has exploited the apparent universality of COG0325 functions (for which the recent studies on *S. elongatus* provided strong confirmation [11]) and the sequence conservation between *S. elongatus* PipY and human PROSC (Supplementary Fig. S1). In this way we provide valuable hints on the respective ability and lack of ability of the Pro87Leu and Arg241Gln PROSC mutants to override the mutant phenotype of the *yggS*⁻ *E.coli* mutant strain [17], as well as on the possible mechanism of disease causation by the Arg241Gln PROSC mutation.

MATERIALS AND METHODS

Production of wild-type and mutant PipY

The *S.elongatus* *pipY* open reading frame (gene *Synpcc7942_2060*, http://genome.microbedb.jp/cyanobase/GCA_000012525.1/genes/Synpcc7942_2060), PCR-amplified from genomic DNA using Deep Vent DNA polymerase (from New England Biolabs) and the primers 5'-GTCAGAGGCACATATGGCCCAAATTGC-3' and 5'-GGCCAAGACTCGAGCGATCGCGGC-3', was digested with *NdeI/XhoI* and ligated into the similarly digested pET-22b+ plasmid (from Novagen). The resulting pET-22b(+)-2060 plasmid, isolated from *E.coli* DH5 α , was used for site-directed mutagenesis using the Quickchange approach (Stratagene, La Jolla, CA) and the following forward and reverse mutagenic primers: for the Pro63Leu mutation, 5'ACGGATCTGCTGGATTTGACTTGG3' and 5'AAGTCAAATCCAGCAGATCCGTCAG3'; and for the Arg210Gln mutation, 5'GCAACTTGGATTCAAGTCGGAACC3' and 5'TTCCGACTTGAATCCAAGTTGCC3'. The presence of the correct sequences and of the desired mutations was verified by Sanger sequencing of the plasmids.

To produce PipY with a C-terminal -GluHis6 tag, the pET-22b(+)-2060 plasmid was transformed into *E.coli* BL21(DE3) (from Novagen), growing the transformed cells in 1-L liquid LB-ampicillin (0.1 mg/ml) at 37 °C, to ~0.8 OD⁶⁰⁰, followed by 3-hour induction with 1 mM isopropyl β-D-1-thiogalactopyranoside. All subsequent steps were at 4 °C. After centrifugation, the cell pellet, suspended in 30 ml of 25 mM HEPES pH 7.5/0.4 M NaCl/1 mM dithiothreitol (DTT)/1 mM phenyl methyl sulphonyl fluoride/20 mM imidazole, was sonicated to disrupt the cells, the suspension was centrifuged, and PipY was purified from the supernatant by Ni-affinity chromatography and imidazole gradient elution (His-Trap HP 1-ml column in an Åkta-FPLC, both from GE Healthcare). The resulting preparation was concentrated to ~10 mg/ml (Bradford assay [20] using bovine serum albumin as standard) and placed in 20 mM HEPES pH 7.5/16 mM NaCl/1mM DTT (storage buffer) by centrifugal ultrafiltration (Amicon Ultra, from Millipore). The Pro63Leu and Arg210Gln mutants were purified in the same way.

PipY crystallization, X-ray diffraction and structure determination.

Crystals were grown by vapour-diffusion (21°C) in 0.8 µl sitting drops made of 1/1 mixtures of protein solution and crystallization solution. The protein solution contained 12.6 mg/ml PipY in storage buffer supplemented with either 10 mM DCS (PipY-Apo) or 10 mM D-Ala (PipY-PLP). The crystallization solution was, for PipY-Apo, 0.1 M Tris-HCl pH 8.5/0.2 M MgCl₂/25% w/v polyethylene glycol (PEG) 3350; and for PipY-PLP, 0.1 M Hepes pH 7.5/0.2 M calcium acetate/18% w/v PEG 8000. Crystals were flash-frozen in liquid nitrogen using as cryobuffers the crystallization solutions (enriched with 5% extra PEG 3350 for PipY-Apo). Synchrotron (Table 1) X-ray diffraction was at -170°C, using Pilatus3 6M detectors. Datasets at 1.90/1.93 Å resolution (PipY-Apo/PipY-PLP) were processed with XDS [21], scaled with AIMLESS [22], (Table 1) and phased

by molecular replacement with Phaser [23], using for the PipY-Apo dataset a poly-alanine model of YggS from *E.coli* (PDB entry 1W8G) lacking residues 1-20; and utilizing for PipY-PLP the A subunit of PipY-Apo once it was traced and refined. Refmac5 [24] was used for optimizing the positioning of the two PipY chains present in the asymmetric units of both crystals (Table 1), using rigid body refinement. It was also used for model refinement, alternating iteratively cycles of restrained refinement and of manual model building with Coot [25]. This last program was used for incorporating PLP (taken from the PDB) to the PipY-PLP model. Isotropic B factors and Translation/Libration/Screw (TLS) were used in the last refinement steps [26], choosing the TLS groups with the TLSMD server (<http://skuld.bmsc.washington.edu/~tlsmd>) [27]. All diffraction data were used throughout the refinement processes, except for the randomly selected 5% of the data used for R_{free} calculation. The stereochemistry of the model was checked and improved with PDB_REDO [28]. Analysis with Rampage [29] of the geometry of the main chain torsion angles revealed excellent values for both models. Graphical representations of the structures were generated with PyMOL (Pymol Molecular Graphics System, Version 1.6, Schrödinger, LLC, www.pymol.org).

Other techniques

Size exclusion chromatography was performed on a Superdex 200 10/300 column fitted on an ÄKTA FPLC (both from GE Healthcare), at 0.5 ml/min flow rate, using a solution of 10 mM Na phosphate pH 7.0/0.4 M NaCl/5 mM MgCl_2 /5 mM EDTA/0.2 mM DTT, and monitoring the absorbance of the effluent at 280 nm. The mass of PipY was estimated from a semi-logarithmic plot prepared with protein standards of known masses.

Table 1. X-Ray diffraction data and refinement statistics of PipY models

Crystal	PipY-PLP (PDB 5NM8)	PipY-Apo (PDB 5NLC)
Data collection		
Synchrotron/Beamline	Diamond/ I03	ALBA/ BL13 - XALOC
Wavelength (Å)	0.9763	0.9792
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁
Unit cell a, b, c (Å)	46.6, 98.9, 103.7	67.6, 41.8, 77.8
α, β, γ, (°)	90, 90, 90	90, 108.2, 90
Solvent (%)	50.1	37.7
Matthews coefficient	2.43	1.97
Resolution range ^a (Å)	71.60-1.93 (1.97-1.93)	73.89-1.90 (1.94-1.90)
Reflections ^a	469259/ 37132	150200 / 32801
total/ unique	(26106 / 2409)	(7964 / 2072)
Completeness ^a (%)	99.9 (98.7)	99.7 (99.2)
Multiplicity ^a	12.6 (10.8)	4.6 (3.8)
I/σ ^a	22.3 (4.5)	20.6 (4.0)
CC _{1/2} ^a (%)	99.9 (87.3)	99.8 (93.5)
Wilson B-factor (Å ²)	30.1	21.0
R _{pim} ^{a,b} (%)	2.0 (19.4)	2.3 (17.0)
Refinement		
Resolution ^a (Å)	28.39-1.93 (1.97-1.93)	44.61-1.90 (1.94-1.90)
Reflections, work/test	35275 / 1857	31161 / 1640
R _{factor} ^c /R _{free} ^d (%)	19.4 / 23.0	15.7 / 20.5
r.m.s.d. from ideal value		
Bond length (Å)	0.0096	0.0180
Bond angle (°)	1.4116	1.8560
Number of:		
Polypeptide chains	2	2
Protein atoms	3322	3358
Water molecules	162	375
PLP molecules	2	0
Average B-factor (Å ²)		
Protein	46.0	26.3
H ₂ O	44.8	34.7
PLP	49.6	-
Ramachandran plot ^e (%)		
Favoured	98.8	97.9
Allowed	1.2	2.1
Outliers	0.0	0.0

^aValues in parentheses are the data for the highest-resolution shell.

^b $R_{pim} = \sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \langle I_{hkl} \rangle$.

^c $R_{factor} = \sum_{hkl} | |F_{obs}| - |F_{calc}| | / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

^d R_{free} is the same as R_{factor} but calculated for the 5% data omitted from the refinement.

^eCalculated with RAMPAGE [29]

Spectrophotometric determinations were carried out with a NanoDrop 1000 spectrophotometer (Thermo Scientific) on 2 μ l samples. All reagents and PipY were in storage buffer. Optical absorption values are given for 1-cm light path. Incubations of PipY with PLP or DCS were performed for 15 min in 50 μ l solutions at 25 °C, with controls lacking PipY carried in parallel and subtracted. For incubation with DCS, PipY had been incubated previously with 0.34 mM PLP and then freed from unbound PLP by centrifugal gel filtration [30]. Plots of the absorbance at 425 nm of PipY versus PLP or DCS concentrations were fitted to hyperbolae using GraphPadPrism (GraphPad Software, San Diego, CA).

CD spectra in the far-UV region (195–250 nm) were obtained at 21 °C with a Jasco J-810 spectropolarimeter in 0.2 ml of 50 mM Na phosphate pH 7.5 and 0.25 mg/ml protein in a 0.1 cm-pathcell. Each spectrum was the average of five scans.

The thermal stabilities of wild-type and mutant PipY proteins were investigated by the thermofluor approach [31] as previously reported [32], using 0.25 mg/ml PipY in a solution of 22 mM Hepes pH 7.5 with the indicated additions.

RESULTS

The *pipY* protein product

PipY was produced in *E. coli* from an expression plasmid and was obtained in pure form (Fig. 1A, inset). Similarly to the *E. coli* and yeast *COG0325* gene products, YggS and YBL036C, [12,13,19], PipY was monomeric, judged from its behaviour in size exclusion chromatography assays (Fig. 1A). As expected for a protein forming a protonated Schiff base with PLP [33], pure PipY was yellow due to a broad optical absorption peak having its maximum absorption at ~425 nm (Fig. 1B, black continuous line), a longer wavelength than the maximum for

free PLP (Fig. 1B, dotted line). PipY incubation with excess PLP, followed by removal of the PLP excess by centrifugal gel filtration [30] increased the size of the 425-nm peak (Fig. 1B, dashed line), as if the pure protein was incompletely saturated with PLP, perhaps because of partial dissociation of PLP during protein isolation and storage. In fact, prolonged PipY storage (several months) decreased the height of the 425-nm peak (Fig. 1B, grey line), as expected for PLP dissociation with time. Titration with PLP of this aged protein restored maximal absorbance with a K_D value for PLP of $\sim 30 \mu\text{M}$ (Fig. 1C). As is typical for protein-bound PLP [34], the 425-nm peak disappeared with the addition of DCS (Fig. 1D), while absorbance at shorter wavelength (around 335 nm) increased. Titration of the decrease in absorbance at 425 nm versus DCS concentrations (Figs. 1D and 1E) yielded a K_D for DCS in the mM range, two to three orders of magnitude higher than K_D values for PLP-containing proteins that bind DCS as an analogue of an amino acid substrate (the case for alanine racemase [15]). This low affinity for DCS excludes specific binding of this antibiotic to PipY, in agreement with the reported lack of alanine racemase activity of PipY [11] and of the *COG0325* gene products of other species [13].

Overall characteristics of crystalline PipY

PipY crystal structures were determined at $\sim 1.9 \text{ \AA}$ resolution with and without PLP (PipY-PLP and PipY-Apo, respectively; Table 1). No electron density that could correspond to D-Ala or DCS was observed in the crystals, despite the fact that these compounds were present at 10 mM concentrations in the respective crystallization solutions (see Materials and Methods). Each crystal had two protein chains in the asymmetric unit. These chains were related by twofold non-crystallographic symmetry. The small contact surfaces ($\sim 5\%$ of the exposed surface) and the predictions of the PISA server (www.ebi.ac.uk/pdbe/pisa/) did not support a homodimeric nature of PipY in

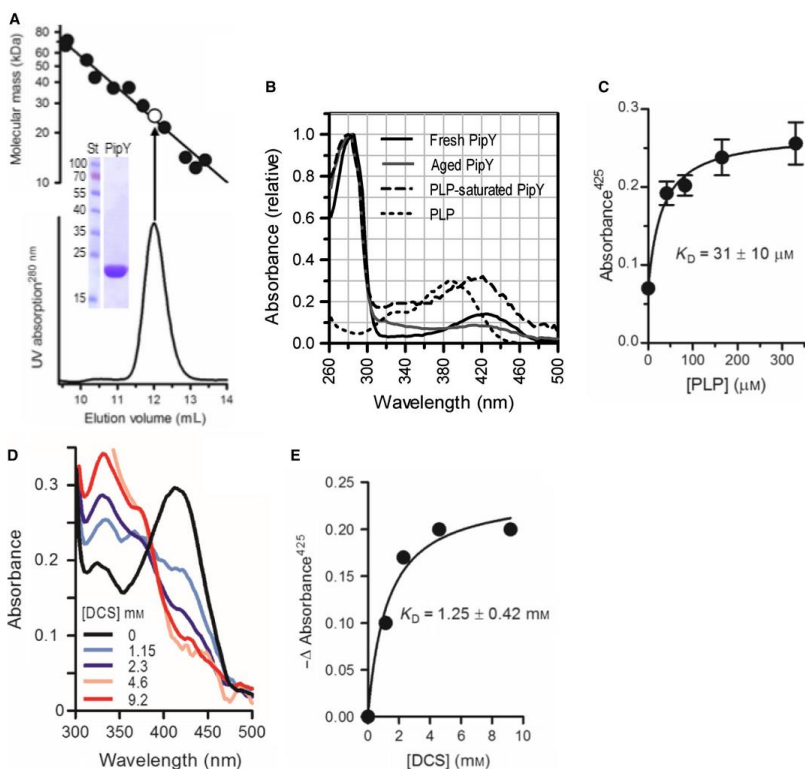


Figure 1- PipY properties. (A) Size exclusion chromatography of pure PipY with elution monitored as UV absorption (bottom) and plotted semilogarithmically (top) together with the elution positions of molecular mass standards (closed circles). The open circle, corresponding to PipY, has been placed at the intersection of the PipY elution position and its sequence-deduced mass. The protein standards used and their masses (in parenthesis, in kDa) were *Enterococcus faecalis* carbamate kinase (71.3) [44], bovine serum albumin (66.4), *Escherichia coli* N-acetyl-L-glutamate kinase (54.3) [45], chicken ovalbumin (42.7), *Synechococcus elongatus* PII (37.2) [46], *Pseudomonas aeruginosa* PII (36.9) (our own unpublished data), bovine erythrocyte carbonic anhydrase (29), soybean trypsin inhibitor (21.5), cow's lactalbumin (14.2), pancreatic ribonuclease A (13.7) and horse heart cytochrome C (12.3). *Inset*, SDS/PAGE of pure PipY (St, protein standards of the indicated masses). (B) Optical absorption spectrum of PipY, freshly prepared (black continuous line), after aging for several months (gray continuous line), or following a 15-min incubation of $2 \text{ mg}\cdot\text{mL}^{-1}$ PipY with 0.34 mM PLP, followed by excess PLP removal by centrifugal gel filtration [30] (broken line, large trace). Spectra are normalized relative to the absorption at 280 nm . For comparison, the spectrum of PLP, normalized to a maximum absorbance of 0.3, is shown (dotted line). (C) Influence of added PLP on the height of the protein-bound PLP peak of $0.66 \text{ mg}\cdot\text{mL}^{-1}$ of aged PipY. (D,E) Influence of variable DCS concentrations on the absorption spectrum of PipY ($0.91 \text{ mg}\cdot\text{mL}^{-1}$, incubated with 0.34 mM PLP and then freed from unbound PLP by centrifugal gel filtration) in the $300\text{--}500 \text{ nm}$ range (D) or at 435 nm (E), fitting the data to a hyperbola.

solution, agreeing with the size exclusion chromatography results that showed that PipY is a monomer (Fig. 1A). Except for the initial and final 2-3 residues and the C-terminal His₆-tag, the entire polypeptide chains were visible in the crystal structures. In addition, one PLP molecule was found binding per chain in the PipY-PLP crystal. Both monomers in each asymmetric unit were essentially identical (r.m.s.d. <0.6 Å, Supplementary Table S1), with somewhat lower identity when PLP-containing and PLP-lacking PipY molecules were compared (r.m.s.d. 0.62-0.98 Å, Supplementary Table S1). There was high similarity with the reported structure [19] of the yeast *COG0325*-encoded protein YBL036C or with the structures of another three *COG0325* products from other bacteria that are deposited in the PDB (Supplementary Table S1; r.m.s.d. 1.29-1.75 Å; and Supplementary Fig S2A).

The structure of PipY

PipY has (Fig. 2A) the characteristic type-III fold of PLP-dependent enzymes [19,35], a modified TIM barrel [36] having an extra N-terminal α -helix preceding the first of the eight β/α repeating units and lacking the helix of the last repeat. Unlike the paradigms of this fold, alanine racemase, ornithine decarboxylase and broad specificity amino acid racemase [37-39], all of them composed of two-domain subunits forming homodimers, PipY is a single-domain monomer consisting exclusively of the barrel domain (Supplementary Fig. S2A). This is also the case for the highly similar structures previously reported for YBL036C [19] or deposited in the PDB for another three bacterial *COG0325* proteins (Supplementary Table S1). All these *COG0325* protein structures have a one-turn imperfect α helical extension of the C-terminal β -strand (helix 9, Fig. 2A and Supplementary Fig. S2A). This extension binds the phosphate of PLP (see below). However, PipY differs from all the other *COG0325*-encoded protein products of known structure in the length of the first α -helix (Supplementary

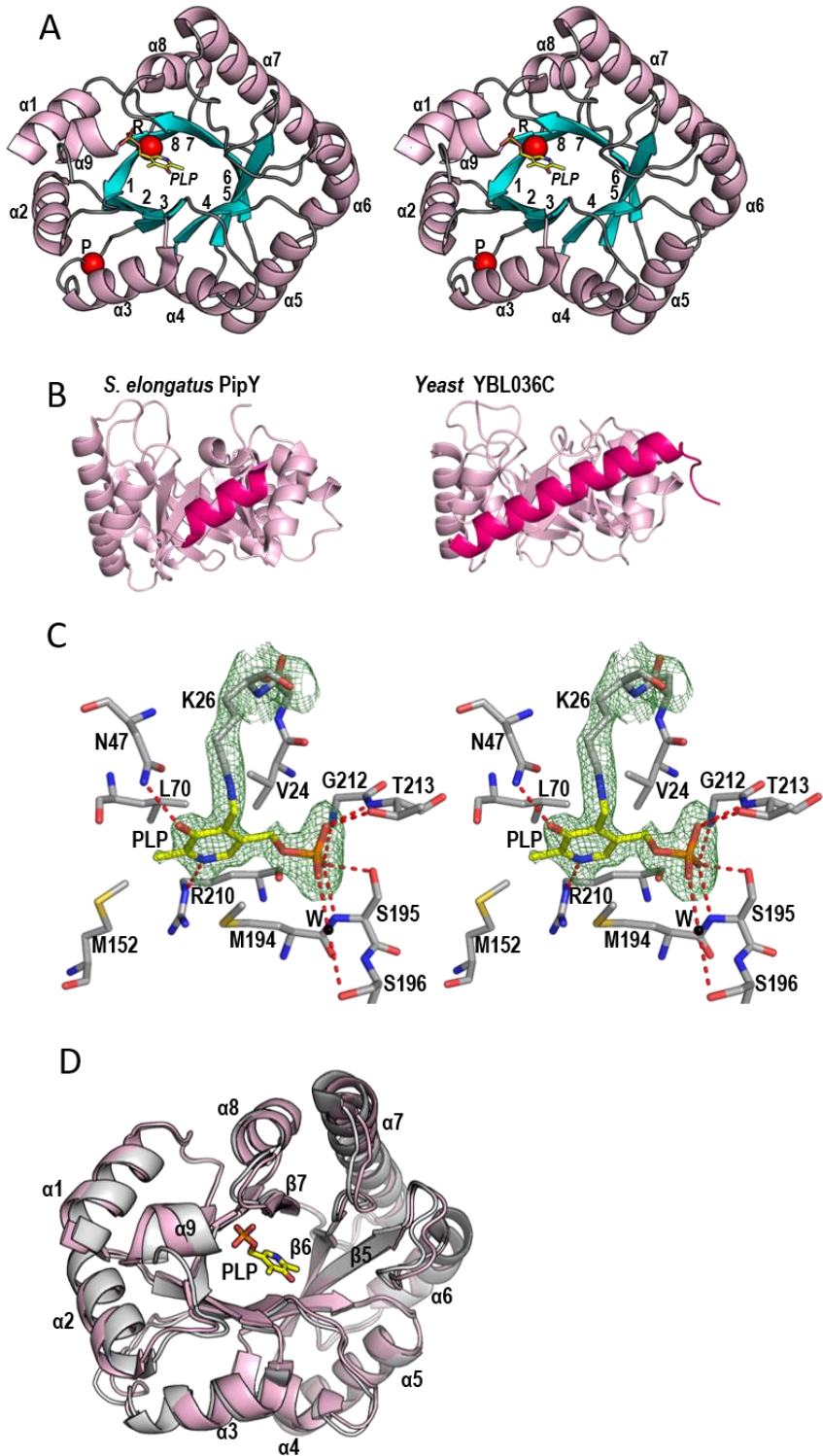


Figure 2 (previous page). PipY structure. Panels A and C are stereo views. Beta strands are labeled with numbers. Helices are numbered also with numbers preceded by an α . When using sticks representation (always the case for PLP), O, N, S, and P atoms are colored red, blue, yellow, and orange, respectively, and C atoms are yellow for PLP and gray for the protein. (A) Cartoon representation of the structure of PipY complexed with PLP. Helices are pink and strands are blue. Red spheres mark the locations of Pro63 and Arg210, which correspond to human PROSC Pro87 and Arg241. (B) Cartoon side views (90° rotation around the X axis, relative to the view in panel A, with the C-edge of the β sheet pointing to the top) of PipY and the yeast *COG0325* gene product YBL036C (PDB [1CT5](#)), to highlight the differences in the length of helix 1, which is deeply colored. (C) The PLP binding site. Side-chains are shown in sticks and labeled in single letter notation. The grid around the PLP-Lys26 adduct shows the Fo-Fc omit electron density map for this adduct at 2.5σ . Broken red lines represent hydrogen bonds of PLP with the protein (D) PipY movements upon PLP binding shown by superimposing PipY-Apo (pink) and PipY-PLP (gray) in cartoon representation. Selected secondary structure elements are labeled.

Fig. S2), which in PipY spans less than three full turns, while in all these other proteins is much longer on both the N- and C-terminal ends of this helix, which spans up to 6 turns (Fig. 2B). Perhaps this helix, which is lacking in canonical TIM barrels, could be involved in interactions with other proteins, with specific requirements for COG0325 proteins of different species.

PLP binding

PLP sits towards the centre of the barrel, on the C-edge side of the parallel β -sheet (Fig. 2A). It is found in the same site and orientation as in yeast YBL036C and in the deposited structures of COG0325 proteins. This location is the same that is found in the PLP enzymes presenting the PLP type III fold (Supplementary Fig. S2A-C). However, in PipY and the other COG0325 proteins of known structure the PLP is much more accessible than in the type III fold PLP enzymes [37-39]. In these dimeric enzymes the cofactor is shielded by the two-domain architecture and by the other subunit, being accessible only through a tunnel (not shown). The exposed PLP of PipY and of other COG0325 proteins should facilitate PLP exchange with other molecular partners. The PLP forms a

Schiff base with Lys26 (Fig. 2C; and Supplementary Fig. S2C). Its phosphate is anchored via six hydrogen bonds on helix 9 (residues Gly212 and Thr213) and on the $\beta 7$ - $\alpha 8$ turn (Ser195 and, indirectly through water, Ser196). The cofactor makes other non-covalent interactions with six of the eight β/α repeats of the modified TIM barrel. Only β/α repeats 4 and 5 do not contact the PLP. The pyridoxal ring is partially sandwiched between Leu70 from $\beta 3$ and Met194 from the $\beta 7$ - $\alpha 8$ turn, lying through its border on Arg210 from $\beta 8$, with which an NH \cdots N hydrogen bond is made. The exposed part of the pyridoxal ring makes contacts through its C2-methyl group with Met152 from $\beta 6$, hydrogen bonding its C3-OH with Asn47 from $\beta 2$, and contacting Val24 of $\beta 1$ via its C4 carbaldehyde and C5 hydroxymethyl groups. Except for the indirect interaction with Ser196 (Fig. 2C), all PLP interactions found in PipY are present in the other COG0325 structures, either with identical or conservatively replaced residues (Supplementary Fig.S2A).

Changes associated with PLP binding in the context of a role of PipY in PLP homeostasis

Our determination of the structures of PLP-bound and PLP-free forms of PipY allowed characterization of the changes associated with PLP binding (Fig. 2D). These changes are not large, as shown by the modest r.m.s.d. increase when comparing PipY-PLP and PipY-Apo (Supplementary Table S1). The largest changes are experienced by helix 9 (Fig. 2D), which is displaced towards the PLP phosphate that anchors on it. This helix drags in its movement the residues following it, which are displaced by up to 4.5 Å. Helices 1 and 2 are also dragged towards the phosphate, while Ser195 and the $\beta 7$ - $\alpha 8$ loop are displaced by the phosphate, with concomitant small displacement of helix 8, which drags with it helix 7 and the $\beta 6$ - $\alpha 7$ loop, and even helix 6 and the $\beta 5$ - $\alpha 6$ loop.

The movement of helix 9 might be the best candidate for triggering the PLP exchange required for the proposed role of PipY in PLP homeostasis [11-14,17,18]. Due to its exposed nature helix 9 could interact with other macromolecular players, acting as a trigger for PLP exchange, since a change in the position of helix 9 could importantly weaken PLP binding to PipY, given the PLP-anchoring role of this helix.

Exploiting PipY for investigating the disease-causing mechanisms of PROSC mutations

Recently, vitamin B₆-dependent epilepsy was reported in association with null mutations in the human *COG0325* gene *PROSC* [17,18]. However, some patients carried the missense *PROSC* mutations Pro87Leu and/or Arg241Gln, two mutations for which a loss-of-function effect was at best hypothetical [18]. In addition, there was no correspondence between the degree of clinical severity attributed to each of these mutations on the basis of the observations on the patients [18] and the ability of each mutant form to complement the increased pyridoxine toxicity phenotype of the *yggS*⁻ *E.coli* mutant [17]. Since the *PROSC* Pro87 and Arg241 residues are conserved in *S.elongatus* PipY (respective PipY counterparts, Pro63 and Arg210, Supplementary Fig. S1), we investigated the effects of the corresponding PipY mutations on recombinantly produced PipY as an approximation to the effects of *PROSC* mutations. Both mutant proteins could be purified (Fig. 3A, bottom panel) and were monomeric (determined by size exclusion chromatography, data not shown). The Pro63Leu mutation did not negatively affect any of the parameters assessed: yield (Fig. 3A, top panel), thermal stability (Fig. 3B), proper fold of the pure protein (estimated by CD, Fig. 3C) and PLP content (assessed by the absorption spectrum, Fig. 3D). The latter was to be expected, since Pro63 sits on the opposite face of the TIM barrel than PLP (Fig. 2A) and, therefore, the mutations

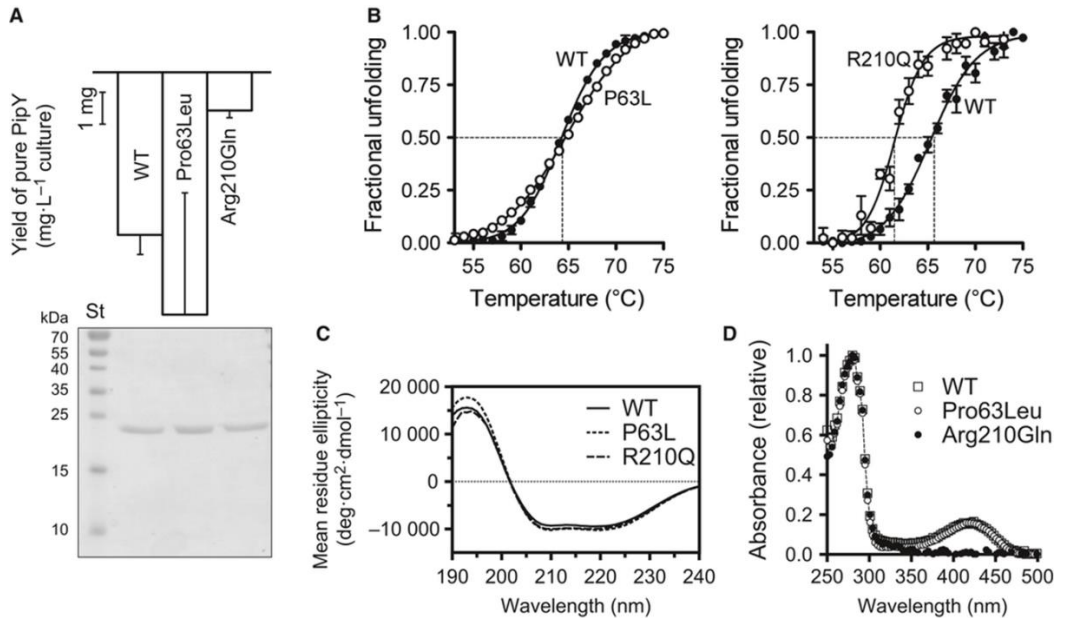


Figure 3- Comparisons between wild-type and Pro63Leu and Arg210Gln mutants of PipY. (A) Yield (top; mean \pm SE; $n \geq 2$) and purity (bottom; coomassie-stained SDS/PAGE, 15% gel) of the final protein preparations. (B) Thermal stability of the wild-type (closed circles) and the two mutants (open circles), estimated in thermofluor assays (see Materials and Methods). Buffer, pH, and protein concentrations were the same for both panels, but the solution contained, in addition, 0.5 m NaCl and 20 mm imidazole (left panel) or 16 mm NaCl and 1 mm DTT (right panel). The broken lines mark for each protein form the temperature at which the fluorescence change is 50% of the maximum. (C) CD spectra to exclude gross misfolding of the mutants. (D) Optical absorption spectrum to show that the Arg210Gln mutant lacks the peak of bound PLP. Spectra are the average of ≥ 2 preparations. They are normalized relative to the absorption at 280 nm.

of this residue should not directly hamper PLP binding. In contrast, the Arg210Gln mutant, also obtained in pure (Fig. 3A, bottom panel) and well-folded form (Fig. 3C) was produced in quite low yield (Fig. 3A, top panel), exhibited decreased thermal stability (nearly 5 °C decrease, Fig. 3B) and had little if any PLP, judged from its low optical absorption in the 425 nm region (Fig. 3D). This last observation fits the fact that the mutation affects the invariant arginine shown by the structure to run parallel to the buried border of

the pyridoxal ring, with which it makes an NH··N hydrogen bond (Fig. 2A,C). Our findings explain the results of complementation assays for the Pro87Leu and Arg241Gln PROSC mutants [17] and the disease-causality of the Arg241Gln mutation (see Discussion)

DISCUSSION

The complementation of *yggS*⁻ *E.coli* by *COG0325* genes of other species [12,13,17] and the phenotypic similarities of this *E.coli* mutant and of the *S.elongatus pipY*⁻ mutant [11] are strong indications that the function of *COG0325* genes is conserved across phyla. The alleviation by PLP or by vitamin B₆ supplementation of the alterations observed in these mutants [11-14], best illustrated by the vitamin B₆ dependence of the epilepsy due to PROSC deficiency [17,18], indicates that this conserved function is related to PLP, in line with the fact that *COG0325* products are PLP-containing proteins. Yet, previous work [11-13] has not favoured an enzymatic role of *COG0325* proteins. In our own work, we have failed to generate crystals of PipY bound to D-Ala or to DCS despite extensive crystallization attempts with these compounds. These two ligands have been found binding in crystals of the structurally similar alanine racemase [40]. Furthermore, as exemplified in the PipY structure, *COG0325* proteins lack key constant and obligatory features of type III-fold PLP enzymes [35]. These missing traits include the dimeric character, the presence of a second domain and the existence of an invariant substrate-binding arginine (Arg129 of Ala racemase, Supplementary Figs. S2A,C) and of a negative charge that stabilizes this arginine [37-41]. Thus, *COG0325* proteins are believed to be involved in PLP homeostasis rather than being catalysts [11-14,17,18].

A PLP-homeostatic role would be instrumental to prevent the toxicity of free B₆-vitamers [13,42, and see 17] and would imply that *COG0325* proteins deliver

PLP to the apoenzymes of PLP-dependent enzymes. Such role fits: i) the observations of increased toxicity of pyridoxine with *yggS*⁻ *E. coli* [12,13] and of pyridoxine and PLP-targeting antibiotics with *pipY*⁻ *S. elongatus* [11]; ii) the interactions of these traits with the supply of amino acids to these bacterial mutants [11-13]; iii) the observation of pleiotropic amino acid metabolism changes in *COG0325* mutants [12-14,16-18]; iv) the synthetic lethality caused in *S. elongatus* by the *pipY/cysK* double mutation [11]; and v) particularly, the finding of amino acid levels in the cerebrospinal fluid of a patient with PROSC deficiency that are highly suggestive of multiple transaminase deficiency [18]. In line with a homeostatic role of PipY, in the structures of PipY (present data) and of other *COG0325* proteins ([19], and see PDB files given in Supplementary Table S1) PLP is exposed as if it was poised for transfer. Furthermore, the relatively large changes in the orientation of the outwards-protruding helix 9 depending on the presence or absence of PLP, taken together with the key role of this helix in PLP anchoring, renders this helix an obvious candidate for being a trigger that could be operated by partners to weaken the interactions of *COG0325* proteins with PLP, promoting PLP transfer.

The alternative possibility that *COG0325* proteins, although not catalytic by themselves, could be subunits of hypothetical heterodimeric PLP enzymes in which the *COG0325* subunit would act as a common PLP-containing module appears unlikely. The pleiotropic metabolic effects of the null *COG0325* gene mutations demand multiple alternative partners for the *COG0325* protein. It would be unlikely not to detect any of these partners in yeast-two-hybrid assays using PipY as the bait, but, in fact, no partner was detected [11]. Furthermore, the increased amino acid levels in the cerebrospinal fluid of a PROSC-deficient patient [18] fitted multiple transaminase deficiency. However, transaminases do not utilize for binding PLP the modified TIM barrel fold used

by COG0325 proteins [35]. Similarly, a role of COG0325 in the regulation of a specific PLP enzyme, as exemplified for ornithine decarboxylase with its enzyme/antizyme/antizyme inhibitor system [43], cannot be excluded, but, again, appears unlikely, largely because it would not fit the multi-target pleiotropy reflected in the metabolic phenotypes of COG0325 mutants.

Strong support for a PLP homeostatic role of COG0325 proteins was provided by observations made in vitamin B₆-dependent epilepsy associated to PROSC mutations [17,18]. By using PipY as a model of PROSC, we attempted here to shed light on why the PROSC Pro87Leu mutant complemented and the Arg241Gln mutant did not complement the pyridoxine toxicity phenotype of an *E.coli* *yggS*⁻ strain [17]. The abundant production in *E.coli* of PLP-carrying Pro63Leu PipY could explain the complementation observed with the PROSC Pro87Leu mutant [17]. In contrast, lack of complementation would be expected for the poorly produced and PLP-lacking Arg210Gln mutant of PipY, as observed for the corresponding human PROSC mutant Arg241Gln [17], in line with the prior observation that there was no complementation by YggS when it carried a mutation that made it unable to bind PLP [13]. Concerning disease-causation and clinical severity, the lack of bound PLP and the decreased stability and yield of PipY carrying the Arg210Gln mutation clearly support a disease-causing role of the corresponding Arg241Gln PROSC mutation, in agreement with the finding of this last mutation in patients with vitamin B₆-dependent epilepsy. However, the PROSC deficiency due to the Arg241Gln mutation may be partial if the mutant can still bind PLP at increased concentrations of this cofactor, what could explain the relatively mild clinical severity associated with this mutation [18]. In contrast to the Arg210Gln PipY mutant, the PipY Pro63Leu mutant that mirrors the PROSC Pro87Leu mutant appeared normal. Thus, studies using the PROSC Pro87Leu mutant itself expressed in human cells might be necessary to test experimentally the disease-causation and clinical severity

attributed to this mutation [18]. In any case, it is interesting that three of the six missense PROSC mutations observed thus far in patients with vitamin B6-dependent epilepsy replace proline by leucine or the reverse (Supplementary Fig. S2A) [17,18]. Given the structural uniqueness of proline as a protein amino acid, it is conceivable that these changes could lead to increased PROSC misfolding, raising the possibility of using pharmacological chaperones to try to alleviate the deleterious effects of these missense mutations.

PROTEIN DATABANK FILES

Protein DataBank accession codes 5NM8 and 5NLC

ACKNOWLEDGEMENTS

We thank MA Castells (UA), JL Llacer, S. Ventas (IBV-CSIC) and N. Gougard (CIBERER-IBV) for technical support, A. García and M. Orzáez (CIPF-Valencia) for CD facilities, and the IBV-CSIC crystallographic facility for crystal growth. Supported by grants from the Valencian Government (PrometeoII/2014/029) and Spanish Governments (BFU2014-58229-P to V.R.; BFU2012-33364 and BFU2015-66360-P to A.C.; FPI contract to LT) and to EC FP7/2007-2013 BioStruct-X (grant agreement N°283570, proposal 7687) for synchrotron access. We are grateful to Diamond (Oxfordshire, UK) and Alba (Barcelona, Spain) synchrotrons for access and for staff support. The authors declare no conflict of interest.

REFERENCES

1. Margulis L, Schwartz KV (1982) Five kingdoms. An illustrated guide to the phyla of life on earth. WH Freeman & Company, San Francisco.

2. Garcia-Pichel F, Belnap J, Neuer S, Schanz F (2003) Estimates of global cyanobacterial biomass and its distribution. *Algol Stud* 109:213–217
3. Hamilton TL, Bryant DA, Macalady JL (2016) The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environ Microbiol* 18:325-340.
4. Luque I, Forchhammer K (2008) Nitrogen assimilation and C/N balance sensing. In *The Cyanobacteria: Molecular Biology, Genomics and Evolution* (Herrero A and Flores E, eds.). Caister Academic Press, Poole, UK .
5. Llácer JL, Espinosa J, Castells MA, Contreras A, Forchhammer K, Rubio V (2010) Structural basis for the regulation of NtcA-dependent transcription by proteins PipX and PII. *Proc Natl Acad Sci U S A* 107:15397-15402.
6. Forcada-Nadal A, Palomino-Schätzlein M, Neira JL, Pineda-Lucena A, Rubio V. (2017) The PipX protein, when not bound to its targets, has its signaling C-terminal helix in a flexed conformation. *Biochemistry*, in press.
7. Espinosa J, Rodríguez-Mateos F, Salinas P, Lanza VF, Dixon R, de la Cruz F, Contreras A. (2014) PipX, the coactivator of NtcA, is a global regulator in cyanobacteria. *Proc Natl Acad Sci U S A* 111:E2423-E2430.
8. Burillo S, Luque I, Fuentes I, Contreras A (2004) Interactions between the nitrogen signal transduction protein PII and N-acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J Bacteriol* 186:3346-3354.
9. Espinosa J, Forchhammer K, Burillo S, Contreras A (2006) Interaction network in cyanobacterial nitrogen regulation: PipX, a protein that interacts in a 2-oxoglutarate dependent manner with PII and NtcA. *Mol Microbiol* 61:457-469.
10. Labella JI, Obrebska A, Espinosa J, Salinas P, Forcada-Nadal A, Tremiño L, Rubio V, Contreras A. (2016) Expanding the cyanobacterial nitrogen regulatory network: The GntR-like regulator PlmA interacts with the PII-PipX complex. *Front Microbiol* 7:1677.

11. Labella JI, Cantos R, Espinosa J, Forcada-Nadal A, Rubio V, Contreras A (2017) PipY, a member of the conserved COG0325 family of PLP-binding proteins, expands the cyanobacterial nitrogen regulatory network. *Front Microbiol*, 8:1244.
12. Prunetti L, El Yacoubi B, Schiavon CR, Kirkpatrick E, Huang L, Bailly M, ElBadawi-Sidhu M, Harrison K, Gregory JF, Fiehn O, Hanson AD, de Crécy-Lagard V (2016) Evidence that COG0325 proteins are involved in PLP homeostasis. *Microbiology* 162:694-706.
13. Ito T, Iimori J, Takayama S, Moriyama A, Yamauchi A, Hemmi H, Yoshimura T (2013) Conserved pyridoxal protein that regulates Ile and Val metabolism. *J Bacteriol* 195:5439-5449.
14. Ito T, Yamauchi A, Hemmi H, Yoshimura T. (2016) Ophthalmic acid accumulation in an *Escherichia coli* mutant lacking the conserved pyridoxal 5'-phosphate-binding protein YggS. *J Biosci Bioeng* 122:689-693.
15. Neuhaus FC (1967) D-Cycloserine and O-carbamyl-D-serine. In *Antibiotics I. Mechanism of action* (Gottlieb D and Shaw PD, eds.) Springer Verlag Berlin Heidelberg, pp 40-83.
16. Kolodkin-Gal I, Romero D, Cao S, Clardy J, Kolter R, Losick R. (2010) D-Amino acids trigger biofilm disassembly. *Science* 328:627-629.
17. Darin N, Reid E, Prunetti L, Samuelsson L, Husain RA, Wilson M, El Yacoubi B, Footitt E, Chong WK, Wilson LC, Prunty H, Pope S, Heales S, Lascelles K, Champion M, Wassmer E, Veggiotti P, de Crécy-Lagard V, Mills PB, Clayton PT. (2016) Mutations in PROSC disrupt cellular pyridoxal phosphate homeostasis and cause vitamin-B6-dependent epilepsy. *Am J Hum Genet* 99:1325-1337.
18. Plecko B, Zweier M, Begemann A, Mathis D, Schmitt B, Striano P, Baethmann M, Vari MS, Beccaria F, Zara F, Crowther LM, Joset P, Sticht H, Papuc SM, Rauch A (2017) Confirmation of mutations in PROSC as a novel cause of vitamin B6-dependent epilepsy. *J Med Genet pii: jmedgenet-2017-104521*.

19. Eswaramoorthy S, Gerchman S, Graziano V, Kycia H, Studier FW, Swaminathan S. (2003) Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta Crystallogr D Biol Crystallogr* 59:127-135.
20. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal Biochem* 72:248–254.
21. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125-132.
22. Evans PR and Murshudov GN (2013) How good are my data and what is the resolution? *Acta Cryst D Biol Crystallogr* 69:1204–1214.
23. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
24. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355-367.
25. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
26. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 374:300–321.
27. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr* 39:109–111.
28. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund AC, Blanchet C, Bongcam-Rudloff E, Combet C, Da Costa AL, Deleage G, Diarena M, Fabbretti R, Fettahi G, Flegel V, Gisel A, Kasam V, Kervinen T, Korpelainen E, Mattila K, Pagni M, Reichstadt M, Breton V, Tickle IJ, Vriend G (2010) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr* 42:376-384.

29. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* 50:437–450.
30. Penefsky HS (1977) Reversible binding of Pi by beef heart mitochondrial adenosine triphosphatase. *J Biol Chem* 252:2891-2899.
31. Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD, Nordlund P, Sundstrom M, Weigelt J, Edwards AM (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* 103:15835-15840.
32. Sancho-Vaello E, Marco-Marín C, Gougéard N, Fernández-Murga L, Rüfenacht V, Mustedanagic M, Rubio V, Häberle J (2016) Understanding N-acetyl-L-glutamate synthase deficiency: Mutational spectrum, impact of clinical mutations on enzyme functionality, and structural considerations. *Hum Mutat* 37:679-694.
33. Fasella P (1967) Pyridoxal phosphate. *Annu Rev Biochem* 36:185-210.
34. Cook SP, Galve-Roperh I, Martínez del Pozo A, Rodríguez-Crespo I (2002) Direct calcium binding results in activation of brain serine racemase. *J Biol Chem* 277:27782-27792.
35. Percudani R, Peracchi A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep* 4:850-854.
36. Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE, Priddle JD, Waley SG. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5Å resolution: using amino acid sequence data. *Nature* 255:609-614.

37. Shaw JP, Petsko GA, Ringe D (1997) Determination of the structure of alanine racemase from *Bacillus stearothermophilus* at 1.9-Å resolution. *Biochemistry* 36:1329-1342.
38. Kern AD, Oliveira MA, Coffino P, Hackert ML (1999) Structure of mammalian ornithine decarboxylase at 1.6 Å resolution: stereochemical implications of PLP-dependent amino acid decarboxylases. *Structure* 7:567-581.
39. Espaillat A, Carrasco-López C, Bernardo-García N, Pietrosemoli N, Otero LH, Álvarez L, de Pedro MA, Pazos F, Davis BM, Waldor MK, Hermoso JA, Cava F (2014) Structural basis for the broad specificity of a new family of amino-acid racemases. *Acta Crystallogr D Biol Crystallogr* 70:79-90.
40. Wu D, Hu T, Zhang L, Chen J, Du J, Ding J, Jiang H, Shen X. (2008) Residues Asp164 and Glu165 at the substrate entryway function potently in substrate orientation of alanine racemase from *E. coli*: Enzymatic characterization with crystal structure analysis. *Protein Sci* 17:1066-1076
41. Watanabe A, Yoshimura T, Mikami B, Hayashi H, Kagamiyama H, Esaki N (2002) Reaction mechanism of alanine racemase from *Bacillus stearothermophilus*: X-ray crystallographic studies of the enzyme bound with N-(5'-phosphopyridoxyl) alanine. *J Biol Chem* 277:19166-19172.
42. Ghatge MS, Contestabile R, di Salvo ML, Desai JV, Gandhi AK, Camara CM, Florio R, González IN, Parroni A, Schirch V, Safo MK (2012) Pyridoxal 5'-phosphate is a slow tight binding inhibitor of *E. coli* pyridoxal kinase. *PLoS One* 7:e41680.
43. Wu HY, Chen SF, Hsieh JY, Chou F, Wang YH, Lin WT, Lee PY, Yu YJ, Lin LY, Lin TS, Lin CL, Liu GY, Tzeng SR, Hung HC, Chan NL (2015) Structural basis of antizyme-mediated regulation of polyamine homeostasis. *Proc Natl Acad Sci U S A* 112:11229-11234.
44. Marina A, Alzari PM, Bravo J, Uriarte M, Barcelona B, Fita I, Rubio V (1999) Carbamate kinase: New structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine. *Protein Sci* 8: 934-940.

45. Gil F, Ramón-Maiques S, Marina A, Fita I, Rubio V. (1999) N-Acetyl-L-glutamate kinase from *Escherichia coli*: cloning of the gene, purification and crystallization of the recombinant enzyme and preliminary X-ray analysis of the free and ligand-bound forms. *Acta Crystallogr D Biol Crystallogr* 55:1350-1352.
46. Llácer JL, Contreras A, Forchhammer K, Marco-Marín C, Gil-Ortiz F, Maldonado R, Fita I, Rubio V (2007) The crystal structure of the complex of PII and acetylglutamate kinase reveals how PII controls the storage of nitrogen as arginine. *Proc Natl Acad Sci USA* 104:17644-17649.

SUPPORTING INFORMATION (CAPÍTULO 1)

**Studies on cyanobacterial protein PipY shed light on structure,
potential functions and vitamin B6-dependent epilepsy**

**Lorena Tremiño¹, Alicia Forcada-Nadal^{1,2}, Asunción Contreras² and Vicente
Rubio^{1*}**

¹ Instituto de Biomedicina de Valencia (IBV-CSIC), and CIBER de Enfermedades Raras (CIBERER-ISCIH), Valencia, Spain.

² Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain.

Contains:

Supplementary Table S1

Supplementary Figure S1

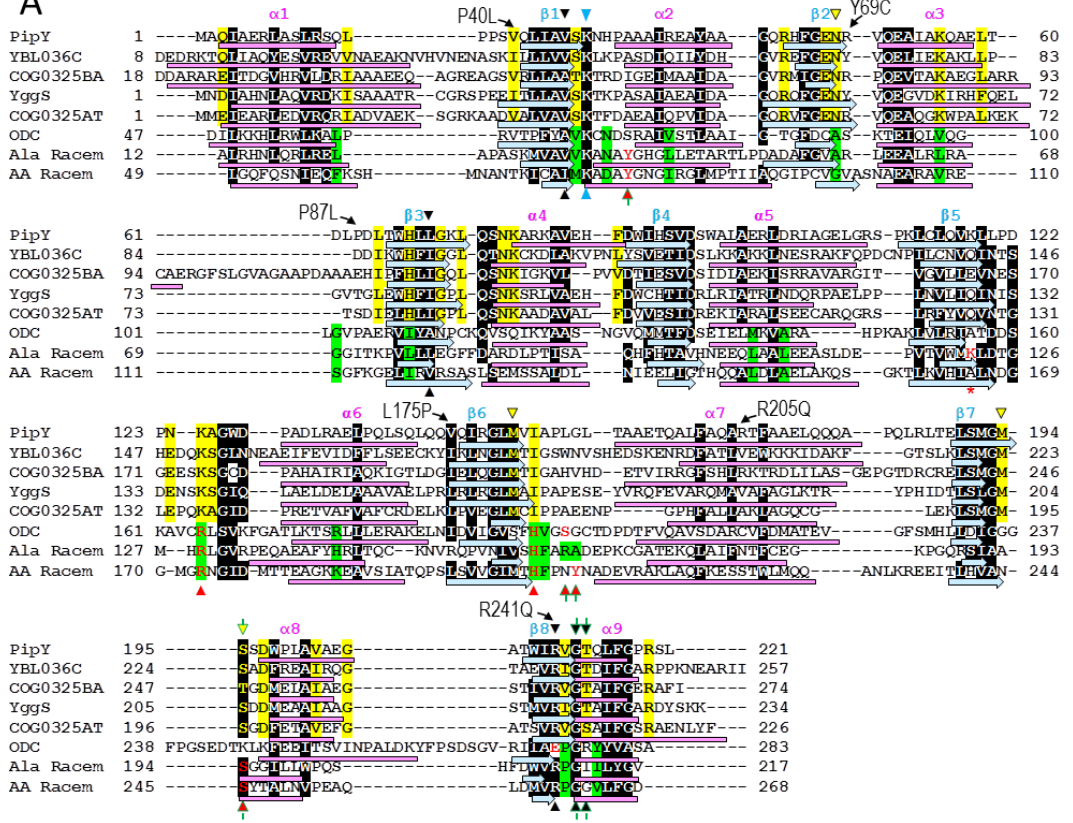
Supplementary Figure S2

Supplementary Table S1. Degree of similarity among the structures of PipY monomers studied here, and with the structures of other COG0325 gene products, and of TIM barrel domains of amino acid racemases and ornithine decarboxylase.

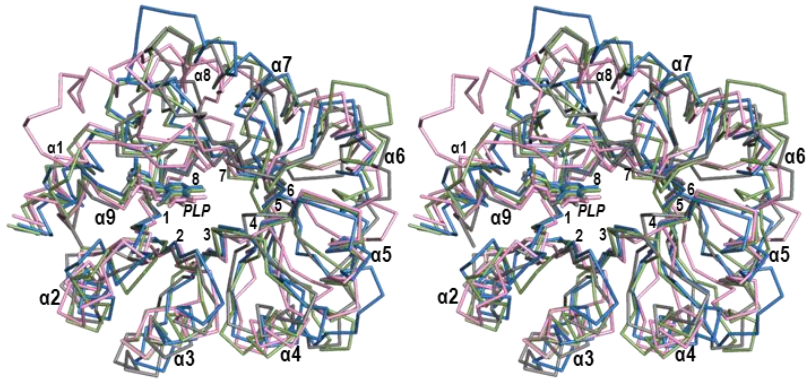
	Root mean square deviation for the superimposition of the indicated number of C α atoms ^a (Å / number C α atoms)			
	PipY-Apo A	PipY-Apo B	PipY-PLP A	PipY-PLP B
PipY-Apo B	0.59/215	-	-	-
PipY-PLP A	0.90/213	0.62/215	-	-
PipY-PLP B	0.98/210	0.70/212	0.39/212	-
COG0325 yeast YBL036C (1CT5)	-	-	1.63/195	1.61/194
COG0325 <i>E. coli</i> YggS (1W8G)	-	-	1.50/205	1.52/205
COG0325 <i>B. adolescentis</i> (3CPG)	1.75/202	1.69/202	-	-
COG0325 <i>A. tumefaciens</i> (3R79 A)	-	-	1.29/199	1.42/196
Alanine racemase (2RJG)	-	-	2.39/191	-
Broad spectrum AA racemase (4BEU)	-	-	-	2.31/200
Ornithine decarboxylase (7ODC)	-	-	2.60/186	-

^aThe two chains in the same asymmetric unit are distinguished by the letters A and B. *B. adolescentis*, *Bifidobacterium adolescentis*; *A. tumefaciens*, *Agrobacterium tumefaciens*; AA, amino acids. The alanine racemase, broad spectrum amino acid racemase and ornithine decarboxylase are from *E. coli*, *Vibrio cholerae* and mouse, respectively. Protein DataBank accession numbers are given between parentheses. Except in the case of the protein from *B. adolescentis*, which contained no PLP and thus was compared with the monomers of PipY-Apo, all other proteins contained PLP and have been compared with PipY-PLP. In the cases of racemases and ornithine decarboxylase only the result of the best superimposition of PipY with one subunit of the indicated enzymes is shown.

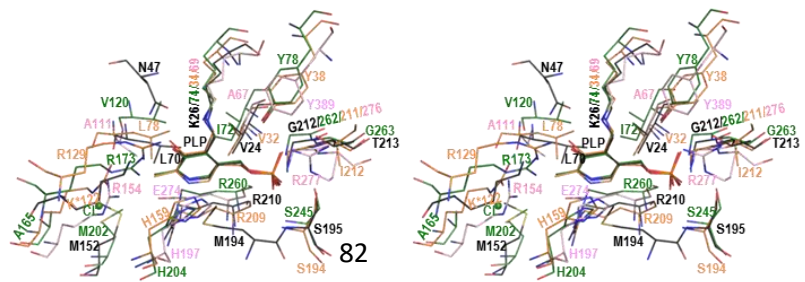
A



B



C



Supplementary Figure S2 (previous page). Sequence and structural comparisons of PipY with other proteins. **(A)** Structure-aided sequence alignment of the five *COG0325* gene products of known structure (*S. elongatus* PipY and the products of yeast, *Bifidobacterium adolescentis*, *E. coli* and *Agrobacterium tumefaciens*, abbreviated PipY, YBL036C, COG0325BA, YggS and COG0325AT, with PDB file identifiers 5NM8, 1CT5, 3CPG, 1W8G and 3R79, respectively) and of the TIM barrel domains of mouse ornithine decarboxylase (ODC; PDB file 7ODC), *E. coli* alanine racemase (Ala Racem, PDB 2RJG) and *Vibrio cholerae* broad specificity amino acid racemase (AA Racem; PDB 4BEU). Residues conserved or conservatively replaced in at least 5 of the 7 aligned sequences are highlighted by black shadowing over white lettering. When the conservation is restricted to the 5 COG0325 products or to the 3 enzymes, the residues are highlighted yellow and green, respectively. Horizontal arrows (in light blue) and rectangles (pinkish) under each sequence mark β strands and α helices, respectively, with labelling of these secondary structure elements at the top of the alignment (blue and pink labels). The triangles at the top and bottom signal residues that make contacts with PLP in the COG0325 products and the enzymes, respectively, being blue for the Schiff base-making lysine, arrow-shaped and green-lined for phosphate-interacting residues, black if the interaction is conserved in all the aligned sequences, or yellow or red if conserved only in COG0325 products or the enzymes, respectively. In addition, yellow and red lettering is used in some occasions to confirm the interaction of a given residue with PLP when the interaction is not fully conserved. The red asterisk denotes the lysine that is carboxylated in Ala racemase. The curved small arrows and the adjacent mutations (shown in single letter code) above the PipY sequence indicate the corresponding single amino acid changes found in vitamin B6-dependent epilepsy [17, 18]. **(B)** Stereo view of the superposition of the structures (backbone representation) of PipY-PLP (grey) and the barrel domain of broad spectrum amino acid racemase (green; PDB entry 4BEU), alanine racemase (blue, 2RJG) and ornithine decarboxylase (pink, 7ODC). The bound PLP molecules are shown in sticks representation and are coloured as the corresponding protein chain. Beta strands are labeled with numbers while helices are labelled with an α followed by a number. **(C)** Detailed stereo view (sticks representation) of the superimposed PLP binding sites of PipY-PLP (black) and of the barrel domains of broad spectrum amino acid racemase (green; PDB 4BEU), alanine racemase (orange, 2RJG) and ornithine decarboxylase (pink, 7ODC). O, N, S and P atoms are coloured red, blue, yellow and orange, respectively. Residue labels are in single letter amino acid code and are in the same color as the corresponding structures. The pink asterisk denotes that lysine 122 is carboxylated in alanine racemase [41]. The green sphere is the Cl^- anion in broad spectrum amino acid racemase [39]. Tyr389 of ODC belongs to the other domain of the same subunit of this enzyme. Not shown in the figure because they obliterate the view of other residues are Tyr299 and Tyr255 from the second domain of the other subunit of amino acid racemase and alanine racemase, respectively. These residues extend from the front of the figure towards the O atom bound to the C3 atom of the pyridoxal ring, forming a hydrogen bond with it via their phenolic O atoms.

Capítulo 2: Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations.

Autores: Lorena Tremiño[†], Alicia Forcada-Nadal[†] y Vicente Rubio^{*}

Referencia: *Hum. Mutat.* (2018) 39, 1002-1013

Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations

Lorena Tremiño^{1†}, Alicia Forcada-Nadal^{1†} and Vicente Rubio^{1,2*}

¹Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain; ²Group 739 of the Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER-ISCI), Spain.

† These authors contributed equally to this work.

*Corresponding author:

Vicente Rubio

Instituto de Biomedicina de Valencia

C/Jaime Roig 11

Valencia-46010, Spain.

Email: rubio@ibv.csic.es

Received 19 January 2018;

revised: 3 April 2018;

accepted 20 April 2018;

first published: 24 April 2018

Abstract. Vitamin B₆-dependent genetic epilepsy was recently associated to mutations in *PLPBP* (previously *PROSC*), the human version of the widespread *COG0325* gene that encodes TIM-barrel-like pyridoxal phosphate (PLP)-containing proteins of unclear function. We produced recombinantly, purified and characterized human PROSC (called now PLPHP) and its six missense mutants reported in epileptic patients. Normal PLPHP is largely a monomer with PLP bound through a Schiff-base linkage. The PLP-targeting antibiotic D-cycloserine decreased the PLP-bound peak as expected for pseudo-first-order reaction. The p.Leu175Pro mutation grossly misfolded PLPHP. Mutations p.Arg241Gln and p.Pro87Leu decreased protein solubility and yield of pure PLPHP, but their pure forms were well folded, similarly to pure p.Pro40Leu, p.Tyr69Cys and p.Arg205Gln mutants (judged from CD spectra). PLPHP stability was decreased in p.Arg241Gln, p.Pro40Leu and p.Arg205Gln mutants (thermofluor assays). The p.Arg241Gln and p.Tyr69Cys mutants respectively lacked PLP or had a decreased amount of this cofactor. With p.Tyr69Cys there was extensive protein dimerization due to disulfide bridge formation, and PLP accessibility was decreased (judged from D-cycloserine reaction). A 3-D model of human PLPHP allowed rationalizing the effects of most mutations. Overall, the six missense mutations caused ill effects and five of them impaired folding or decreased stability, suggesting the potential of pharmacochaperone-based therapeutic approaches.

Keywords: COG0325 family proteins, PLPBP gene, inborn errors, PROSC, PLPHP recombinant production, PLPHP characterization, pyridoxal phosphate, PLP, site-directed mutagenesis, structure-function correlations.

INTRODUCTION

Very recently a novel form of vitamin B₆-dependent genetic epilepsy of recessive inheritance was reported (Darin et al., 2016; Plecko et al., 2017), associated to mutations in the *PROSC* gene (MIM# 617290; this gene is now renamed *PLPBP*, from pyridoxal phosphate binding protein; www.genenames.org). *PLPBP* is the human version of *COG0325*, a gene that is highly conserved and that is very widely distributed among all life forms, from bacteria to humans (Ito et al., 2013; Prunetti et al., 2016; Labella et al., 2017). As indicated by its name, *PLPBP* encodes a pyridoxal phosphate (PLP)-containing protein of uncertain function that, having been proposed to be involved in pyridoxal phosphate (PLP) homeostasis (Darin et al., 2016; Ito et al., 2013; Plecko et al., 2017; Prunetti et al., 2016), is now called PLPHP (<http://www.uniprot.org/uniprot/O94903>; old name, *PROSC*). Although this protein has been found not to be essential for life in several organism (Ito et al., 2013; Labella et al., 2017; Prunetti et al., 2016), null mutations of the *COG0325* gene associate in bacteria with pleiotropic phenotypes that include increased sensitivity to pyridoxine toxicity and to PLP-targeting antibiotics, as well as various amino-acid metabolism-related aberrations (Ito et al., 2013; Ito et al., 2016; Kolodkin-Gal et al., 2010; Labella et al., 2017; Prunetti et al., 2016). In humans, of the eleven patients that have been reported thus far with vitamin B₆-dependent epilepsy associated to *PLPBP* mutations (Darin et al., 2016; Plecko et al., 2017), four patients carried null mutations in homozygosis due to protein truncation at residues 71 or 78 (Darin et al., 2016), quite upstream in the 275-residue PLPHP polypeptide chain. Another patient was a compound heterozygote for two different splice-site changes that should have resulted in decreased levels of normal PLPHP (Darin et al., 2016). From these truncating or splicing mutations it is clear that the loss of PLPHP function triggers early in life

nervous system pathology manifested as epilepsy with developmental delay if not effectively treated with vitamin B₆-supplementation.

In addition to these patients carrying truncating or splicing mutations, six patients carried single amino acid substitutions of uncertain effects on PLPHP. Table 1 lists these mutations and includes the second mutant allele and a summary of the reported clinical data for each patient carrying one of these mutations (Darin et al., 2016, Plecko et al., 2017). Of these mutations, p.Leu175Pro was previously shown (Darin et al., 2016) to decrease *PLPBP* mRNA production by ~75% in the fibroblasts from a patient carrying this mutation in homozygosity, allegedly because of the creation of a novel siRNA site. However, the effect of this mutation at the protein level was even more drastic, since no PLPHP could be detected in western blots of these fibroblasts (Darin et al., 2016).

The effects of another two missense mutations (p.Pro87Leu and p.Arg241Gln) were studied in complementation assays in *yggS Escherichia coli* (Darin et al., 2016). This mutant bacterial strain lacks the corresponding COG0325 protein, YggS, and exhibits a phenotype of decreased resistance to pyridoxine toxicity that can be restored to normal by complementation with wild-type *PLPBP* (Darin et al., 2016; Ito et al., 2013; Prunetti et al., 2016). However, the results of these complementation assays did not parallel the clinical severity inferred for these two mutations (Darin et al., 2016), what prompted us to introduce these mutations into the recombinantly expressed COG0325 gene of the cyanobacterium *Synechococcus elongatus*, using the produced protein, PipY, as a model of human PLPHP (Tremiño et al., 2017).

We now produce recombinantly and purify human PLPHP, investigating the effects on protein properties of all the *PLPBP* missense mutations reported thus far in patients with vitamin B₆-dependent epilepsy (Darin et al., 2016;

TABLE 1. Summary of PLPBP missense mutations reported by Darin et al. (2016) (report 1) and Plecko et al. (2017) (report 2) in patients with vitamin B₆-dependent epilepsy

Mut. #	Sequence substitution cDNA ^a	Protein ^b	Effects on PLPBP ^c	Report/ Patient ^d	2nd allele in same patient		Seizures: onset age/ type at presentation ^e	B ₆ vitamer for: seizure control		Developmental outcome ^e	
					cDNA	Protein		Form ^f	mg/day		Response
1	c.119C>T	p.Pro40Leu	↓ thermal stability	2/1	c.722G>A	p.Arg241Gln	d7/myoclonia	PN	300-400	Good	Normal at y12.5
2	c.206A>G	p.Tyr69Cys	Dimerization ↓ PLP content ↓ PLP accessibility	2/4	c.206A>G	p.Tyr69Cys	d9/tonic head and eye deviation	PN	300	Good	Learning difficulties, y30
3	c.260C>T	p.Pro87Leu	↓ solubility	2/3	c.260C>T	p.Pro87Leu	d3/GTC	PN	450	Good	Marked delay at m27
4	c.524T>C	p.Leu175Pro	Misfolding	1/4	c.722G>A	p.Arg241Gln	m1/GTC	PN	250	Good	Normal schooling and life, IQ60, y17
5	c.614G>A	p.Arg205Gln	↓ thermal stability	2/2	c.524T>C	p.Leu175Pro	d1/GTC	PLP	45 kg ⁻¹	Partial	Marked delay at y3.5
6	c.722G>A	p.Arg241Gln	↓ solubility ↓ thermal stability Protein lacks PLP	2/1 1/7	c.249_252del	p.Ser84Cysfs*21	d6/clonic arms, legs	PN	100	Good	Normal at y15.5

^a Respective GenBank reference sequences for the human PLPBP gene and for its mRNA, NC_000008.11 and NM_007198.3. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the mRNA reference sequence, with the initiation codon as codon 1.

^b Protein reference sequence NP_009129.1; Uniprot KB O94903.

^c Results from this work. For each mutation, the effect that might have the most important negative impact is highlighted in bold-type. For mutation 3 the importance of the reported effect on disease causation is uncertain.

^d Patients are identified by their numbers in the original reports of Darin et al. (2016) (report 1) and Plecko et al. (2017) (report 2).

^e The letters preceding the figures denote days (d), months (m) or years (y). GTC, generalized tonic-clonic seizures. For further clinical details, see the original report in which each patient was described.

^f PN, pyridoxine. PLP, pyridoxal phosphate.

Plecko et al., 2017), introduced by site-directed mutagenesis. Actually, we characterize here for the first time the properties of normal human PLPHP, determining its oligomeric state in solution, its spectroscopic properties reflecting PLP binding, its thermal stability using thermofluor assays, and its sensitivity to the PLP-targeting antibiotic D-cycloserine (DCS) (Neuhaus, 1967). This characterization has been essential for assessing experimentally the effects of these missense mutations introduced into the recombinant protein. A structural model of PLPHP based on the experimental structures of other COG0325 proteins (including PipY) has helped rationalize these effects. In this way, we have found for all the mutations differential traits relative to the wild-type protein that can account or contribute to disease causation.

MATERIALS AND METHODS

Recombinant production of wild-type and mutant forms of human PLPHP.

The human *PLPBP* coding sequence (HGNC ID: 9457; *GenBank* gene reference sequence NC_000008.11; mRNA reference sequence NM_007198.3, numbering +1 the A of the ATG translation initiation codon, which is codon 1; protein reference sequence NP_009129.1; Uniprot KB entry O94903) was provided by GenScript (Piscataway, USA) as a codon-optimized (for *E.coli* expression) synthetic gene inserted into the KpnI-BamHI sites of the pGS-21a plasmid. After biological amplification in DH5 α cells, this plasmid was used for introduction of the desired *PLPBP* missense mutations by the Quickchange approach (Stratagene, La Jolla, CA, USA) using Deep Vent DNA polymerase (New England Biolabs, Ipswich, MA, USA) and the indicated (Table 2) forward and reverse mutagenic primers. After digestion of the parental plasmid by DpnI digestion, the mutant plasmids were used to transform *E. coli* DH5 α cells, from

which they were isolated and corroborated to host the correct mutant sequences (Sanger sequencing). For protein expression, the *PLPBP* coding sequences from the wild type and the mutant pGS-21a plasmids were PCR-amplified (Deep-Vent DNA polymerase and primers PLPHP-1-F and PLPHP-275-R, Table 2) and cloned into NcoI/PmeI-digested pOPIN-E, using the In-Fusion technology (In-Fusion HD Enzyme Premix, Clontech, Madrid, Spain), followed by transformation into DH5 α cells, plasmid isolation and verification (Sanger sequencing with primers PLPHP-1-F and PLPHP-275-R, Table 2). *E. coli* BL21 (DE3) cells (Invitrogen) transformed with the appropriate pOPIN-derived plasmid were grown at 37 °C in aerated 0.5 L of LB-ampicillin (0.1 mg/ml) to 0.6-0.8 OD⁶⁰⁰, then 1 mM isopropyl- β -D-1-thiogalactopyranoside was added and the culture was continued for 3 hours. Wild-type and mutant PLPHP forms were purified in the same way at 4°C. After the induction period, the cells were

TABLE 2. Oligonucleotides used in the present study. The bases that are mutated are highlighted in bold type, and the mutated codon is underlined. Sequences used for cloning on pOPINE are shown in low case.

Name	Sequence (5'- 3')
PLPHP-P40L-F	CGATCCAGCTGCGTCTG
PLPHP -P40L-R	AGACGCAGCTGGATCGCC
PLPHP -Y69C-F	GTGAAA <u>ACTG</u> IGTTCAAGAG
PLPHP -Y69C-R	TCTTGAACACAGTTTTCCACC
PLPHP -P87L-F	AGCCTGTGCCTGGAAATCAAG
PLPHP -P87L-R	GATTTCCAGGCACAGGCTCAG
PLPHP -L175-F	CCGAACCCGGAGTTC
PLPHP -L175-R	CTCCGGGTTCTGGGC
PLPHP -R205Q-F	AGCCTGCAGGAGGAACTG
PLPHP -R205Q-R	AGTTCCTCCTGCAGGCTC
PLPHP -R241Q-F	AACGTT <u>CAG</u> ATCGGTAGC
PLPHP -R241Q-R	ACCGATCTGAACGTTGGT
PLPHP -1-F	aggagatataccatgTGGCGTGCGGGTAG
PLPHP -275-R	gtgatggtgatgtttGTGTTCTGCGCAAC

harvested by centrifugation, disrupted by sonication in 15 ml of lysis buffer [25 mM HEPES pH 7.5, 0.5 M NaCl, 1 mM dithiothreitol (DTT) and 1 mM phenyl methyl sulphonyl fluoride (PMSF)], and the suspension was clarified by 30-min centrifugation at $11,000 \times g$. The supernatant was applied to a 1-ml Ni-chelate HisTrap-HP column fitted in an ÄKTA-FPLC system (both from GE Healthcare). After a 30 ml-wash with column buffer (25 mM HEPES pH 7.5, 0.5 M NaCl and 20 mM imidazole), a 80-ml 20 mM-to-500 mM linear imidazole gradient was applied in the same buffer, monitoring the optical absorption of the effluent at 280 nm, collecting fractions. PLPHP, eluted at about 120 mM imidazole, was pure (monitored by SDS-PAGE). It was concentrated to >2 mg protein/ml, placing it simultaneously in 20 mM HEPES pH 7.5, 16 mM NaCl, using centrifugal ultrafiltration (10 kDa cutoff membrane, Amicon Ultra, from Merck-Millipore, Merck Chemicals and Life Sciences, Madrid), and storing it at -20 °C until use.

Electrophoretic techniques

SDS-PAGE (Laemmli, 1970) was performed in 12% polyacrylamide gels, using Coomassie Blue staining to visualize protein bands. When indicated, the 3% (v/v) β -mercaptoethanol (MSH) normally present in the sample buffer was omitted.

Western blotting was used to detect PLPHP in the supernatant and precipitate obtained after centrifugation of the initial crude cell extracts. Equal volumes of the supernatant and of the precipitate reconstituted in the original volume of lysis buffer were subjected to SDS-PAGE followed by western blotting to nitrocellulose membranes and immunodetection as previously reported for PipY (Labella et al., 2017). The proper electrophoretic transfer to the membranes was verified by observing the similar protein banding pattern of wild type and mutant proteins in the membranes after staining them with

Poinceau red. The primary antibody (anti-PipY rabbit polyclonal antibody) was used at higher concentration (1:750 dilution) than for PipY (Labella et al., 2017), since although this antibody cross-reacts with human PLPHP, its potency for this protein in western blots was ~2-fold lower than for PipY (data not shown).

Optical methods

Circular dichroism spectra were recorded at 21 °C with a Jasco J-810 spectropolarimeter (Jasco, Easton, MD, USA) on 0.1 cm-path cuvettes containing wild type or mutant forms of PLPHP in 50 mM Na phosphate pH 7.5. Each spectrum was the average of five scans. All samples that were purified were analyzed at 5 and 10 µM concentration of PLPHP polypeptide chains. Closely similar results were obtained for these two concentrations for each protein form. Data were processed with the K2d server (available in DichroWeb, <http://dichroweb.cryst.bbk.ac.uk>).

Optical absorption spectra over the range 220-748 nm were determined on 2 µl of protein solution using a NanoDrop™ 1000 V3.8.1 microspectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). The protein was dissolved in 20 mM HEPES pH 7.5 and 16 mM NaCl. When indicated, the solutions containing 5-7 mg/ml PLPHP (wild type or mutant forms) as well as negative control solutions lacking the protein were incubated 20 min at 25 °C with variable concentrations of DCS and then the absorption spectra in the UV-visible range were determined. The spectra due to the protein was determined by difference of the spectra of the parallel protein-containing and protein-free incubations

Size-exclusion chromatography

Size-exclusion chromatography of purified PLPHP variants (0.1-0.3 mg; ≤50 µl injections) or of protein standards of known mass was carried out on a

Superdex™ 75 5/150 GL column (GE Healthcare) mounted on an ÄKTA FPLC system, monitoring the OD²⁸⁰ of the effluent. The solution used to equilibrate and to run the column at 25 °C and at a flow rate of 0.4 ml/min contained 10 mM Na phosphate pH 7, 0.4 M NaCl, 5 mM MgCl₂, 5 mM EDTA and 0.2 mM DTT. The standards used (masses in kDa between parentheses) were rabbit phosphorylase b (97.4), bovine serum albumin (66.4), chicken ovalbumin (45), maltose binding protein (42), bovine erythrocyte carbonic anhydrase (29), soybean trypsin inhibitor (21.5) and horse heart cytochrome C (12.3).

Thermal stability assays

Thermofluor assays (Vedadi et al., 2006) were performed in triplicate on a real-time PCR instrument (7500 model from Applied Biosystems, ThermoFisher Scientific, Alcobendas, Madrid, Spain), monitoring the increase in SYPRO Orange (from Invitrogen, Carlsbad, CA) fluorescence upon gradual (1°C/min) temperature increase of solutions of 15-20 μM PLPHP (wild type or mutant forms) in 20 mM HEPES pH 7.5/16 mM NaCl and SYPRO Orange (a 1:1000 dilution of the commercial preparation).

Structural model generation

An initial model for the PLPHP 3-D structure was generated with I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Yang et al., 2015). Automated and manual refinement cycles with ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) (Xu & Zhang, 2011) and COOT (Emsley, Lohkamp, Scott & Cowtan, 2010) were used in succession to eliminate structural clashes and to improve the model. When the model appeared satisfactory and presented good stereochemistry (tested with RAMPAGE, Lovell et al., 2003), a PLP molecule was incorporated by

superimposing this structure on the experimental structure of PipY [Protein Databank (PDB) entry 5NM8; www.rcsb.org] (Tremiño et al., 2017). Then refinement cycles were repeated again as above. The final model with PLP bound presented no clashes or bad contacts (tested with MolProbity, Williams et al., 2017) and exhibited excellent stereochemistry, with no outliers in the Ramachandran plot.

Other techniques

Pure PipY was prepared as previously reported (Tremiño et al., 2017). The amount of pure PLPHP protein was determined from its optical absorption at 280 nm for sequence-deduced (EXPASY ProtParam tool, <https://www.expasy.org>) absorbance values for 0.1% solutions (1-cm optical path) of 0.636 (p.Tyr69Cys mutant) or 0.678 (all other PLPHP forms).

Plots, numerical calculations and curve fittings were performed with program GraphPad Prism 5 (GraphPad Software, San Diego, CA, USA). Figures of protein structure were prepared with PyMOL (Pymol Molecular Graphics System, Version 1.6, Schrödinger, LLC, www.pymol.org).

Preparations of wild-type and mutant proteins and experiments carried out with them were repeated at least twice. The electrophoretic and size-exclusion chromatography profiles shown in Figs. 1 and 2 illustrate individual experiments but are representative of all the experimental repeats.

RESULTS AND DISCUSSION

Production and characteristics of pure wild-type human PLPHP

We produced recombinantly in *E. coli* human PLPHP with a C-terminal 6-His tag. Although many human proteins are not expressed well in *E. coli*, this was not the case for PLPHP, which was highly soluble, as shown by its recovery largely in

the supernatant of the disrupted bacterial cells expressing it (Fig. 1A, tracks labeled WT), and which could be purified to homogeneity (purity monitored by SDS-PAGE, Fig. 1B, top panel, WT track) in quite a good yield (close to 10 mg/L of culture, Fig. 1C, bar labeled WT). Its CD spectrum (Fig. 1D, left panel, continuous line) closely resembled that of its cyanobacterial homologue PipY (Tremiño et al., 2017) indicating that PLPHP is well folded and agreeing with a similar composition of α helices and β strands of PLPHP and PipY, thus supporting the folding of PLPHP essentially as PipY. This last protein was shown crystallographically to present the modified TIM barrel fold of type III PLP proteins (Tremiño et al., 2017), a fold also exhibited by the structural model of PLPHP built here (see below Fig. 3A).

As expected for a PLP-containing protein, PLP solutions had yellow color (picture inset in Fig. 1C) and presented in the optical absorption spectrum the characteristic peak at about 425 nm (Fig. 1E) that reflects the protonated Schiff base of protein-bound PLP (Fasella, 1967), thus justifying the inclusion of bound PLP in the structural model of PLPHP (see below Fig. 3A).

Previously, PipY was found to be monomeric in solution as revealed by size exclusion chromatography (Tremiño et al., 2017). The same technique also supported that PLPHP is a monomer (Fig. 2A, left). However, while with PipY only the monomer peak was observed (Fig. 2A, left) (Tremiño et al., 2017), a minor peak was seen with PLPHP in the UV absorption profile of the column effluent, with the elution position of its maximum fitting the position expected for a dimer (Fig. 2A, left). The monomer and the dimer peaks of PLPHP exhibited the same optical absorption spectrum (not shown) including the 425 nm peak, indicating that both of them had PLP bound. Rechromatography of fractions collected from each of the peaks gave (Fig. 2A, right, top and middle panels) only the original peak in each case (except for minor amounts of the

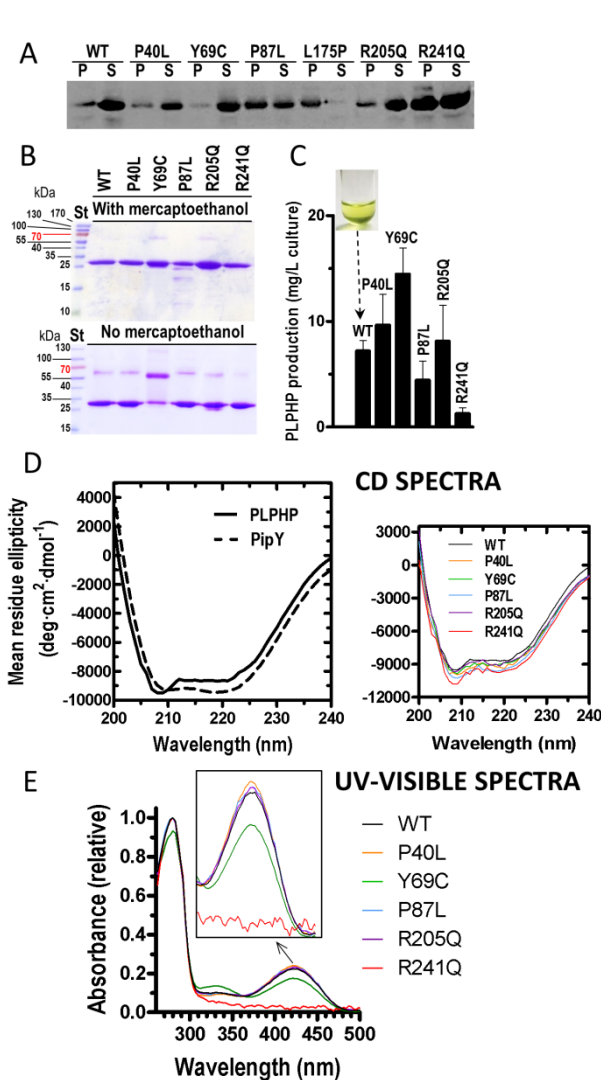


Figure 1. Production and characterization of wild-type and mutant forms of PLPHP. For shortness, mutations are abbreviated in single-letter amino acid code. WT, wild type. **(A)** Western blot of PLPHP in the precipitate (P) and the supernatant (S) following centrifugation of bacterial cell extracts (see Materials and Methods). The amounts of precipitate and supernatant applied for a given mutant represent the same volume of original extract. **(B)** SDS-PAGE (Coomassie staining) of the indicated purified PLPHP variants. Top gel, standard sample treatment for SDS-PAGE (which includes 3% MSH). Bottom gel, MSH omitted. **St**, Page Ruler Prestained Protein Ladder (Thermo Fisher Scientific) with masses (in kDa), given at the side. **(C)** Yield of pure wild type or mutant PLPHP (mean \pm SEM, $n=4$ for WT and $n=2$ for the

mutants). Only the yield of the p.Arg241Gln mutant differs significantly ($p<0.01$, Student's *t*-test) from the yield obtained for WT PLPHP. The test tube picture demonstrates the yellow color of pure WT PLPHP. **(D)** Representative CD spectra for (*left panel*) PLPHP and *S. elongatus* PipY (data for PipY are taken from Tremiño et al., 2017) or for (*right panel*) the indicated pure human PLPHP variants. The traces shown were obtained with 10 μ M (0.31 mg/ml) PLPHP polypeptide chains. **(E)** Optical absorption spectra of wild type and mutant forms of PLPHP, normalized to identical protein concentration (value of 1 given to the optical absorption at 280 nm, excepting the p.Tyr69Cys mutant, for which the relative OD²⁸⁰ maximum was given a value of 0.94 (see Materials and Methods). Data are the averages for 4 samples (WT) or 2 samples (all other forms). *Inset*: zoom on the 425 nm PLP peak of the various PLPHP protein variants.

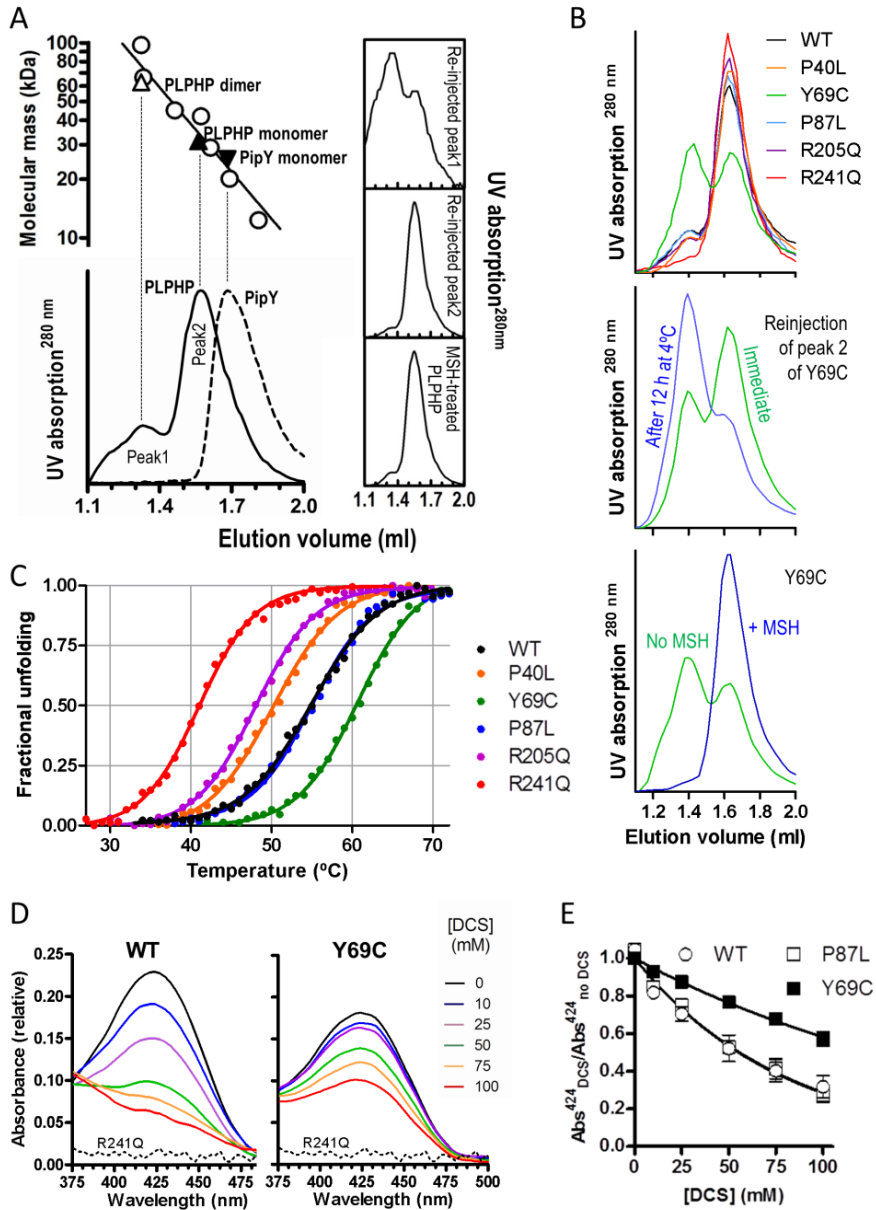
other peak possibly derived from contamination in the fractions used for rechromatography), excluding that the monomer and dimer were in rapid equilibrium. This suggested that the dimer was formed by covalent binding of two monomers, possibly by a disulfide bridge. This possibility appeared likely, since, in contrast with PipY, which has a single cysteine, PLPHP has 5 cysteine residues, three of them exposed in its 3D structural model (cysteines 15, 209 and 261, see below Fig. 3A). Since no dimer was observed on standard SDS-PAGE of PLPHP (Fig. 1B, top), which includes mercaptoethanol (MSH) in the sample solution, we repeated SDS-PAGE of this protein omitting the addition of this disulfide reducing agent, revealing a small amount of dimer (Fig. 1B, bottom), agreeing with the size exclusion chromatography results and with the hypothesis of a disulfide bond being responsible for dimer formation. Similarly, brief preincubation of native PLPHP with 10 mM mercaptoethanol prior to size exclusion chromatography greatly reduced the amount of the early-eluting peak (Fig. 2A, right, bottom panel). Therefore, although the majority of PLPHP is monomeric, it appears that this protein can dimerize stably by forming a disulfide bond between its exposed cysteines. As will be seen, dimerization is much more prominent for the p.Tyr69Cys mutant.

Effects of the disease-associated PLPHP amino acid substitutions on folding, absorption spectra and oligomeric state of the protein

Site-directed mutagenesis was used to introduce each one of the 6 missense PLPHP mutations reported in patients with vitamin B₆-induced epilepsy (Table 1). Only one of these mutations, p.Leu175Pro, caused gross misfolding, judged from the complete insolubility of the mutant protein when it was expressed in *E. coli* (Fig. 1A). This was revealed by specific detection of the protein in western blots using a polyclonal rabbit antibody against PipY that cross-reacts

specifically with PLPHP. Clearly, the p.Leu175Pro mutant was in the precipitate of the initial cell extract, with no detection of any PLPHP in the corresponding supernatant. With the other five mutants, soluble protein was always found, although the data (Fig. 1A) revealed lower solubility of the p.Pro87Leu and p.Arg241Gln mutants compared with wild-type PLPHP. Nevertheless, the decreased solubility of these last two mutants did not prevent their purification, although in lower yield than the wild-type protein or the other mutants (Fig. 1C). Circular dichroism spectra of the purified mutant proteins were very similar to that for the wild-type protein, showing that these purified mutant forms were well folded (Fig. 1D, right panel).

Naked-eye observation of the preparations of the purified mutants showed that all of them, similarly to the wild-type enzyme, were colored yellow, with the exception of the p.Arg241Gln mutant, which was colorless. In addition, the p.Tyr69Cys mutant appeared to have a fainter tone than preparations of wild-type PLPHP having similar protein concentration. These observations were confirmed by examining the absorption spectra of these mutant proteins (Fig. 1E). While the spectra were essentially identical for wild type PLPHP and for the p.Pro40Leu, p.Pro87Leu and p.Arg205Gln mutants, there was no peak at 425 nm in the spectrum of the p.Arg241Gln mutant, indicating lack of bound PLP, strongly suggesting an important decrease in the affinity of this mutant apoprotein for PLP. In turn, the p.Tyr69Cys mutant showed a decrease of ~30% in the height of the 425 nm peak (Fig. 1E). This indicates a decrease in the amount of the protonated Schiff base of protein-bound PLP formed by the reaction of PLP with Lys46 (residue identified from sequence alignment with COG0325 proteins of known structure, see below Fig. 3B). Thus, this mutant protein has a reduction in the amount of normally bound PLP, what could have a negative impact on PLPHP function.



We also assessed the oligomeric nature of the different mutants by size exclusion chromatography (Fig. 2B, top). Excepting the p.Tyr69Cys mutant, all the mutants were eluted as wild-type PLPHP, as monomers with a minority dimer content. The dimer peak appeared somewhat smaller in the case of the

Figure 2 (previous page). Size exclusion chromatography, thermal stability and reactivity with DCS of wild-type and mutant forms of PLPHP. Abbreviations are as in Fig. 1. (A) Size exclusion chromatography of pure PipY and PLPHP proteins. The UV absorption of the effluent was monitored (*bottom*). A semilogarithmic plot (*top*) relating the mass of protein standards (*open circles*; listed in Materials and Methods) with their elution volumes is shown on the top. *Open* and *closed upright triangles* mark the respective elution positions of the two PLPHP peaks, attributed the sequence-deduced masses for dimers and monomers, respectively. *Inverted closed triangle*, PipY monomer. *Insets* to the right depict the peaks observed after rechromatography of samples taken from PLPHP peaks 1 (*top*) or 2 (*middle*), or the elution pattern of PLPHP following its incubation for 20 min at 25 °C with 10 mM MSH (*bottom*). (B) *Top*: Elution of wild-type and mutant forms of PLPHP from the gel filtration column. *Middle*: Rechromatography of a sample from peak 2 of the first chromatography of the p.Y69C mutant. A part of the sample was analyzed immediately after the first chromatography (*green*) whereas another part was left 12 hours at 4 °C prior to application to the column (*blue*). *Bottom*: Observed elution pattern of the p.Y69C mutant after its preincubation (20 min, 25 °C) with 10 mM MSH (*blue*). For comparison, the elution profile for the same mutant when MSH was omitted is shown in *green*. (C) Thermofluor assays (means of 3 replicas) of the indicated PLPHP forms. (D) Effects of 20-min incubations with variable DCS concentrations on the 424-nm optical absorption peak of the WT and p.Y69C forms of PLPHP. To illustrate the expected background absorbance for complete PLP removal, the spectrum for the p.R241Q mutant, which has no PLP bound, is also shown. The traces for a representative experiment are shown. (E) Plot of the decrease in OD^{424} for the indicated PLPHP forms as a function of DCS concentration in the 20-min incubation with the protein. Data are the means \pm SEM for three (WT) or two (mutants) separate experiments. The lines fitted are exponential decays for concentrations of DCS giving 50% decay (C_{50}) after 20-min incubation, of 55 mM (WT and p.Pro87Leu) and 128 mM (p.Tyr69Cys). The differences between these values of C_{50} is significant ($p < 0.001$, Student's t-test).

p.Arg241Gln mutant, possibly reflecting conformational changes between the PLP-containing and PLP-free forms (as reported with PipY, Tremiño et al., 2017) that could influence the accessibility in the PLP-free form of the exposed cysteines involved in disulfide bridge formation. However, the largest difference between mutant and wild-type was observed with the p.Tyr69Cys mutant. The fresh protein containing this mutation was eluted from the gel filtration column as two large peaks, one corresponding to the dimer and the other to the monomer (Fig. 2B, top). When a part of the monomer peak was reinjected, it was predominantly monomeric (Fig. 2B, middle), although there

was an important amount of dimer that could have arisen from contamination of the original sample with dimer and/or from spontaneous dimer formation from the monomer. To evaluate how fast dimers were generated from monomers, we left a fraction of the monomer peak at 4 °C for 12 hours. After this period, the majority of the protein was dimeric (Fig. 2B, middle). Thus, it appears that the dimer and the monomer are not in rapid equilibrium, and that the dimer increases with time, as expected from dimerization due to disulfide bridge formation by the novel cysteine introduced by the mutations. This was confirmed by the elution of the protein as a monomer from the gel filtration column when the protein had been treated with 10 mM MSH (Fig. 2B, bottom). Further confirmation was provided by SDS-PAGE of the p.Tyr69Cys mutant without or with treatment with MSH (Fig. 1B). When subjected to this reducing treatment (Fig. 1B, top) all the protein moved as a monomer, whereas when not treated with MSH (Fig. 1B, bottom) much of the p.Tyr69Cys mutant migrated as a dimer. This contrasts with the behavior of all other mutants and of wild type PLPHP, which, when not treated with MSH, migrated largely as monomers, with only a minority as dimers (Fig. 1B, bottom). Also in agreement with the gel filtration results, SDS-PAGE without MSH confirmed that the dimer was less abundant in the p.Arg241Gln mutant than with the other mutants or with wild-type PLPHP (Fig. 1B, bottom). It is to be noted that the dimer of the p.Tyr69Cys mutant moved in SDS-PAGE slightly faster than the dimers for the other forms (Fig. 1B, bottom), whereas the movement of the monomer was identical for all forms. Thus, even in the presence of SDS there appear to be structural differences between the p.Tyr69Cys dimer and the dimers observed with the other forms of PLPHP. These differences may reflect the formation with the p.Tyr69Cys mutant of a novel type of dimer involving a disulfide bridge

between two monomers mediated by the newly created cysteine found in the mutant.

Influence of the mutations on PLPHP thermal stability

We compared in thermofluor assays the thermostability of the different mutants with that of wild-type PLPHP (Fig. 2C). For wild-type PLPHP the change in fluorescence with heating attained 50% of its maximum at about 56 °C (T_m value) (Fig. 2C). The same T_m temperature was observed with the p.Pro87Leu mutation, while the P40L, R201Q and particularly the R241Q mutants exhibited decreased thermal stabilities, corresponding to T_m values of approximately 50, 49 and 42 °C, respectively (Fig. 2C). These reductions in T_m values are quite large, making it likely that the decrease in thermal stability could be an important disease-causing element for these mutants. In contrast with the observations with the p.Pro40Leu, p.Arg205Gln and p.Arg241Gln mutations, the p.Tyr69Cys mutant was substantially more stable in the thermofluor assay than the wild-type protein (T_m , 60.7 °C, nearly 5 °C higher than the T_m of wild-type PLPHP) (Fig. 2C). Apparently, the dimerization stabilizes the protein, protecting it from thermal unfolding.

Decreased PLP accessibility in the p.Tyr69Cys mutant, assessed using D-cycloserine (DCS).

The existing evidence supports a role of PLPHP in PLP homeostasis (Darin et al. 2016; Ito et al., 2013; Plecko et al., 2017; Prunetti et al., 2016; Tremiño et al., 2017), whereby PLPHP could deliver PLP to target PLP-binding proteins. Dimerization of PLPHP could hamper such delivery and thus could be an important determinant of disease causation. Since the antibiotic DCS targets protein-bound PLP, we used DCS to monitor the degree of accessibility of PLP, since such accessibility could be an important factor for PLP delivery. Thus, we

compared the dependency on the concentration of DCS of the reaction of this antibiotic with the PLP of wild-type PLPHP or of the p.Tyr69Cys mutant (Fig. 2D). We also tested in the same assays the p.Pro87Leu mutant (not shown), a mutant for which no ill-effect of the mutation had been found other than a decrease in protein solubility. In these assays we monitored the decrease in the PLP absorption peak after 20-min reaction with variable concentrations of DCS. While in these assays the wild-type and the p.Pro87Leu mutant reacted with DCS to similar degrees, the reaction of the PLP in the p.Tyr69Cys mutant was approximately 2.5-fold slower (Fig. 2E), indicating that the p.Tyr69Cys substitution decreases the availability of PLP for reaction with DCS. Therefore, it appears that PLP donation by PLPHP could be importantly hampered by the p.Tyr69Cys mutation.

Overall considerations

All the missense mutations studied here caused ill effects on PLPHP (Table 1, 4th column), strongly suggesting that the epilepsy resulted from decreased PLPHP function. This agrees with the results of *in silico* predictions that predominantly suggested probably damaging effects for the mutations p.Pro40Leu, p.Tyr69Cys, p.Arg205Gln and p.Arg241Gln (Plecko et al., 2017) as well as for the p.Leu175Pro mutation (our own analysis with PolyPhen-2 and MutPred servers, <http://genetics.bwh.harvard.edu/pph2/>, <http://mutpred.mutdb.org/>; Adzhubei et al., 2010; Li et al., 2009). Only for the p.Pro87Leu mutation the *in silico* predictions were ambiguous (Plecko et al., 2017).

Misfolding/decreased stability appear to be the major pathogenic determinants (Table 1, highlighted in bold-type) for the mutations p.Pro40Leu, p.Leu175Pro, p.Arg205Gln and p.Arg241Gln. Gross misfolding is reflected in the

complete lack of solubility of the p.Leu175Pro mutant. The p.Arg241Gln mutations also appeared to cause substantial misfolding, reflected in an important increase in the fraction of insoluble PLPHP protein (Fig. 1A), which was translated in our *in vitro* expression assays into low recovery of pure soluble PLPHP protein (Fig. 1C). In addition, the thermal stability of the p.Arg241Gln mutant was found to be drastically decreased, as it also was the case (although to a lesser extent) for the p.Pro40Leu and p.Arg205Gln mutants (Fig. 2C). The p.Pro87Leu mutation also increased the fraction of the protein that was insoluble, decreasing the yield of pure PLPHP. Therefore, for five of the six reported PLPHP missense mutations the structural stability of the protein fold was decreased, indicating that PLPHP deficiency due to these mutations could be a disorder of folding/stability. Thus, a search for pharmacological chaperones (Gómez et al., 2017) should be considered as a potential alley towards therapy of destabilizing PLPHP missense mutations. PLP could be such a chaperone at least in the case of the p.Arg241Gln mutation, since this mutation primarily hampered PLP loading of PLPHP (Fig. 1E and Table 1) and the apo form of PLP-proteins can be less stable than the holoproteins (Relimpio et al., 1981). Perhaps the increased PLP levels observed in these patients and their fibroblasts upon administration of pharmacological doses of pyridoxine (Darin et al., 2016; Plecko et al., 2017) can overcome the PLP loading deficiency of the p.Arg241Gln mutant, increasing *in vivo* the level of this mutant protein by a chaperoning effect, improving the ability of this mutant to carry out its PLP-homeostatic role. This could contribute to the relatively good outcomes of the two patients found to carry this mutation (patients 2/1 and 1/7, Table 1; both were compound heterozygotes of two missense mutations), particularly in the case of patient 1/7 in whom the second mutation, p.Pro87Leu, is considered to be very severe given the presentation observed in an homozygous patient for this last mutation (Table 1, patient 2/3, Plecko et al.,

2017). This severity of the p.Pro87Leu mutation is hard to explain on the bases of our present experimental results, which only documented a substantial but incomplete decrease in the solubility and the yield of *E. coli*-produced recombinant mutant protein. The fact that the p.Pro87Leu mutant protein was produced in substantial level in *E. coli*, together with its normal loading with PLP, explain the complementation by this mutant of *E.coli yggS*⁻ for its pyridoxine toxicity phenotype (Darin et al., 2016), which might depend on buffering down the intrabacterial PLP level (Prunetti et al., 2016). This buffering effect in *E. coli* would not be possible with the p.Arg241Gln mutant because of the intrinsic and importantly decreased affinity for PLP of this mutant. The severity of the p.Pro87Leu mutation might reflect selective loss in the human of the function of this mutant, perhaps because of accelerated mutant protein destruction by an animal-specific degradative process that would not be active against this mutant in *E. coli*. Alternatively, Pro87 might be important for interaction of PLPHP with its human protein targets, so that the p.Pro87Leu mutation might deprive PLPHP of at least part of its PLP homeostatic function.

As in the case of the p.Arg241Gln mutation, but to a much lesser extent, the p.Tyr69Cys mutation also hampered PLP loading of PLPHP. Perhaps a more significant effect of this mutation was the fact that the novel cysteine also caused covalent dimer formation of the p.Tyr69Cys mutant, resulting in decreased PLP accessibility, judged from the reaction of this mutant with DCS (Fig. 2C, E). This decreased accessibility might be a key factor in disease-causation by this mutant form of PLPHP (Table 1, highlighted in bold-type). In any case, the results with p.Arg241Gln and p.Tyr69C support the importance of PLP in the function of PLPHP, as would be expected if this function were PLP homeostasis (Darin et al., 2016, Ito et al., 2013; Labella et al., 2017, Prunetti et al., 2016, Tremiño et al., 2017). For this function, difficulties in loading PLPHP

with its PLP cargo, most prominent in the case of the p.Arg241Gln mutant, or hampering of the release of this cargo as it appears to occur with the p.Tyr69Cys mutant are likely causes of decreased PLPHP function.

3-D structural modelling yielded for PLPHP (Fig. 3A and insets therein) the expected modified TIM barrel fold that has been observed in other COG0325 members of known structure (Eswaramoorthy et al., 2003; Tremiño et al., 2017; and PDB files 1W8G, 3CPG, & 3R79), and allowed characterization of the mutation's effects for the p.Leu175Pro, p.Arg241Gln and p.Tyr69Cys mutations. The structural model of PLPHP suggests (Fig. 3A, inset A1) that the p.Leu175Pro mutation is structurally very disturbing because it causes a very drastic change (Pro is smaller than Leu and it does not have a ramified hydrophobic side-chain; its imino acid nature imposes very important conformational restrictions to the polypeptide chain; Creighton, 1993) of a virtually constant (only substituted by Ile or Val, Fig. 3B) Leu in a tightly packed and crowded hydrophobic nest that glues the core β sheet of the barrel with the outlying helix layer. The lack of PLPHP in the fibroblasts of a patient that was homozygous for this mutation (Darin et al., 2016) strongly suggests that this misfolding, with removal of the protein by the proteostatic machinery, is the reason for this lack of the mutant protein despite the fact that the fibroblasts still contained substantial amount of the mRNA for *PLPBP* (Darin et al., 2016). The absolute lack of PLPHP in this patient fits the severity of the presentation, which required PLP instead of pyridoxine for seizure control (Table 1) as was found in patients with obligatorily null mutations of both *PLPBP* alleles (Darin et al., 2016).

The lack of PLP in the p.Arg241Gln mutation fits the fact that this mutation affects an invariant arginine (Fig. 3B) shown by the structures of COG0325 proteins to be involved in PLP binding (Fig. 3A, inset A2), since its side chain runs along the buried border of the PLP pyridoxal ring, making with it extensive van der Waals contacts and a N-H-N hydrogen bond (Tremiño et al., 2017).

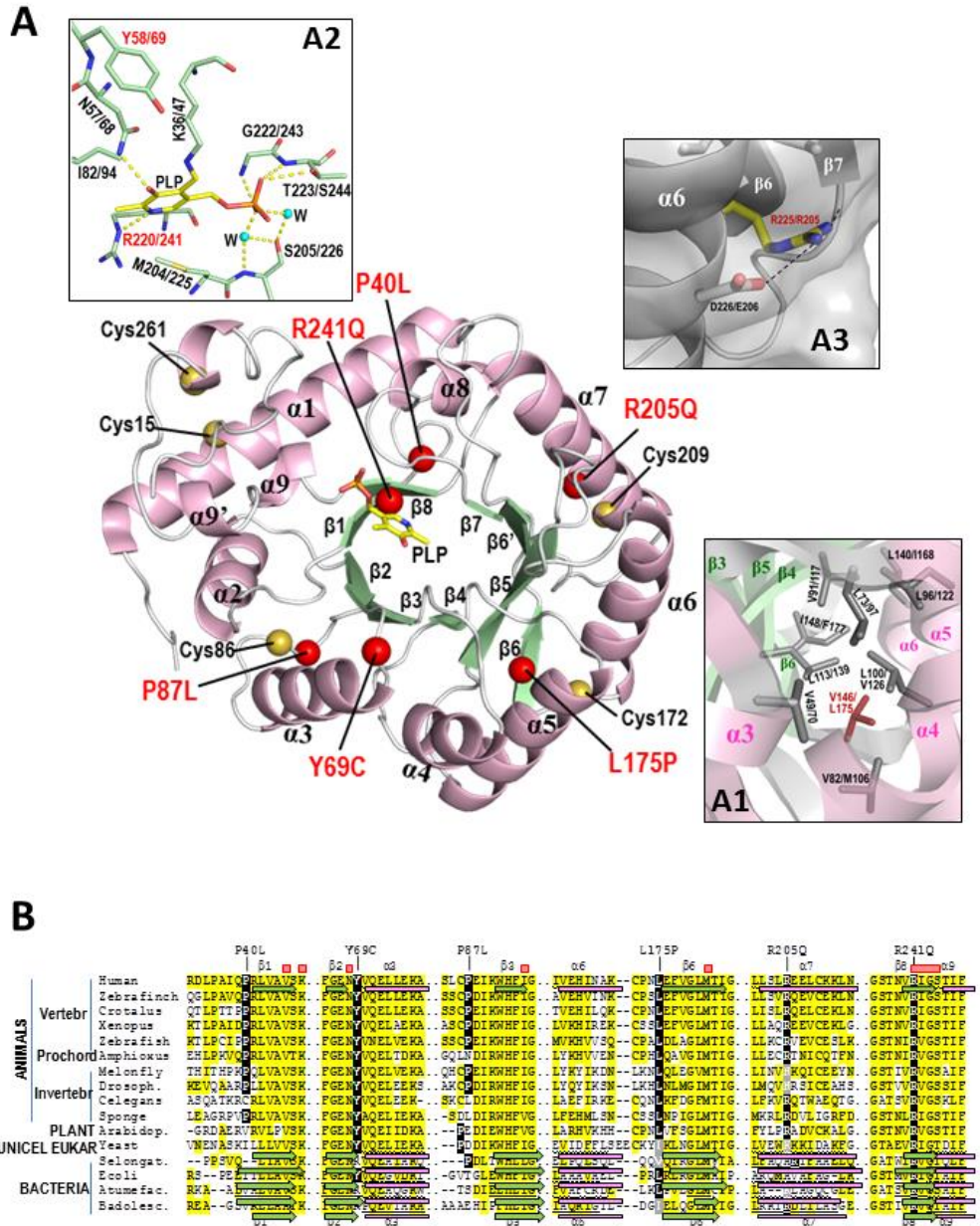


Figure 3. Structural mapping and conservation of the PLP-dependent enzymes residues found to host single amino acid substitutions in patients with vitamin B₆-dependent epilepsy. (A) 3-D model of PLP-dependent enzyme built with I-Tasser, in cartoon representation, with α helices (ribbons) and β strands (arrows) colored pink and green, respectively, and labeled. The structure is viewed along the barrel axis, with the C-edge of the circular β sheet being closest to the viewer. The residues hosting the mutations (labeled in red in single amino acid notation) are shown as red spheres. Cysteine residues are shown as yellow spheres and

are identified by the residue number. The bound PLP is also shown in sticks representation and labeled, with C, O, N and P atoms colored yellow, red, blue and orange, respectively. **Insets** zoom on specific regions of the protein, showing the residue rotamers found in the experimental 3D-structures of bacterial COG0325 proteins exhibiting full conservation of the mutated residues. Double residue numbering is given (bacterial residue/human residue). Residues hosting mutations are highlighted by giving their numbers in red. **Inset A1.** Hydrophobic nest that is disturbed by the p.Leu175Pro mutation, modeled on the basis of the structure of PipY (PDB file 5NM8), in ribbon representation, with hydrophobic side-chains in grey sticks (except Leu175, which is red-colored). **Inset A2.** The PLP site, modeled on the corresponding site for *E. coli* YggS (PDB 1W8G). **Inset A3:** Arg205 of PLPHP, modeled from the structure of the COG0325 protein of *B. adolescentis* (PDB entry 3CPG). **(B)** Alignment of short regions of the PLPHP sequence (in single letter amino acid code) where the patients' missense mutations map with the corresponding regions of representative members of the COG0325 family from bacteria to humans. The individual species and the taxonomic group are given to the left. The Clustal ω alignment has been manually modified to allow the stacking of the same secondary structure elements revealed in crystal structures of COG0325 family members and in the corresponding structural model of human PLPHP (see Materials and Methods). α helices and β strands are respectively mapped in the sequences as horizontal pink bars and green arrows, respectively. They are labeled above and below the stacked sequences according to their numbering given in panel A. Human PLPHP residues hosting the missense changes identified in patients are numbered according to their positions in the sequence, and, together with the identical or conservatively replaced corresponding residues of other COG0325 proteins, are shown in white lettering over a black background (invariance) or a grey background (conservative replacement). Identity or conservative replacement of other residues in a majority of the aligned sequences is highlighted by a yellow background. Residues that contact PLP in the crystal structures are marked at the top with red squares or, for several adjacent PLP-binding residues, with a red rectangle. The aligned COG0325 sequences correspond (from top to bottom) to the following species, Uniprot KB entries and, when existing, PDB identifiers: human, 094903; the bird *Taeniopygia guttata*, HOZ2K7; snake *Crotalus adamanteus*, J3S9B4; frog *Xenopus tropicalis*, F6TK15; zebrafish (*Danio rerio*), B1H1M3; prochordate *Branchiostoma floridae* (also known as amphioxus), C3Y5V7; insects *Zeugodacus cucurbitae* (melon fruit fly) and *Drosophila melanogaster*, A0A0A1WY34 and Q9VA97, respectively; worm *Caenorhabditis elegans*, P52057; sponge *Amphimedon queenslandica*, A0A1X7VVM2; plant *Arabidopsis thaliana*, Q9SZ26; and the unicellular organisms baker's yeast (*Saccharomyces cerevisiae*), P38197, PDB 1CT5; *Synechococcus elongatus* PCC7942, Q31LH9, PDB 5NM8; *Escherichia coli*, P67080, PDB 1W8G; *Agrobacterium tumefaciens*, A9CHE9, PDB 3R79); and *Bifidobacterium adolescentis*, A1A3G9, PDB 3CPG.

This explains the failure of this mutant to complement the pyridoxine toxicity phenotype of *yggS*⁻ *E. coli* (Darin et al., 2016) since this complementation is

known to depend on proper PLP binding (Ito et al., 2013). The somewhat decreased PLP content of the p.Tyr69Cys mutation may reflect the fact that Tyr69 is adjacent to an invariant PLP-binding residue (Asn68, Fig. 3A, inset A2; and Fig. 3B), covering it as a lid with its flat and exposed phenolic side chain. This lid also covers partially the extended side-chain of the invariant lysine that anchors covalently the PLP to the protein (Lys47 in PLPHP, Fig. 3A, inset A2). Concerning the covalent dimer formation with the p.Tyr69Cys mutant, such dimerization is understandable on the basis of the introduction of a highly exposed cysteine that most likely reacts with the same neo-cysteine of another subunit. As already indicated, this dimer formation might explain the decrease in the availability of the PLP for reaction with external agents, such as DCS, possibly reflecting also hampering of donation of PLPHP-bound PLP to its potential physiological partner PLP-requiring proteins.

For the other three missense mutations, p.Pro40Leu, p.Pro87Leu and p.Arg205Gln, the structure supports their proper binding of PLP, as observed, given their normal visible absorption spectra (Fig. 1E), since these mutations do not affect residues of the PLP site. The structure could suggest some destabilizing effect of the p.Arg205Gln mutation, which maps in helix 7 of the barrel scaffold and thus could de-structure somewhat this helix (Fig. 3A, inset A3). Effects for the p.Pro40Leu and p.Pro87Leu mutations are more difficult to infer from the structure, since these changes, although drastic at the amino acid level (Creighton, 1993), map in loops closely preceding β strands (β 1 and β 3, respectively), on the N-edge of the β -sheet, the opposite side of the barrel with respect to the side where PLP sits. Further structural studies involving X-ray crystallography may be necessary to fully unveil the structural consequences of these two mutations.

ACKNOWLEDGMENTS

We thank A. García and M. Orzáez (CIPF-Valencia) for CD facilities, and Silvia Ventas for technical help. This work was supported by grants from the Valencian (PrometeoII/2014/029) and Spanish (BFU2014-58229-P and BFU2017-84264-P) governments.

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249.
- Creighton, T.E. (1993). *Proteins: structures and molecular properties*. New York, NY: W.H. Freeman and Company.
- Darin, N., Reid, E., Prunetti, L., Samuelsson, L., Husain, R.A., Wilson, M., ... Clayton, P.T. (2016). Mutations in PROSC disrupt cellular pyridoxal phosphate homeostasis and cause vitamin-B₆-dependent epilepsy. *Am J Hum Genet* 99:1325–1337.
- Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
- Eswaramoorthy, S., Gerchman, S., Graziano, V., Kycia, H., Studier, F.W. & Swaminathan, S. (2003). Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta Crystallogr D Biol Crystallogr* 59, 127–135.
- Fasella, P. (1967) Pyridoxal phosphate. *Annu Rev Biochem.* 36:185-210.
- Gámez, A., Yuste-Checa, P., Brasil, S., Briso-Montiano, Á., Desviat, L.R., Ugarte, M., ... Pérez, B. (2017). Protein misfolding diseases: Prospects of pharmacological treatment. *Clin Genet*. In press, doi: 10.1111/cge.13088.
- Ito, T., Iimori, J., Takayama, S., Moriyama, A., Yamauchi, A., Hemmi, H. & Yoshimura, T. (2013). Conserved pyridoxal protein that regulates Ile and Val metabolism. *J Bacteriol* 195:5439–5449.

Ito, T., Yamauchi, A., Hemmi, H. & Yoshimura, T. (2016). Ophthalmic acid accumulation in an *Escherichia coli* mutant lacking the conserved pyridoxal 5'-phosphate-binding protein YggS. *J Biosci Bioeng* 122:689–693.

Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R. & Losick, R. (2010). D-amino acids trigger biofilm disassembly. *Science* 328:627–629.

Labella, J.I., Cantos, R., Espinosa, J., Forcada-Nadal, A., Rubio, V. & Contreras, A. (2017). PipY, a member of the conserved COG0325 family of PLP-binding proteins, expands the cyanobacterial nitrogen regulatory network. *Front Microbiol* 8:1244.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N, ... Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25:2744–2750.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., ... Richardson, D.C. (2003). Structure validation by C-alpha geometry: phi, psi and C-beta deviation. *Proteins* 50:437–450.

Neuhaus, F.C. (1967) D-Cycloserine and O-carbamyl-D-serine In D. Gottlieb & P.D. Shaw (Eds.), *Antibiotics I. Mechanism of Action* (pp. 40–83). Berlin, Heidelberg: Springer Verlag.

Plecko, B., Zweier, M., Begemann, A., Mathis, D., Schmitt, B., Striano, P., ... Rauch, A. (2017). Confirmation of mutations in PROSC as a novel cause of vitamin B₆-dependent epilepsy. *J Med Genet*. 54:809–814.

Prunetti, L., El Yacoubi, B., Schiavon, C.R., Kirkpatrick, E., Huang, L., Bailly, M., ... de Crécy-Lagard, V. (2016). Evidence that COG0325 proteins are involved in PLP homeostasis. *Microbiology* 162:694–706.

Relimpio, A., Iriarte, A., Chlebowski, J.F. & Martinez-Carrion M. (1981) Differential

scanning calorimetry of cytoplasmic aspartate transaminase. *J Biol Chem* 256:4478-4488.

Tremiño, L., Forcada-Nadal, A., Contreras, A. & Rubio, V. (2017). Studies on cyanobacterial protein PipY shed light on structure, potential functions, and vitamin B₆-dependent epilepsy. *FEBS Lett* 591:3431-3442.

Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Wasney, G.A., ... Edwards, A.M. (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* 103:15835–15840.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., ... Richardson, D.C. (2017). MolProbity: More and better reference data for improved all-atom structure validation. *Prot Science* 27:293-315.

Xu, D. & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* 101:2525–2534.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12:7–8.

Capítulo 3: Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria.

Autores: Lorena Tremiño, Javier Espinosa, Asunción Contreras y Vicente Rubio*

Manuscript to be submitted

Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria.

Lorena Tremiño¹, Javier Espinosa², Asunción Contreras² and Vicente Rubio^{1*}

¹ Instituto de Biomedicina de Valencia (IBV-CSIC), and CIBER de Enfermedades Raras (CIBERER-ISCIH), Valencia, Spain.

² Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain.

Manuscript to be submitted

Abstract

CutA, a non-canonical member of the PII signaling protein superfamily, is widely distributed among life domains. Unlike canonical PII, it is present in animals (including humans). Its gene was identified in *Escherichia coli* as a part of a Cu²⁺ protection locus. CutA is widely annotated as a Cu²⁺/divalent cation tolerance protein. We challenge such annotation since its gene *cutA1* failed to increase Cu²⁺ resistance of an *E.coli cutA1* mutant. In addition, in the phylogenetically distant and much more Cu²⁺-sensitive (relative to *E.coli*) cyanobacterium *Synechococcus elongatus* PCC7942, we knocked out the orthologous gene, confirming its non-essentiality. The mutant did not exhibit increased sensitivity to Cu²⁺, Zn²⁺ and Co²⁺. Thus, the *cutA*/CutA annotation in databases appears incorrect. Nevertheless, given the existence of reports of interaction with Cu²⁺ of CutA from *E. coli*, *Pyrococcus horikoshi* and *Arabidopsis thaliana* and a crystal structure of *P. horikoshi* CutA with Cu²⁺ bound, we examined the binding of Cu²⁺ to purified *S. elongatus* CutA, both in solution assays and by X-ray crystallography. *S. elongatus* CutA aggregated upon Cu²⁺ addition. Turbidimetric monitoring suggested a single type of Cu²⁺ site with an approximate K_D of 60 μM. High-resolution crystal structures of *S. elongatus*

CutA failed to reveal bound Cu^{2+} , showing for CutA the expected PII-like homotrimeric structure with the ferredoxin-type ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$) subunit fold which in CutA is followed by a $\beta_5\alpha_3$ C-terminal projection. This projection and the lack of the long flexible T-loop are the more conspicuous structural differences with PII. A Bis-Tris buffer molecule was bound to the intersubunit pockets. We managed to obtain CutA structures without Bis-Tris and with 1, 2 and 3 pockets occupied by this buffer. Despite earlier proposals of a paramount role of pocket filling for a potential allosterically-modulated signaling function of CutA, no conformational changes were observed upon pocket filling with Bis-Tris. The lack of structural flexibility and the high resolution of the structures fit our thermofluor observation of an extremely high thermal stability of *S. elongatus* CutA, agreeing with observations with CutA proteins from other organisms.

Keywords. Copper resistance, PII superfamily, *Synechococcus elongatus* PCC7942, thermostability.

INTRODUCTION

PII proteins are highly widespread and very ancient signaling molecules first reported in *Escherichia coli* as nitrogen metabolism controllers (reviewed in Stadtman, 2001). They are characteristic homotrimers presenting a $(\beta\alpha\beta)_2$ ferredoxin subunit fold (Xu et al., 2003). Their targets are other protein molecules, either enzymes, transporters or gene expression regulators (reviewed in Forcada-Nadal et al., 2018). Initially grouped into GlnK and GlnB classes depending on whether they were or were not involved on ammonia transport control (Thomas et al., 2000), they nucleate a superfamily that contains canonical PII proteins of known function as well as non-canonical members (Kinch and Grishin, 2002; Forchhammer and Lüddecke, 2016).

Among the non-canonical members of the PII superfamily CutA occupies a prominent place, given its near-universal occurrence (Perrier et al., 2000), inferring the presence of CutA already in very ancient life forms and certainly in the Last Universal Common Ancestor (LUCA). In fact, unlike canonical PII, which is not present in animals, CutA is constant in the animal kingdom, being also found in mammals including humans (Perrier et al., 2000). Interestingly, despite the close structural similarity of PII and CutA (Arnesano et al., 2003), CutA lacks significant sequence identity with canonical PII, having been identified as a member of the PII superfamily because of its structural similarity with PII and also because it was “fished” in transitive BLAST searches using as baits the sequences of some *nif3* proteins (COG3323 entry in the EggNOG database <http://eggnogdb.embl.de/#/app/results>), which are proteins that do exhibit significant homology with canonical PII (Kinch and Grishin, 2002).

There is little clear-cut and well-characterized information on CutA functions. CutA is widely annotated in databases as "*Divalent-cation tolerance protein CutA*", and "*Involved in resistance toward heavy metals*", because of the identification of the *Escherichia coli* gene encoding CutA, *cutA1*, in studies of metal sensitivity of this enterobacterium (Fong et al., 1995). In these studies, this gene was found as a part of a locus shown to influence the sensitivity of *E.coli* to Cu^{2+} and to other divalent metals. Of the three genes found in this locus, the most upstream gene was *cutA1*, which encodes CutA and is clustered with *cutA2* (Fong et al., 1995). This last gene, which encodes the periplasmic membrane-bound thiol-disulfide interchange protein DsbD (Missiakas et al., 1995), enhances the metal resistance of *E. coli*, possibly because of the importance of thiol groups for interaction with metals (Metheringham et al., 1996). We have re-analyzed the published data and found no clear-cut evidence for the proposed role of *cutA1* by itself on metal resistance.

Therefore, we have examined here whether *cutA1* is or is not important for metal resistance of *E. coli*, by carrying out Cu^{2+} sensitivity experiments of an *E. coli* strain allegedly having only *cutA1* inactivated, but found also by us to lack the beginning of *cutA2*. To clarify with this mutant the impact on metal sensitivity of *cutA1* alone, we have carried out complementation assays with this isolated gene (carried in a plasmid). The results of these assays have not supported a role of *cutA1* in metal protection of *E. coli*.

With the purpose of generalization of the *E. coli* findings to other bacterial species, we have inactivated the *cutA* orthologous gene of the cyanobacteria *Synechococcus elongatus* PCC 7942 (from now on *S. elongatus*), monitoring potential changes in metal sensitivity of this organism. *S. elongatus* was chosen because of its phylogenetic distance from *E. coli* (Hedges, 2002) and also because we could not find within its genome an orthologous gene for *cutA2* (BLASTP search using *E. coli* DsbD protein sequence; data not shown), thus excluding any *cutA2* influence, and also because in preliminary experiments we found *S. elongatus* to be much more sensitive to divalent metals including Cu^{2+} than *E. coli* (see Results below). We were also inclined to use *S. elongatus* because it has become a true paradigm for the PII regulatory network, with a very well characterized regulatory universe including many crystal structures of PII and the complexes that belong to this regulatory network (reviewed in Forcada-Nadal et al., 2018). We felt that, for the purpose of completeness of characterization of this model organism, it could be worth it to study in it, too, non-canonical members of the PII superfamily, a study already started in cyanobacteria with another non-canonical member, SbtB (Selim et al., 2018). Again, we report here lack of effect of CutA inactivation on the sensitivity of *S. elongatus* to divalent metals.

Our finding of lack of metal protection by *cutA1* puzzled us because previously some CutA proteins had been reported to bind Cu^{2+} (Arnesano et al., 2003; Burkhead et al., 2003; Tanaka et al., 2004a) and even CutA from *Pyrococcus horikoshi* had been crystallized and found to host in its structure (determined by X-ray crystallography) bound Cu^{2+} ions (Tanaka et al., 2004a) although in a more recent publication (Bagautdinov, 2014) no divalent metal was found in crystals of *P. horikoshi* CutA that were grown in the presence of Cu^{2+} .

We therefore decided to examine the binding of Cu^{2+} to recombinantly expressed and purified *S. elongatus* CutA, both in solution assays and by X-ray crystallography. The results of these studies are presented here. They have provided the crystal structure of highest resolution reported thus far for a CutA protein, although no structure could be produced with bound Cu^{2+} , a cation that caused aggregation. These studies have also revealed that, as in the case of other mesophiles (Matsuura et al., 2018), CutA had very high thermal stability. One interesting observation made in our crystal structures that may have an important bearing on the potential functions of CutA was the observation of cargo molecules in the three intersubunit cavities found in the *S. elongatus* CutA trimer. Although these molecules were non-natural and did not trigger any structural changes on CutA, our finding might be a first step towards other studies to find natural ligands for these pockets that could possibly induce structural changes conducting to a proposed (Arnesano et al., 2003; Laskowski et al., 2003) signaling role of CutA.

MATERIALS AND METHODS

***E. coli* copper tolerance assays and complementation studies.**

Wild type *E. coli* (strain BW25113) and the corresponding *cutA* lacking strain JW4097 of the Keio collection (Baba et al., 2006), in which *cutA1* is replaced by a kanamycin resistance cassette, were provided by Dr. F.J. Mojica, Universidad de Alicante. Both strains were lysogenized for phage λ DE3 using the λ DE3 Lysogenization Kit (as recommended in the kit, from Novagen), thus to allow T7 promoter-dependent expression of pET15b-carried genes. *E. coli cutA1* and *cutA2* genes (EG12177 and EG12178 Ecogene identification codes, <http://ecogene.org/>), encoding respectively CutA and DsbD, were separately PCR-cloned from genomic DNA of *E. coli* K12 (using Deep Vent polymerase and the primer pairs *cutA1-NdeI-F/cutA1-BamHI-R* and *cutA2-NdeI-F/cutA2-BamHI-R*, Table 1) into the *NdeI-BamHI* sites of pET15b, to yield plasmids pET*cutA1* and pET*cutA2*, which encode the corresponding proteins N-terminally fused to the His₆-tag MGSSHHHHHSSGLVPRGSH. Automated fluorescent Sanger sequencing was used to confirm the correctness of the constructions and the absence of mutations as well as to determine the insertion point of the kanamycin resistance cassette in the genome of the JW4097 *E. coli* strain and the lack of insertion of this cassette in the wild-type strain (primers for PCR amplification of the region including the insertion point, P1 and P2, Table 1). Standard procedures were used for making electrocompetent cells (Green and Sambrook, 2012) and for electrotransformation with plasmids (Fiedler and Wirth, 1988), selecting transformants by culturing under aeration at 37°C in LB-ampicillin.

For Cu²⁺ resistance assays the λ DE3-lysogenized BW25113 (WT) and JW4097 cells were grown to identical cell densities (estimated from OD^{600(1cm)}), assuming that 1 OD corresponds to 10⁹ cells/ml; Elbing and Brent, 2002) and then they were diluted to the indicated concentrations in sterile water. Equal size drops of these dilutions were seeded in LB-agar supplemented with or without 3 mM

CuSO₄. Growth in individual drops was checked by naked eye after 18 hours of plate incubation at 37 °C. Similar assays were carried out with the lysogenized JW4097 strain that had been transformed with either "empty" parental pET15b or the pET*cutA1* or pET*cutA2* plasmids.

Table 1. Oligonucleotides used in the present study. Restriction sites for NdeI and BamHI are underlined.

Name	Sequence (5' - 3')
<i>cutA1</i> -NdeI-F	TCCGCT <u>CATATG</u> CTTGATGAAAAAAGTTTCG
<i>cutA1</i> -BamHI-R	TATTAAGGATCCTCAGCGTAAAGATGCGTTGAGC
<i>cutA2</i> -NdeI-F	TCCGCT <u>CATATG</u> GCTCAACGCATCTTTACG
<i>cutA2</i> -BamHI-R	TATTAAGGATCCTCACGGTTGGCGATCGCGC
P1	TAGCTTCATGCTGTAATGATCAATCGCG
P2	TAGAGGTAGTAACCGTCTTTGATCTGCC
1F	GAAGGCGGTGCTG
1R	AATGTATCAGAAGCAACAG
CutA-NdeI-F	ATCACAC <u>CATATG</u> ACTGACCTCAGTTCCCTG
CutA-BamHI-R	TGATCTGGATCCTTAGGAGAGGCTCGAACG

Preparation of the *S. elongatus* Δ *cutA* mutant and determination of metal tolerance in this organism.

To inactivate *cutA* in *S. elongatus* PCC 7942, the wild-type strain was transformed with cosmid 8S34-E4 taken from the Unigene set, an existing insertion mutant library for *S. elongatus* (Holtman et al., 2005) which carries the CutA ORF (gene identifier *Synpcc7942_2261*) inactivated with a Tn5 transposon conferring resistance to kanamycin. Transformation of *S. elongatus* was achieved using standard protocols (Clerico et al., 2007) and resulted in two classes of transformants: double recombinants, that substitute the wild-type

allele by the mutant allele, and single recombinants that result from the integration of the entire cosmid with duplication of the targeted gene. The two types of transformants were distinguished by the Cm marker on the cosmid backbone, which is present only in single recombinants. Two clones Km^R Cm^S were isolated and the correct inactivation of *cutA* was verified by PCR with primers 1F and 1R (Table 1).

S. elongatus strains were routinely grown photoautotrophically on BG11 medium while shaking at 30 °C under continuous illumination conditions (~30 μE m⁻² s⁻¹) (Labella et al., 2017), checking cell density as OD⁷⁵⁰. Culture biomass was estimated as chlorophyll content from the OD⁶⁶⁵ of a centrifuged methanolic extract of 0.1 ml of culture (Elbing and Brent, 2002). Cultures were adjusted to 10 μg chlorophyll/mL and then serial dilutions of the cell suspension were prepared and drop-seeded as for *E.coli* (see above) in solid 1% agar plates of BG11 medium supplemented with metals as indicated. Plates were photographed after 5 days of growth at 30 °C under constant illumination.

Production of recombinant *S. elongatus* CutA.

The *S. elongatus cutA* gene (Cyanobase gene identifier, *Synpcc7942_2261*), PCR-amplified from genomic DNA using Deep-Vent DNA polymerase (New England Biolabs) and the CutA-NdeI-F and CutA-BamHI-R oligonucleotides (Table 1), was cloned into pJET1.2/blunt with the CloneJET PCR Cloning Kit (ThermoFisher Scientific). Then, after release from this plasmid by *NdeI/BamHI* double digestion, it was subcloned directionally into the corresponding sites of pET28a (Novagen) using Quick Ligase (New England Biolabs). The resulting plasmid encoding *S.elongatus* CutA N-terminally fused to the MGSSHHHHHSSGLVPRGSH tag, was used to transform *E. coli* BL21(DE3) cells (Invitrogen), which were grown at 37°C to a cell density of 0.7 OD⁶⁰⁰ in 0.5-L well-aerated cultures in liquid LB-kanamycin (50 μg/ml) medium. Then the

temperature was lowered to 20°C, 0.1 mM IPTG was added, and the culture was continued at 20°C for 20 hours.

Subsequent steps were at 0°C. Centrifugally harvested cells from a 500-ml culture were suspended in 12 ml of 0.1 M Tris-HCl pH 8, 0.15 M NaCl, 1 mM EDTA and were disrupted by sonication, centrifuged (30 min, 11,000×g), and the supernatant was applied to a 1-ml Ni-chelate HisTrap-HP column fitted in an ÄKTA-FPLC system (both from GE Healthcare), followed by a 30-ml wash with the same buffer with the EDTA replaced by 20 mM imidazole. Cut A was eluted with an 80-ml linear gradient of this buffer containing 20–500 mM imidazole, collecting fractions. The purest fractions (purity monitored by SDS-PAGE) were pooled, and the imidazole was removed and the protein was concentrated to ≥ 7 mg protein/ml by centrifugal ultrafiltration (Amicon Ultra, from Millipore; 10 kDa cutoff membrane).

Protein crystallization, X-ray diffraction and structure determination.

Robotized crystallization screens by vapor diffusion at 21°C in sitting drops of 0.4 μ l protein solution and 0.4 μ l of reservoir solution (screening solutions from Qiagen and Molecular Dimensions) yielded in < 2 weeks crystals under several conditions. Four crystals that appeared under different conditions were the ones diffracting best X-rays. They are called here L0, L1, L2 and L3. They arose from protein solutions of 7-10 mg/ml, using the reservoir solutions indicated in Table 2. The crystals, flash frozen in liquid nitrogen with no need for additional cryoprotectants, were subjected to diffraction at 100K using synchrotron X-rays (Alba and Diamond synchrotrons, in, respectively, Barcelona Spain, and Oxfordshire, UK). Diffraction data were collected at 1.17-2.00 Å resolution (Table 2) with Dectris PILATUS \times 6 M detectors for L0, L1 and L3, and with PSI Pilatus \times 6M for L2. They were processed with XIA2 (L2 crystal) (Winter et al., 2013) or XDS (Kabsch, 2010), using AIMLESS for scaling (Evans

Capítulo 3

Table 2. Crystallization, X-ray data collection and refinement statistics for the crystals of SeCutA studied here

Crystal	L0	L1	L2 (Cu ²⁺ -crystal)	L3
PDB file identifier	6GDU	6GDV	6GDW	6GDY
Crystallization (21° C, 0.8 µl 1/1 sitting drops)				
Protein solution	7.1 mg/ml CutA, 10 mM of both ATP and MgCl ₂	10 mg/ml CutA	7.1 mg/ml CutA	7.3 mg/ml CutA
Crystallization solution	0.1 M MIB ^a buffer pH 4, 25% PEG 1500	0.1 M Bis-Tris pH 5.5, 0.2 M Li ₂ SO ₄ , 25% PEG 3350	0.1 M Bis-Tris pH 6.5, 29% PEG 3350, 0.22 mM CuSO ₄	0.1 M Bis-Tris pH 6.5, 23% PEG 3350
Data collection ^a				
Synchrotron/Beamline	ALBA / BL13 - XALOC		Diamond/I03	ALBA / BL13- XALOC
Wavelength (Å)	0.9794	0.9794	0.9795	0.9794
Space group	P12 ₁ 1	P12 ₁ 1	P12 ₁ 1	P12 ₁ 1
Unit cell parameters				
a, b, c (Å)	35.2, 96.4, 49.0	34.6, 96.6, 48.9	34.7, 95.8, 48.8	35.6, 96.3, 48.9
α, β, γ, (°)	90.0, 94.2, 90.0	90.0, 97.4, 90.0	90.0, 97.1, 90.0	90.0, 93.2, 90.0
Solvent (%)	42	45	42	44
Resolution range (Å)	48.19-1.75 (1.78-1.75)	48.32-2.00 (2.05-2.00)	47.91-1.80 (1.80 - 1.85)	48.84 -1.17 (1.19-1.17)
Reflections, total/unique	236490/32752 (13239/1772)	55644/20245 (4155/1498)	93757 / 28906 (6714/2129)	504431/110414 (21341/5494)
Completeness (%)	99.9 (100.0)	94.2 (93.2)	98.7 (98.8)	99.9 (99.6)
Multiplicity	7.2 (7.5)	2.7 (2.8)	3.2 (3.2)	4.6 (3.9)
I/σ	11.2 (3.2)	5.6 (1.6)	10.7 (1.8)	21.6 (5.3)
R _{pim} ^b (%)	4.1 (21.8)	7.6 (23.5)	-	1.7 (13.0)
R _{meas} (%)	-	-	10.5 (96.1)	-
Refinement				
Resolution (Å)	48.19 - 1.75 (1.87-1.75)	48.32 - 2.00 (2.05-2.00)	47.91-1.80 (1.85 -1.80)	48.84 -1.17 (1.19-1.17)
Reflections, work/test	31114 / 1368	19233 / 1012	27461 / 1445	104893 / 5521
R _{factor} ^c /R _{free} ^d (%)	19.9 / 23.3	19.2 / 25.5	20.3 / 24.4	14.0 / 16.3
r.m.s.d. from ideal				
Bond length (Å)	0.011	0.013	0.014	0.014
Bond angle (°)	1.457	1.570	1.646	1.646
Number of:				
Polypeptide chains	3	3	3	3
Protein atoms	2462	2521	2447	2554
Solvent molecules	122	135	131	366
Ligand molecules	0	1 Bis-Tris	2 Bis-Tris	3 Bis-Tris
Average B-factor(Å ²)				
Protein	15.49	25.20	31.6	14.88
H ₂ O	33.37	33.30	42.49	30.76
Ligand molecules	-	27.36	48.00	17.93
Ramachandran plot ^e (%)				
Favoured	100.0	100.0	100.0	100.0
Allowed	0.0	0.0	0.0	0.0
Outliers	0.0	0.0	0.0	0.0

^aCommercial proprietary buffer from Molecular Dimensions

^bValues in parentheses are the data for the highest-resolution shell.

^c $R_{pim} = \sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \langle I_{hkl} \rangle$.

^d $R_{factor} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs}$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

^e R_{free} is the same as R_{factor} but calculated for the 5% data omitted from the refinement.

^fCalculated with RAMPAGE.

and Murshudov, 2013). The space group was P12₁1 for all four crystals, with similar cell dimensions and angles in the four cases (Table 2).

Initial phases were obtained for the L1 crystal by molecular replacement with Phaser (McCoy, 2007; McCoy et al., 2007) using a polyalanine model of the structure of subunit A of *E.coli* CutA (Arnesano et al., 2003; Protein DataBank entry 1NAQ), yielding a typical CutA homotrimer in the asymmetric unit. After refinement (see below) to 2.0 Å-resolution, the subunit A was used for phasing all other crystals, revealing in all cases one CutA trimer in the asymmetric unit. Rigid body refinement using Refmac5 (Murshudov et al., 2011) was used for optimizing the positions of the three subunits in each trimer. Refmac5 was also used for model refinement, alternating iteratively cycles of restrained refinement and of manual model building with Coot (Emsley et al., 2010). Ligands, when necessary, were incorporated into the model with Coot. Isotropic B factors and anisotropic refinement (with Refmac5) were used in the last steps of refinement of L2, whereas for L0, L1 and L3, TLS was used in the last refinement steps, the TLS groups being chosen with the TLSMD server (<http://skuld.bmsc.washington.edu/~tlsmd/>). All diffraction data were used throughout the refinement processes, except the 5% randomly selected data used to calculate the R_{free} . The stereochemistry of each model was checked and improved with PDB_REDO (Joosten et al., 2009). Geometry analysis of the protein main chain torsion angles with Rampage (Lovell et al., 2003) revealed excellent values for the models. Graphical representations of the structure were generated with PyMOL (<http://www.pymol.org/>). Root mean square deviation (r.m.s.d.) values were calculated by superimposition of the subunits in Coot. Conservation analysis of CutA protein was performed with ConSurf Server (<http://consurftest.tau.ac.il/>) (Armon et al., 2001; Goldenberg et al., 2009).

Other methods.

Copper-induced aggregation of CutA was monitored turbidimetrically as increase in the absorbance at 400 nm (1-cm light path) with time (5-min intervals) of a solution at room temperature of CutA (80 µg/ml corresponding to 5.4 µM CutA subunits) in 0.1 M Tris-HCl pH 7.5 and 0.15M NaCl containing the indicated concentrations of CuSO₄ (NaOH-neutralized). The program GraphPadPrism was used for representing the data and for fitting sigmoidal curves to the experimental observations.

Size-exclusion chromatography of CutA (0.02-0.1 mg applied in 0.05 ml) or of standards of known mass was carried out on a SuperdexTM 75 10/300 GL column (GE healthcare) mounted on a ÄKTA FPLC system run at 0.4 ml/minute of a solution of 0.1 M Tris-HCl pH 8 and 0.15 M NaCl, with monitoring of the optical absorption of the effluent at 280 nm. A semilogarithmic linear plot of molecular mass (logarithmic, y-axis) versus elution volume (linear scale, X-axis) was used to estimate the oligomeric nature of CutA in solution

Thermal stability of CutA was assessed by the thermofluor approach (Vedadi et al., 2006) performed in duplicate on a real-time PCR instrument (7500 model from Applied Biosystems, ThermoFisher Scientific, Alcobendas, Madrid, Spain), monitoring the increase in the fluorescence due to SYPRO Orange (from Invitrogen, Carlsbad, CA) upon gradual (1°C/min) temperature increase of solutions of 10 µM CutA in 0.1 M TrisHCl pH 8, 0.15 mM NaCl and SYPRO Orange (a 1:1000 dilution of the commercial preparation from Applied Biosciences).

Protein was assayed according to Bradford (1976), with a commercial reagent from Bio-Rad, using bovine serum albumin as a standard. SDS-PAGE was carried out in 15% polyacrylamide gels with Coomassie staining (Laemmli, 1970).

RESULTS AND DISCUSSION

***E. coli* JW4097 strain, a *cutA1* deletion mutant, exhibits decreased tolerance to copper that is restored by transformation with *cutA2* but not with *cutA1*.**

In the original experiments that led to infer that *cutA1* is important for metal resistance of *E. coli* (Fong et al., 1995), one of six mutants causing increased cellular copper sensitivity was found to carry a deletion of a region that encompasses the adjacent genes *cutA1* and *cutA2*. In that work, complementation with lambda clones containing *E. coli* chromosomal DNA fragments clearly showed that a complete *cutA2* gene was required for complementation, but it remained unclear whether *cutA1* was or was not necessary for restoring metal resistance, since the two DNA fragments that exhibited complementing activity contained both *cutA1* and *cutA2* (Fong et al., 1995).

We reassessed the roles of *cutA1* in conferring *E. coli* resistance to copper by using strain JW4097 of the Keio collection (Baba et al., 2006), in which the *cutA1* gene of *E. coli* K-12 is replaced by a kanamycin resistance cassette (Fig. 1A). To enable the use of a pET15b expression plasmid carrying *cutA1* for complementation studies, we first lysogenized the JW4097 cells as well as the parental wild-type strain BW25113 (Baba et al., 2006) with phage λ DE3, which carries the gene for T7 RNA polymerase. This lysogenization had no effect on the metal sensitivity of the wild type or the JW4097 mutant strains (data not shown). The sensitivity to copper ions of the JW4097 strain was considerably higher than that of the parental wild-type strain, as shown in drop cultures in which a 10^4 - 10^5 -fold higher cell inoculum was required with the JW4097 mutant than with the wild-type strain for growth in the presence of 3 mM CuSO_4 (Fig. 1B). However, this finding cannot be attributed conclusively to the

lack of *cutA1*, since the initial 25 bases of the *cutA2* coding sequence overlap with the last bases of *cutA1* (Fig. 1A), and thus a polar effect with lack of *cutA2* expression was anticipated in the mutant.

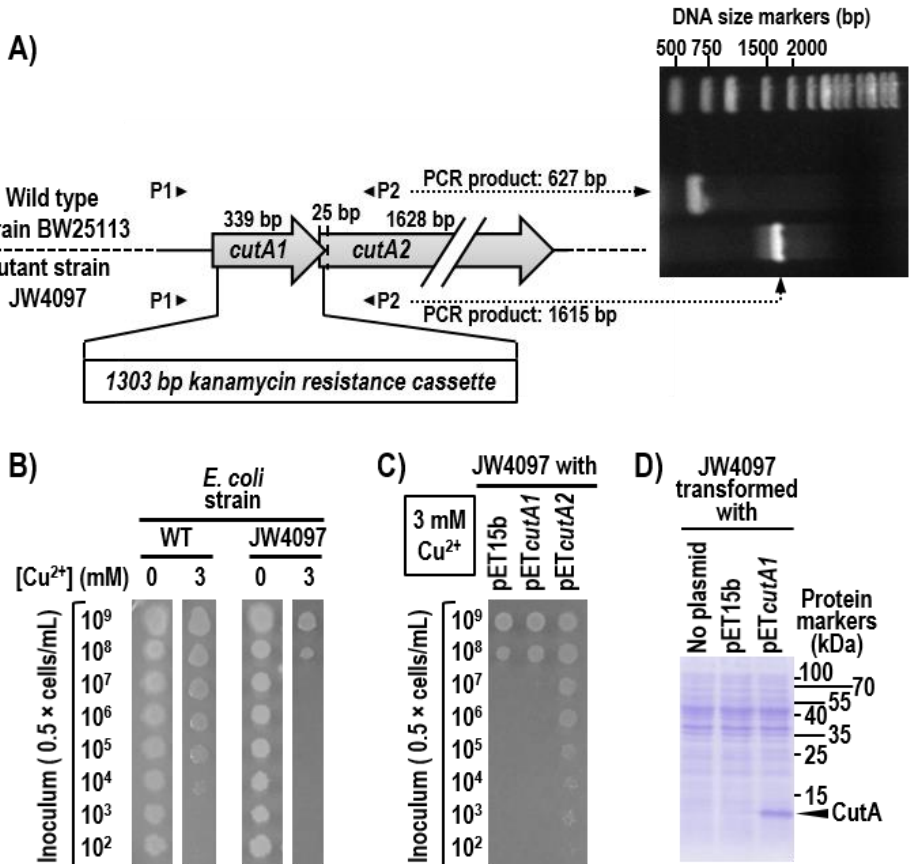


Figure 1. Copper sensitivity of *E. coli cutA1* mutant and lack of complementation by transformation with *cutA1*-carrying plasmid. **(A)** *cutA1-cutA2* gene cluster of *E. coli* K-12 in wild-type (WT) strain BW25113, and in mutant strain JW4097, in which *cutA1* is replaced by a kanamycin resistance cassette. To the right, electrophoretic analysis of the products obtained from the two strains after PCR-amplification of the region encompassed between oligonucleotides P1 and P2 (Table 1). **(B,C)** Drop cultures of decreasing (from top to bottom) cell concentrations of both strains (lysogenized with λ DE3 phage, see Materials and Methods) on agar-LB medium lacking or containing 3 mM copper sulphate as indicated. In **(C)** the cells were transformed with the indicated plasmid (see Table 1 and text). **(D)** SDS-PAGE analysis (15% polyacrylamide gel) of 4×10^8 cells of *E. coli* JW4097 transformed as indicated. The gel was stained with Coomassie blue. The arrow marks the position of CutA protein.

Therefore, we examined separately the ability of *cutA1* and *cutA2* to complement this mutant strain in the copper sensitivity assays by transforming strain JW4097 with pET15 plasmids carrying either *cutA1* or *cutA2* (Fig. 1C). While resistance to Cu^{2+} was restored to wild-type level by transformation with *cutA2* even in the absence of IPTG (the inducer of the T7 promoter; this finding indicates promoter leakiness), *cutA1* did not provide any growth advantage in the presence of 3 mM Cu^{2+} over the same strain transformed with the empty parental plasmid (Fig. 1C). The lack of complementation by *cutA1* was not due to lack of *cutA1* expression, as proven by SDS-PAGE of cell extracts, which showed a band with the expected mass for CutA that was not present in the cells transformed with empty parental plasmid (Fig. 1D).

Deletion of *cutA* increases the sensitivity to metals of *S. elongatus*.

The ability of *cutA2* to completely restore Cu^{2+} resistance to the JW4097 strain indicated that *cutA1* is not involved in determining the sensitivity to Cu^{2+} in *E. coli*. Since *cutA1* orthologs are very widespread among bacterial genomes, we attempted generalization of our results with *cutA1* to other bacterial species by turning to *Synechococcus elongatus* PCC7942 (hereafter *S. elongatus*). This model system appears well chosen for such a study because it is phylogenetically quite distant from *E. coli*, it contains a single *cutA1* ortholog (called here *cutA*; *Synpcc7942_2261* Cyanobase identification code), and since this gene is not physically associated in the genome with an adjacent ortholog of *cutA2*. In fact, there appears not to be any *cutA2* ortholog in *S. elongatus*, as shown by annotation datamining in Cyanobase or by performing a P-blast on the whole genome of this cyanobacterium.

We inactivated the *S. elongatus cutA* gene by allelic replacement by double recombination with the kanamycin-resistance allele *cutA::Tn5* from cosmid 8S34-E4 (Holtman et al., 2005) which is schematically illustrated in Fig. 2A. After

transformation of *S. elongatus* with cosmid 8S34-E4, candidate clones (Km^R Cm^S) were isolated and subjected to PCR analyses. Complete segregation of the inactive *cutA*::Tn5 alleles was easily achieved (Fig. 2A), confirming the prior finding (Rubin et al., 2015) that *cutA* is not an essential gene in *S. elongatus*.

Next, we compared the sensitivity of *S. elongatus* wild-type and *cutA* mutant to three transition metals (Cu²⁺, Zn²⁺ and Co²⁺). To this end, cells grown to mid-exponential phase were harvested and spotted on the surface of BG11-agar plates supplemented or not with the indicated transition metal. As illustrated in Fig. 2B, the addition of these metals significantly impaired cell growth, judged from the number of cells in the inoculum required to observe growth, but no differences in the sensitivities to metals were observed between the wild-type and the *cutA* mutant, indicating that *S. elongatus* CutA (from now on SeCutA) does not have a protective role towards transition metals. Therefore, the annotation of the *Synpcc7942_2261* gene as "*periplasmic divalent cation tolerance protein*" found in Cyanobase appears not to be correct. Interestingly, *S. elongatus* was much more sensitive to Cu²⁺ than *E. coli* (compare Fig. 2B with Fig. 1B) and thus it is tempting to speculate that this sensitivity might be related to the absence of a recognizable ortholog of *E. coli cutA2* in this cyanobacterium.

Evidence for Cu²⁺ binding to *S. elongatus* CutA

Despite our failure to detect decreased Cu²⁺ resistance of the *cutA* mutant of *S. elongatus*, we decided to analyze whether pure *S. elongatus* CutA interacts with Cu²⁺. Thus, we prepared this protein in pure form after expressing it recombinantly in *E. coli* (Fig. 3A). As anticipated, this protein behaved in size exclusion chromatography as expected for a trimer (Fig. 3B).

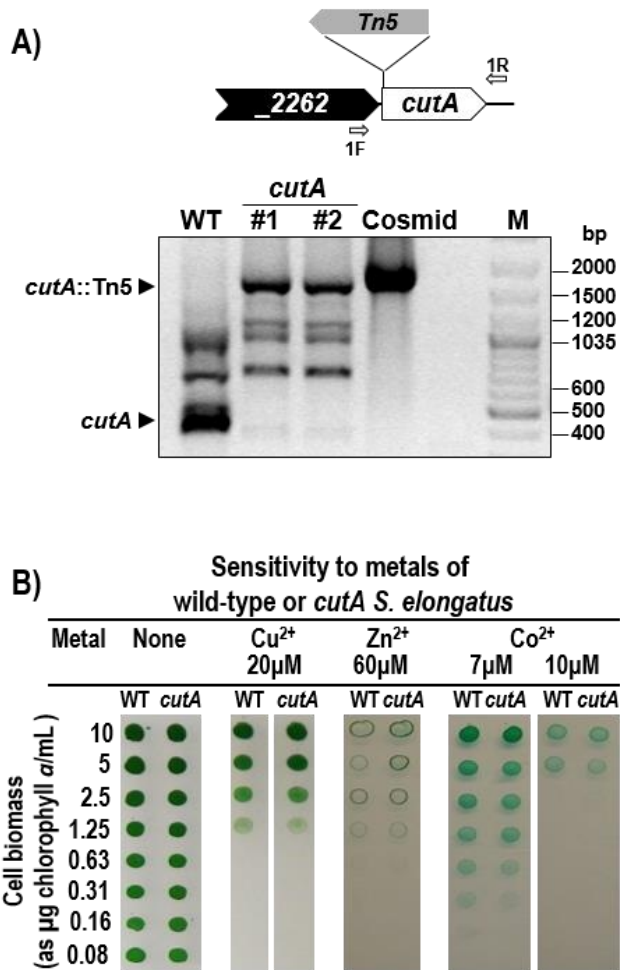


Figure 2. Normal tolerance to metals of *S. elongatus cutA* mutant. (A) Inactivation of *cutA* gene in *S. elongatus*. Schematic representation of the *cutA* genomic region with indication of the site of Tn5 insertion. Two mutant *cutA* clones (#1 and #2) alongside parental *S. elongatus* wild-type (WT) and the cosmid carrying *cutA::Tn5* were PCR-analysed with primers 1F and 1R. Detected alleles and reference size bands are indicated, respectively, on the left and the right. (B) Wild-type and *cutA* mutant growth on BG11 solid media supplemented with either 20 μM CuSO₄, 60 μM ZnSO₄ or the indicated concentrations of CoCl₂. Serial dilutions of a cell cultured (cell biomass assessed as chlorophyll *a* concentration in the culture) were spotted on the surface of agar plates which were photographed after 5 days of incubation at 30 °C under constant illumination conditions.

We obtained qualitative evidence of Cu^{2+} binding by the observation that when Cu^{2+} was added to *S. elongatus* CutA solutions, this protein aggregated. Although this prevented any attempt of spectrophotometric detection of characteristic absorption bands of Cu^{2+} -protein complexes, we took advantage of this aggregation to analyze turbidimetrically (as optical absorption at 400 nm) the formation of protein aggregates with time in diluted solutions of CutA (~5.4 μM CutA polypeptide, corresponding to nearly 2 μM trimers). In all assays, we used Cu^{2+} concentrations in considerable excess over the concentration of CutA subunits. As might have been expected for a process in which aggregates result from accretion of protein to aggregation nuclei that initially give little turbidity, the turbidity in the solution increased upon Cu^{2+} addition following a sigmoidal time-course reaching a maximum at infinite time that was dependent on Cu^{2+} concentration (Fig. 3C). This dependency of the maximum turbidity on the $[\text{Cu}^{2+}]$ could be fitted by a hyperbola (Fig. 3D, open circles) with a K_D for Cu^{2+} (designated as $K_D^{\text{t}\infty}$ in Fig. 3D) of $61 \pm 14 \mu\text{M}$. This value is similar to the K_D of ~0.1 mM reported for Cu^{2+} binding to *E. coli* CutA (Arnesano et al., 2003). In turn, the time required for reaching 50% of the maximum turbidity, $t_{1/2}$, decreased with increasing Cu^{2+} concentration (Fig. 3C, closed circles), with a relation between $t_{1/2}$ and Cu^{2+} concentration fitting hyperbolic decay, although with an apparent K_D for Cu^{2+} (called K_D^{rate} in Fig. 3C) of $0.29 \pm 0.04 \text{ mM}$, a value nearly 5-fold higher than the $K_D^{\text{t}\infty}$. This higher K_D value can be expected if the species that aggregates is the CutA trimer having its three subunits bound to Cu^{2+} , which would be a species being prone to make the extensive intermolecular network required for producing large aggregates. The fractional saturation of trimers with increasing Cu^{2+} concentration should equal the cube of the fractional saturation of the subunit, and its dependency with Cu^{2+} concentration over the range of concentrations used here approxim-

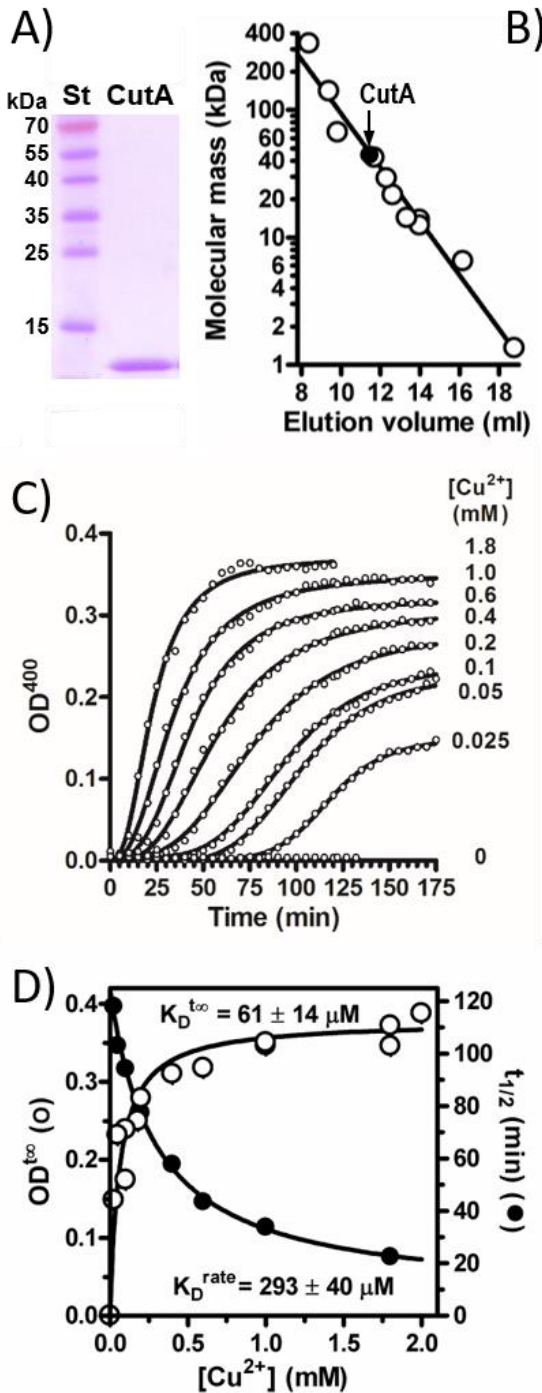


Figure 3. Pure *Synechococcus elongatus* CutA is a homotrimer in solution and binds Cu^{2+} . **(A)** SDS-PAGE (15% polyacrylamide; Coomassie staining) of isolated recombinant *S. elongatus* CutA to illustrate the high degree of purity of the protein (track labeled CutA). St, prestained molecular mass standards, with masses (in kDa) given at the side. **(B)** Plot correlating the masses of protein standards (open circles) with their elution volumes from the size exclusion chromatography column (see Materials and Methods), showing also the elution position of CutA (black circle) assuming that it is a homotrimer (sequence-deduced mass, 44 kDa). The standards used were (masses in kDa in parenthesis) bovine liver glutamate dehydrogenase (332.4), rabbit muscle lactate dehydrogenase (146), bovine serum albumin (66.4), maltose binding protein (42), bovine erythrocyte carbonic anhydrase (29), soybean trypsin inhibitor (21.5), cow's milk α -lactalbumin (14.2), bovine pancreas ribonuclease A (13.7), equine heart cytochrome C (12.4), bovine lung aprotinin (6.5) and vitamin B12 (1.35). **(C)** Time and $[Cu^{2+}]$ dependence of the turbidity (determined as optical absorbance at 400 nm, OD^{400}) of solutions of 84 $\mu\text{g/ml}$ of SeCutA. For details, see Materials and Methods. The lines correspond to sigmoidal dependency with respect to time according to the equation: $OD = OD^{t\infty} \times t^N / (t_{1/2}^N + t^N)$; where OD and $OD^{t\infty}$ are the absorbances at 400 nm at a given time and at $t=\infty$, respectively; $t_{1/2}$ is the time at which $OD=0.5 \times OD^{t\infty}$ and N is an exponent. **(D)** Dependency of $OD^{t\infty}$ and of $t_{1/2}$ on Cu^{2+} concentration. The lines drawn are hyperbolae for the apparent K_D values given.

ates a hyperbola for a K_D of the order of 5-fold the K_D for the fractional saturation of the subunit (not shown), in excellent agreement with our experimental results. In summary, the present results appear to agree with the existence of a single type of site for Cu^{2+} in each subunit of *S. elongatus* CutA having a K_D value for Cu^{2+} of about 60 μM , similar to the K_D for Cu^{2+} of *E. coli* CutA, and they also agree with the view that the Cu^{2+} -saturated trimer (CutA-Cu^{2+})₃ is the main molecular species responsible for the turbidity-producing aggregation.

In any case, the K_D^{too} inferred from our aggregation studies, believed to reflect Cu^{2+} binding to the CutA subunit, is considerably larger than the Cu^{2+} concentrations that are toxic for *S. elongatus* in our *in vivo* studies (Fig. 2B). This difference does not support, in principle, a protective role of Cu^{2+} binding by CutA for this cyanobacterium, in agreement with the lack of a negative effect on Cu^{2+} resistance of knocking out the *cutA* gene of *S. elongatus*.

***S. elongatus* CutA structure fails to reveal Cu^{2+} binding**

In an attempt to identify potential Cu^{2+} binding sites in SeCutA on the protein structure, we tried to get crystals that we could use for soaking in Cu^{2+} solutions. Table 2 summarizes crystallographic data for the four different crystals obtained, called L0, L1, L2 and L3 (depending on the number of ligands bound, see later). The crystals diffracted synchrotron X-rays at resolutions ranging between 1.17 and 2.0 Å, with the best resolution being attained for the L3 crystal. All the crystals had similar monoclinic unit cells (space group P12₁1) that hosted in the asymmetric unit virtually identical CutA homotrimers (Fig. 4 A-C) in which their subunits exhibit (Fig. 4D) the characteristic (Arnesano et al., 2003) ($\beta\alpha\beta$)₂ ferredoxin fold followed by a C-terminal $\beta 5\alpha 3$ extension.

As is typical for CutA proteins (Arnesano et al., 2003; and see also Supplementary Table 1), strands $\beta 2\beta 3$ (the strands that separate the two helices of the ferredoxin fold), are very long, forming a β hairpin (Fig. 4D) that is kinked at the middle, changing its direction by nearly 90° (Fig. 4D,E), with the part after the kink (the distal part) of this hairpin forming one edge of the central hybrid molecular antiparallel β sheets that nucleate the CutA trimer (Fig. 4E). Note that in addition to the already mentioned participation in the sheet of the distal part of the $\beta 2\beta 3$ hairpin, the proximal part of this hairpin from another subunit (the subunit that provides the four central strands of the sheet) also belongs to this sheet (Fig. 4E). In this way, the three sheets become continuous thanks to the $\beta 2\beta 3$ hairpins, since each hairpin participates in two sheets (Fig. 4A). In relation to these sheets, it should also be noted that the only substantial difference between the fold observed in SeCutA and the folds of other CutA proteins of known structure (see data on superimpositions in Supplementary Table 1) is the bending in SeCutA of the N-terminal part of the chain preceding $\beta 1$ (Fig. 4D), so that it interacts antiparallely with $\beta 5$, forming an extra strand (called here $\beta 0$ because it precedes the constant $\beta 1$) in the molecular sheets that nucleate these proteins (Fig. 4E). This extension of the sheet by one strand might increase the stability of this crucial sheet of the trimeric SeCutA structure, thus also possibly increasing trimer stability. The comparison of the $\beta 2\beta 3$ hairpins of SeCutA (and of other CutA proteins) with the corresponding hairpin in canonical PII proteins (Fig. 5A) reveals clearly that the $\beta 2\beta 3$ hairpin, after the kink, is flexible in PII, not forming constantly a hairpin, rather being a large loop (called the T-loop) that is not visualized in many crystal structures of PII because of disorder. The T-loop is plastic, it is typically involved in the interactions of PII with its protein targets, adopting a different conformation with each target, with the propensity of these

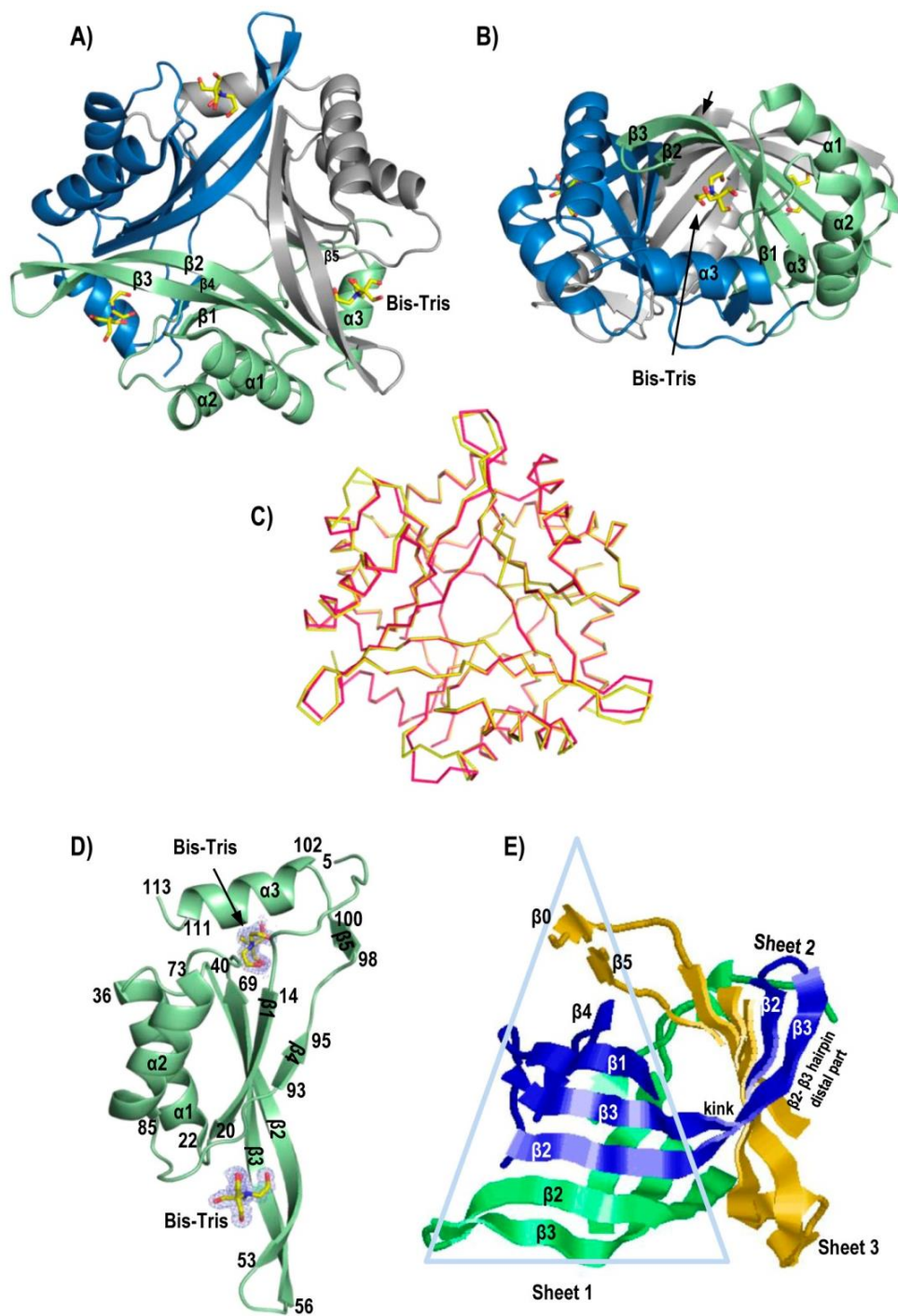


Figure 4 (previous page). Structure of CutA from *S. elongatus*. **A)** and **B)** Cartoon representations of the CutA trimer found in crystal L3, viewed approximately along the threefold axis **(A)** or with this axis vertical **(B)**. Each subunit is colored differently and some secondary structure elements as well as one molecule of the ligand (Bis-Tris; in sticks representation) is labeled. In **C)**, the trimers in the L0 (no ligands, *yellow*) and L3 (three molecules of Bis-Tris bound, *red*) crystals are superimposed in backbone representation. The view is approximately as in **(A)**. **D)** Cartoon representation of one of the three subunits in the L3 crystal. Secondary structure elements (excepting $\beta 0$) are labeled and their starting and ending residues, as well as the initial and final residues of the chain observed in the structure, are given their positions in the amino acid sequence of the protein. The $2Fo-Fc$ map for the two Bis-Tris molecules (labeled and represented in sticks representation) that interact with this subunit are contoured as a grid at $\sigma = 1$. **E)** The β sheets of the trimer to highlight their nuclear nature in forming the trimer. One of the three sheets is enclosed in a blue triangle. The strands of this sheet are labeled.

conformations being strongly influenced by the nucleotide and 2-oxoglutarate, two allosteric effectors that bind over the root of the T-loop (reviewed in Forcada-Nadal et al., 2018). Interestingly, structure-based alignment of PII and CutA sequences (Fig. 5B) shows that the sequences for the $\beta 2\beta 3$ hairpin of CutA or for the corresponding hairpin and T-loop of PII are only marginally different in length (at most 4-residue longer in PII) (Fig. 5B). Thus, it is conceivable that if the distal parts of the $\beta 2\beta 3$ hairpins of CutA were released from the central molecular β sheet, they could also play a signaling role as the T-loop in canonical PII, thus endowing CutA with potential but still non-discovered signaling functions, as suggested by Arnesano et al. (2003). However, another loop of PII, the B-loop, that is also involved (together with the T-loop but much less extensively than it) in the interactions of PII with its partners (Llácer et al., 2007), is virtually non-existent in CutA (Fig. 5A,B). A third difference between canonical PII and CutA proteins is in the C-terminal extension following the end of ferredoxin fold, which in PII is formed by two antiparallel β strands forming the so called mobile C-loop, whereas in CutA it is composed of a β strand and an α helix of 2.5 turns (Fig. 5). As will be seen below, this helix forms the floor

of the intersubunit cavity. In PII this region is important for binding of the allosteric effectors, participating in the accommodation of these ligands, having evolved in plant PII proteins to incorporate after a short second β strand a short helix that is extended with a second 2.5-turn helix for glutamine binding (Chellamuthu et al., 2014).

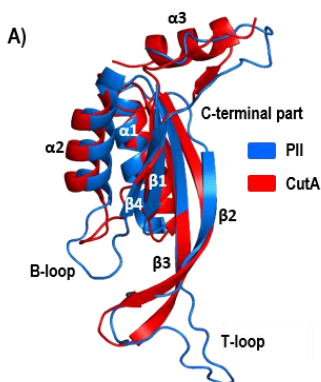
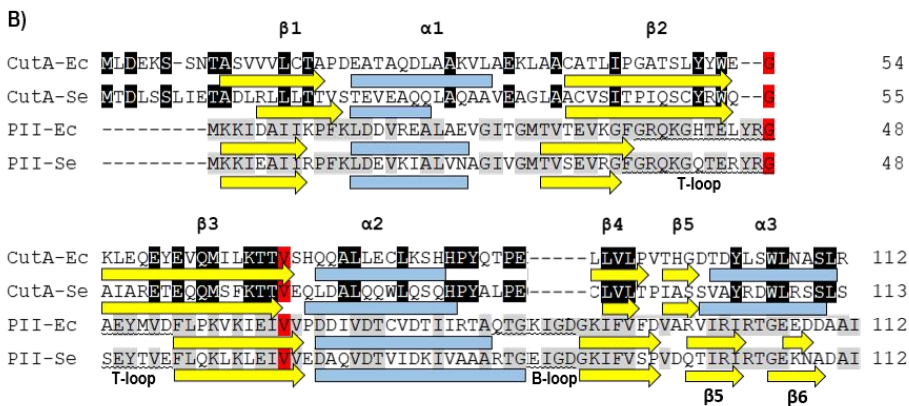


Figure 5. Comparison of the structures of CutA and PII. **(A)** Superimposition of the structure of one subunit of *E. coli* PII protein GlnB (in blue; PDB file 2PII; Carr et al., 1996) and of SeCutA (L3 crystal; in red). α -Helices and β -strands and sites of important conformational differences are labeled. **(B)** Structure-based sequence alignment of *E. coli* and *S. elongatus* CutA and PII. Identities in PII proteins are shadowed black, those in CutA proteins in grey, and those in the four proteins in red. α -Helices, (blue rectangles) and β -strands (yellow arrows) are labeled, as well as the T- and B-loops of PII.



We used the L0 to L3 crystals to carry out Cu^{2+} -soaking experiments to try to ascertain the site of Cu^{2+} interaction with the protein. However, soaking of the crystals with 0.1-1 mM CuSO_4 in the appropriate crystallization solution resulted in very fast crystal destruction. Only when soaking for just a few seconds diffraction studies could be carried out, although with decreased

resolution relative to the non-soaked crystal. However, the structures failed to show any bound Cu^{2+} (not shown). We also carried out co-crystallization attempts adding the maximum possible CuSO_4 concentration that did not cause immediate aggregation. Thus, the maximum CuSO_4 concentration that we could use was 0.44 mM, but crystals were only obtained when the crystallization solution had 0.22 mM CuSO_4 , in fact representing a molar ratio of protein subunits to Cu^{2+} of about 3:1 (L2 crystals, Table 2). Although the resolution was good (1.8 Å), we could not find any electron density suggestive of the presence of bound Cu^{2+} . In prior studies by others, only in one CutA protein structure, from *Pyrococcus horikoshi*, was Cu^{2+} observed (Tanaka et al., 2004a), bound at a twofold symmetry axis of the crystal to the side chain of Asp48, shared by two such residues, one from a different trimer, thus acting as a glue to generate ordered extensive networks of PhCutA molecules in the crystal. Since protein aggregation in solution of *P. horikosi* CutA by the addition of Cu^{2+} was decreased in the Asp48Ala mutant (Tanaka et al., 2004a), the residue corresponding in SeCutA to Asp48 of PhCutA, Glu60, could be a good candidate for a Cu^{2+} binding site. Indeed, this glutamate has its side-chain exposed in our structures, but nevertheless, no density suggesting a Cu^{2+} ion was found bound to it. Another candidate residue for Cu^{2+} binding, Cys41 of SeCutA, an invariant cysteine that sits in the intersubunit pocket and that was proposed to be a potential constant site for Cu^{2+} binding in CutA proteins, since it was found to bind Hg^{2+} (Arnesano et al., 2003), did not bind Cu^{2+} in our L2 crystals, despite the fact that one of the three pockets of the trimer was not occupied with Bis-Tris and Cys41 could have been accessible to Cu^{2+} .

In summary, the *in vivo* experiments carried out here, the poor affinity for Cu^{2+} of SeCutA evidenced by the turbidimetric experiments, and the lack of Cu^{2+} in the SeCutA crystals, taken jointly, militate against the idea that the major

function of CutA in *S. elongatus* (and by extension, in other bacteria) is to prevent Cu²⁺ toxicity, leaving totally open the question of which is the role or roles of this highly widespread protein.

Intersubunit cargo-carrying pockets.

In the SeCutA trimers observed in the different crystals, all of them endowed with nearly perfect threefold symmetry, the three subunits are practically identical (Fig. 6A), as exemplified by the low values of the root mean square deviations (r.m.s.d.) for superimposition of their C α atoms (0.57/0.95/1.28 Å for superimposition of 108/108/109 C α atoms of subunits A-B/A-C/B-C of L3). Nevertheless, there are some conformational differences among subunits in two parts of the protein chain, in particular residues 52-57, in the short β 2- β 3 hairpin tip that could be considered a reminiscence of the T-loop of PII, and at residues 81-89, at the α 2- β 4 connector called in PII the B-loop (Fig. 6A,B), with most of the difference corresponding to Pro86 and Tyr87, which show a displacement of 6.1Å and 4.8Å (Fig. 6B), respectively, between subunits B and C of L3. When residues 52-57 and 81-89 were excluded from the superimposition, r.m.s.d. values for the three subunits of the trimer found in L3 decrease to \leq 0.54Å (93 C α atoms superimposed), illustrating clearly the strong structural stability of each subunit.

It is interesting that these mobile residues correspond to regions of PII that are involved in the interactions with PII targets, although in CutA the part that we call *T-loop reminiscent* is not projected outwards as in PII, being found lining the pocket formed at the junction of two different subunits (Fig. 6C). Since each trimer has three identical intersubunit interfaces, there are three equivalent pockets in the trimer. These pockets were postulated (Laskowski et al., 2003) to be a highly significant structural feature that could provide insight into CutA

function(s). Since in the L3 crystal structure we find in each one of the three pockets a molecule of Bis-Tris (see below), particularly quite close to the T-loop reminiscent, we asked whether the mobility observed could be influenced by the presence of the bound Bis-Tris. Superimposition of the backbone from one subunit of the L3 trimer, which hosts Bis-Tris, with a subunit from the L0 trimer, which has no Bis-Tris, shows structural differences in the same regions (Fig. 6D) and with similar magnitude (not shown) as among the subunits of the L3 trimer. Thus, mobility in these regions appears to be uninfluenced by ligands sitting in the pocket, at least for the Bis-Tris found in our crystal structures.

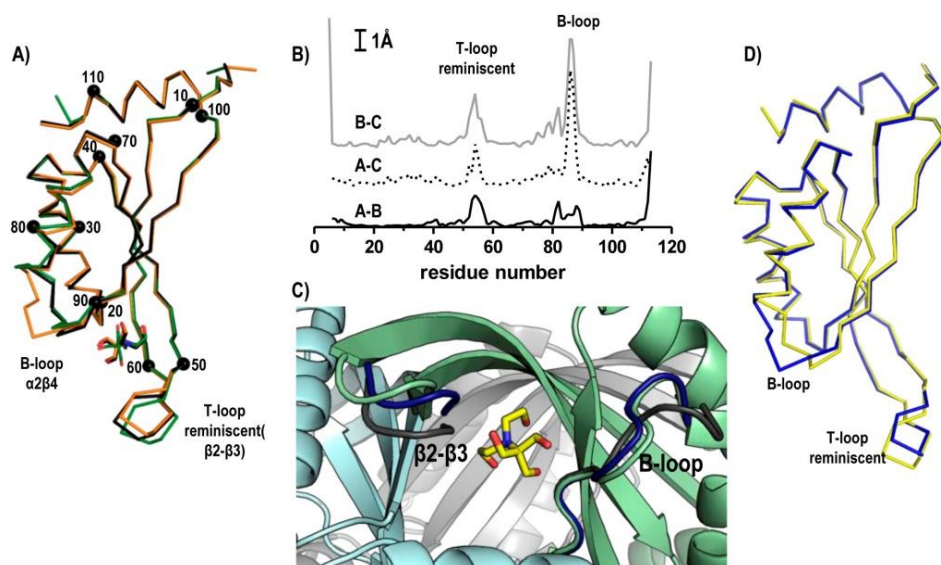
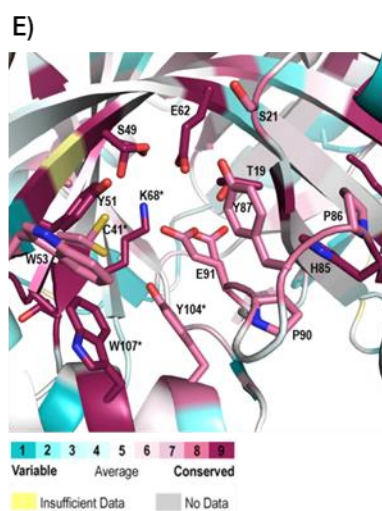
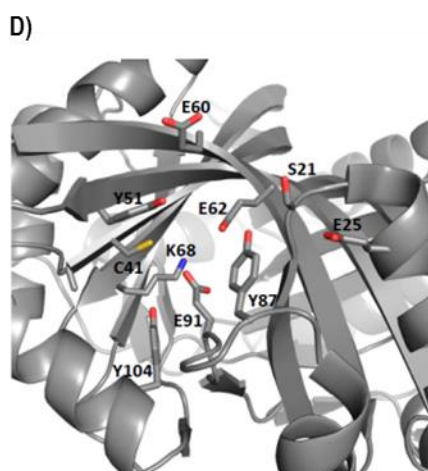
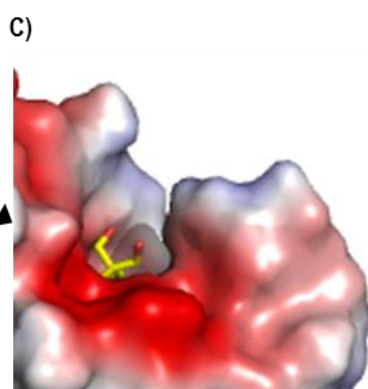
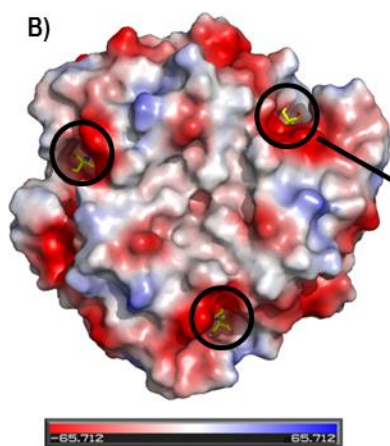
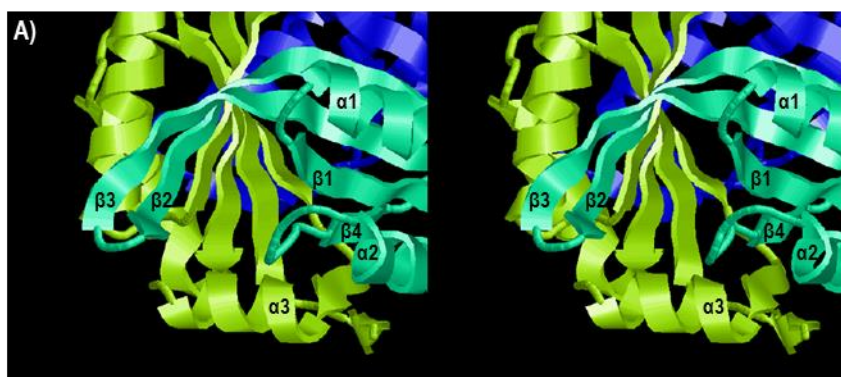


Figure 6. Structural conservation and movement in SeCutA. **A)** Superposition of the backbone of the three CutA subunits of the L3 trimer, with each subunit colored differently. The C α position of each 10 residues is indicated. The Bis-Tris molecules are also shown in sticks representation (O and N atoms red and blue respectively). **B)** Differences in the positions of C α atoms along the sequence for each pair of two subunits. **C)** Cartoon representation of the architecture of the pocket between two subunits (here shown for A and B subunits of the L3 trimer) illustrating the location of a Bis-Tris molecule and superimposing loops β 2- β 3 (T-loop reminiscent) and α 2- β 4 (B-loop) of subunits A (dark blue), B (green) and C (grey). **D)** Same as **(A)**, except that the subunits superimposed are from the L0 (chain B; yellow) and L3 (chain C; blue) trimers.

Each pocket is a crevice in the protein surface corresponding to an internal cavity formed between each pair of adjacent β sheets converging towards the threefold axis of the trimer. This internal cavity, which in the high resolution structure of L3 is found to be mostly filled with water molecules (not shown) is connected to the pocket in the surface by a narrow opening of about 6.5 Å of diameter (not shown). The crevice in the protein surface that we call the pocket is mostly formed (Fig. 7A) between the arched $\beta 2\beta 3$ hairpin of a subunit and the C-terminal helix (helix 3) of the next subunit. In addition, the B-loop ($\alpha 2$ - $\beta 4$ connection) and the $\beta 1$ - $\alpha 1$ connection of the first of these subunits also contribute to delineate this pocket. The pocket is highly polar and negatively charged (Fig. 7B, C) because the opening that connects the pocket with the internal cavity is rimmed by four glutamate residues (Glu25, Glu60, Glu62 and Glu91) and three tyrosines (Tyr51, Tyr87 and Tyr104) (Fig. 7D). Therefore, this pocket cannot accommodate negatively charged compounds like ATP. Somewhat deeper, close to the opening, on the inside face of it, there is one positively charged residue (the only one in the entire pocket and internal cavity), Lys68. Even closer to the opening there is an internal constant cysteine (Cys41) that in the *E. coli* structure was found to interact with the Hg of p-hydroxymercuribenzoate (Arnesano et al., 2003). Many of the residues that form that external pocket are highly conserved in other CutA proteins (Fig. 7E), suggesting important roles for these residues and for the pocket. At the other

Figure 7 (next page). The intersubunit pocket. **A)** Stereo view of the pocket architecture. Each subunit is in a different color. Relevant secondary structure elements are labeled. **B)** Surface representation of the L3 trimer structure, plotting electrostatic potential (negative red and positive blue) calculated with APBS (Baker et al., 2001). The scale gives the range of values for the dimensionless coefficient given by APBS to calculate the surface potential. The three pockets are encircled. **C)** Detail of one of the pockets. **D)** Cartoon representation of a pocket showing polar residues surrounding the opening that connects this pocket to the internal cavity. **E)** Residue conservation according to the indicated color code, estimated by automatic alignment with ConSurf Server.



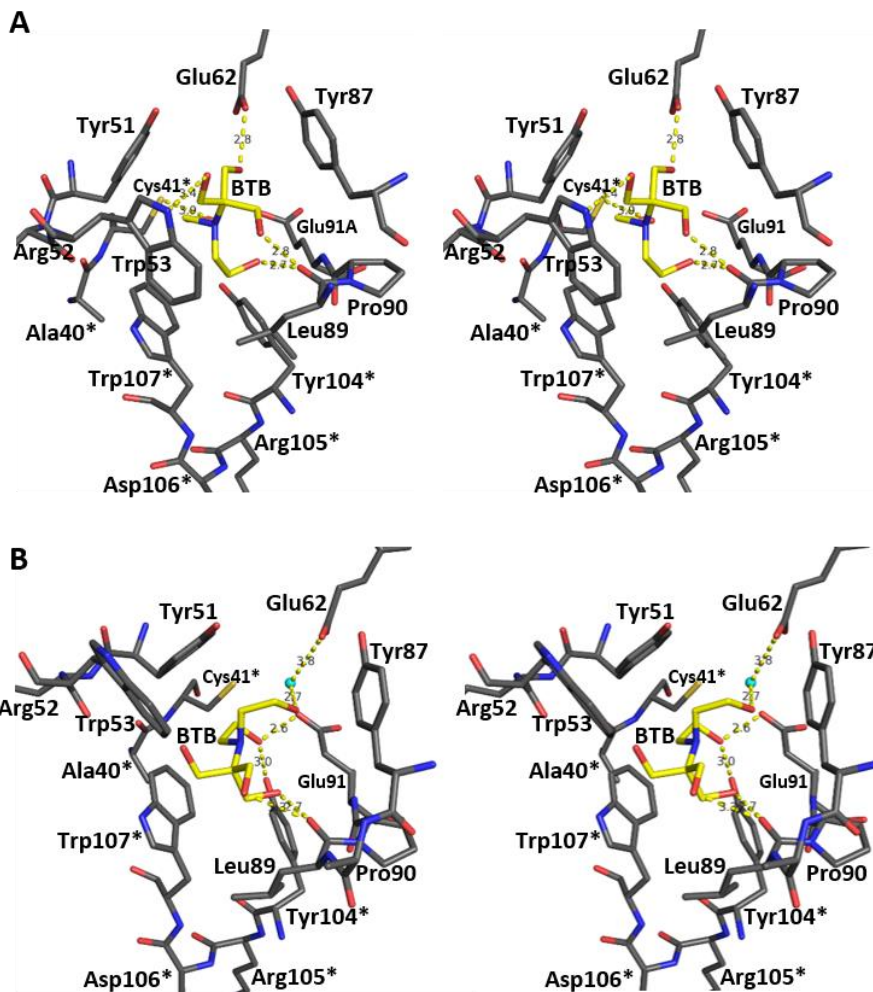


Figure 8. Stereo views of Bis-Tris binding to the pockets formed by subunits C+A (panel A) and subunits A+B (panel B) in the trimer found in the L3 crystal. The broken yellow lines illustrate hydrogen bonds. Asterisks mark residues from subunit C (panel A) or A (panel B). O, N and S atoms are colored red, blue and gold respectively and C atoms are grey. Bis-Tris (abbreviated BTB) is labeled and O, N and C atoms are colored red, blue and yellow, respectively.

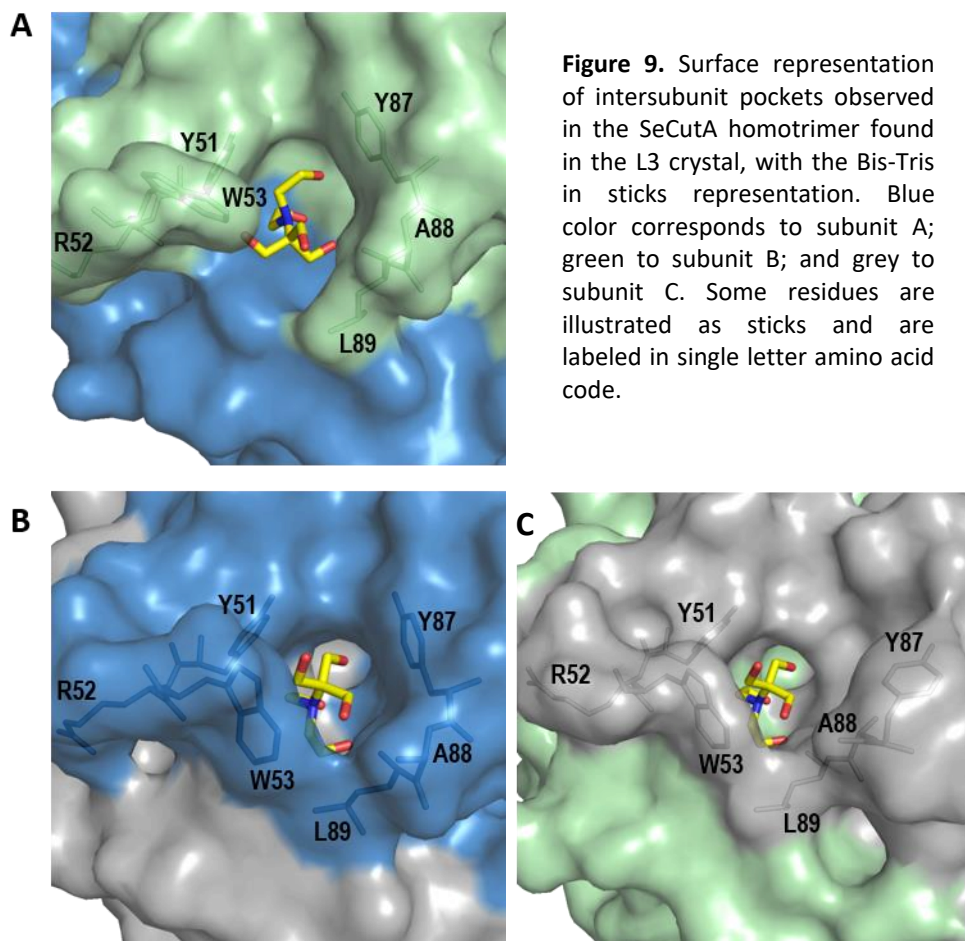
side of the diafragma represented by the opening of the rim, the wide deep cavity on the other side of the opening is largely lined with hydrophobic side-chains, with the N and O atoms of the main chain peptidic bonds of the β strands being the most frequent polar groups (not shown).

As already indicated and shown in Fig. 7B, Bis-Tris derived from the crystallization solution is bound in 3, 2 and 1 pockets of the L3, L2 and L1 trimers, respectively. Bis-Tris is the abbreviation for bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane, a ternary ammonium in which one substituent is an isobutyl trialcohol linked to the N atom by the ternary carbon of the trihydroxyisobutyl group, while the other two substituents are hydroxyethyl groups linked by their C2 atoms to the nitrogen. The pK_a for protonation of the Bis-Tris nitrogen is 6.5. Thus, in the three crystallization solutions in which the L1, L2 and L3 crystals were grown this molecule should be either totally protonated (L1 crystal, pH 5.5) or half of the molecules protonated (L2 and L3 crystals, pH 6.5 in both cases). The combination in the same molecule of an approximately central positive charge and of multiple uncharged but highly polar OH end-groups having ample mobility (given the lack of double bonds that would restrict motions) render Bis-Tris an ideal molecule for interacting with the pocket, which is negatively charged and has many O atoms that could make hydrogen bonds with the multiple hydroxyl groups of the Bis-Tris molecules.

Each Bis-Tris molecule bound to the L3 trimer is encased between the loop at the tip of the $\beta 2$ - $\beta 3$ hairpin that is reminiscent of the T-loop of canonical PII, and the small B-loop of CutA (Fig 6A). Although the Bis Tris is encased in the same part of the pocket in all three pockets of the trimer, it only sits identically in the pockets formed between subunits B and C and between subunits C and A. In these two pockets (Fig. 8A) The $\beta 2$ - $\beta 3$ hairpin loop interacts with the Bis-Tris via the large side-chains of Tyr51 and Trp53, whereas the B-loop provides for interaction the side-chains of Tyr87 and Leu89. Helix 3* of the adjacent subunit (elements from the adjacent subunit are marked by an asterisk) also encloses the Bis-Tris via the side chains of Tyr104* and Trp107*. These six

residues provide large hydrophobic surfaces for van der Waals interactions with carbon atoms of the Bis-Tris molecule. However, the polar hydroxyl groups that tip the five arms of Bis-Tris make hydrogen bonds (Fig. 8A, thin blue lines). These hydrogen bonds are made with the O atom of Leu89, which is the receiver of two bonds, from a hydroxymethyl and a hydroxyethyl group of the ligand; with the δ -carboxylate of Glu62; with the S atom of the invariant Cys41*; and with the indolic N of Trp53.

In contrast, in the pocket formed between the A and the B subunits (Fig. 8B), possibly as a consequence of the movements (Fig. 6A) in the parts of the molecule that line the pocket, the Bis-Tris is rotated approximately around its nitrogen atom and the orientations of its five protruding arms differ from those in the other two pockets. In this way, the enclosure provided by the protein remains the same, but the parts of the Bis-Tris molecule that interact with this enclosure, and particularly the groups making hydrogen bonds, differ from those in the other two pockets. Thus, although the O atom of Leu89 remains the receiver of two hydrogen bonds, these are provided in this case by two hydroxymethyl groups, while the third hydroxymethyl group of Bis-Tris makes no polar interaction. Of the two hydroxyethyl groups, one binds to a water molecule that bridges it to the δ -carboxylate of Glu62, and the other makes two hydrogen bonds with the side chain carboxylate of Glu91 and the OH of Tyr104*. The N atom of Bis-Tris is too deep in the molecule to make any direct contact with the protein in any of the three pockets. However, it is at ion pair distance from Glu62 and Glu91, and therefore it is most likely in the positively charged species. The differences between pockets result in different pocket sizes and shapes when looking at the pockets from the surface, with the differences being particularly patent between the pocket formed between subunits A and B (Fig. 9A) and the other two pockets (Figs. 9B and 9C).



The binding of Bis-Tris to these pockets of CutA reinforces the expectation that a natural biomolecule could bind therein and could play some type of function. The information gathered from the examination of the site and of the mode of Bis-Tris binding suggests that this biomolecule should be either positively charged or neutral and should have polar tips, with many sites capable of making hydrogen bonds or to be able to generate salt bridges. The binding of this molecule perhaps could trigger a signaling function as the function of ATP and 2-oxoglutarate that bind to the corresponding pockets in

PII (Forcada-Nadal et al., 2018). However, since we did not observe any significant change in conformation that could be attributed to Bis-Tris binding, the effects of a bioactive molecule attaching to these pockets cannot be anticipated presently. Of course, given the extensive interactions of bound Bis-Tris with the distal part of the β 2- β 3 hairpin that is fixed in the structure of CutA, the signaling molecule could trigger changes in this distal part of this hairpin that are not triggered by Bis-Tris and that could result in the partial or total release of this distal part of the hairpin, perhaps enabling it to act as the T-loop of PII, triggering still uncharted interactions. This would be plausible, since the structure-based alignment (Fig. 5B) does not predict a mini- or non-existent T-loop in CutA, but rather a T-loop fixed by the trimer structure into a frozen inwards-looking conformation.

A second possibility could stem from the fact that these pockets are really antechambers to larger cavities found within the CutA molecule that appear to be mutually interconnected near the threefold axis as well as connected with a central (although very narrow) opening observed around the threefold axis on one face of the protein (not shown). The cavities inside the molecule could act as storing elements, and the molecules binding to the pocket could, by closing the openings to the inner cavities, or by triggering its opening, prevent or favor entry or exit of the stockpiled compound. This view could have been compatible with a role of CutA in protecting from metals, but now that our results strongly indicate that CutA does not protect bacterial cells from metal toxicity, such a potential role remains at best speculative.

***S. elongatus* CutA shows high thermal stability**

A salient property of CutA that has been described for the protein from different species is its high thermostability, observed in all types of species

from the viewpoint of the temperature of the habitat, from hyperthermophiles to mesophiles and even recorded for a psychrotolerant bacterium (Tanaka et al., 2006; Sawano et al., 2008; Matsuura et al., 2010; Sato et al., 2011; Matsuura et al., 2012). We also tested the thermal stability of *S. elongatus* CutA using thermofluor assays. As for other CutA proteins, the temperature of denaturation as monitored by the increase in fluorescence (Fig. 10) was quite high, with complete unfolding being attained only at about 85 °C. Although this

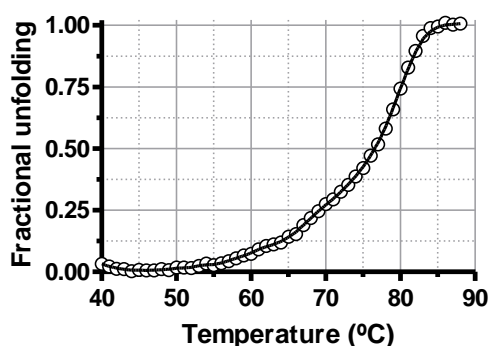


Figure 10: CutA thermal stability. Thermofluor assays of purified CutA from *S. elongatus* (mean \pm SEM, n=2, see materials and methods).

type of assay represents a nonequilibrium technique that is not amenable to rigorous thermodynamic analysis, the curve of increase in fluorescence seems to be a composite of more than one process, perhaps representing first the dissociation of the trimer to monomers and then the unfolding of the monomers. In any case, an operative T_m (temperature at which the increase in the fluorescent signal is one half of the maximum increase attained) of 78°C was estimated (Fig. 10).

These results attest that SeCutA is highly thermostable protein. In this respect SeCutA follows the norm of the CutA proteins of other species, of having a much higher denaturation temperature than the optimal growth temperature (Sawano et al., 2008; Bagautdinov et al., 2015). Charged residues and ion pair formation appear key factors in determining the thermal stability of CutA proteins, as recently proven with *E. coli* CutA (EcCutA) that was made

Table 3. Intra-subunit ion pairs belonging to ion-pair networks in *S. elongatus* CutA. Ion pairs were considered to exist when the distance between the charged groups of different sign did not exceed 6 Å

Donor atoms	Acceptor atoms	Distance (Å)		
		Chain A	Chain B	Chain C
14 ARG NH1	12 ASP OD1	-	-	5.07
14 ARG NH2	12 ASP OD1	-	-	5.24
105 ARG NH1	12 ASP OD1	3.56	3.60	-
105 ARG NH1	12 ASP OD2	2.75	2.72	-
105 ARG NH1	106 ASP OD1	5.92	2.95	2.89
105 ARG NH1	106 ASP OD2	-	4.23	4.14
105 ARG NH2	12 ASP OD1	4.99	4.93	-
105 ARG NH2	12 ASP OD2	4.83	4.77	-
105 ARG NH2	106 ASP OD1	3.92	4.18	4.79
105 ARG NH2	106 ASP OD2	5.92	-	-
109 ARG NH1	12 ASP OD1	4.37	4.54	-
109 ARG NH1	12 ASP OD2	2.84	2.88	-
109 ARG NH1	72 GLU OE1	3.00	2.99	-
109 ARG NH1	72 GLU OE2	3.73	3.72	-
109 ARG NH1	106 ASP OD1	-	-	4.35
109 ARG NH2	12 ASP OD2	4.25	4.37	5.09
109 ARG NH2	72 GLU OE1	3.63	3.50	-
109 ARG NH2	72 GLU OE2	3.44	3.26	-

nearly as thermostable as *P.horikoshi* CutA (150°C denaturation temperature) by increase in the number of charged residues (Matsuura et al., 2018). In this respect, EcCutA and SeCutA have similar denaturation temperatures (90 and 85 °C) and the numbers of charged residues are also similar: EcCutA/SeCutA, D+E 14/12 and K+R+H 11/7. Interestingly, the number of arginine residues (arginine allows the formation of ion pair networks) in EcCutA/SeCutA is 1/6, indicating that a larger number of these networks should be expected in SeCutA than in EcCutA, as corroborated in Table 3, which details residues participating in these networks in SeCutA, whereas the number was restricted to 1 in EcCutA (Tanaka et al., 2006). Thus, these networks might appear to have little influence on the denaturation temperature. However, it is true that EcCutA has more ion pairs

between subunits (17) (Tanaka et al., 2006) than SeCutA (Table 5), so that the combined effects of both factors could result in compensation and in the similarity of the values for the denaturation temperature of EcCutA and SeCutA. This is supported also by the fact that although in the case of *Pyrococcus horikoshi* CutA the number of intrasubunit ion pairs involved in networks is similar to the one in SeCutA (12), the number of intersubunit ion pairs (20) (Tanaka et al., 2006) is much higher than in SeCutA (Table 5). Indeed, *Pyrococcus horikoshi* CutA has a denaturation temperature of 113°C (Sawano et al., 2008), considerably higher than in the cases of EcCutA and SeCutA.

Table 4. Inter-subunit ion pairs in of *S. elongatus* CutA. Ion pairs were considered to exist when the distance between the charged groups of different sign did not exceed 6Å. Calculated using the program Contact (CCP4i), utilizing the structure of the trimer in the L3 crystal.

Chain	Donor atoms	Chain	Acceptor atom	Distance (Å)
A	68 LYS NZ	B	91 GLU OE1	5.55
A	68 LYS NZ	B	91 GLU OE2	4.49
B	68 LYS NZ	C	91 GLU OE1	4.30
B	68 LYS NZ	C	91 GLU OE1	5.31
C	68 LYS NZ	A	91 GLU OE1	4.47
C	68 LYS NZ	A	91 GLU OE2	5.37
C	52 ARG NH1	B	35 GLU OE1	5.78
C	52 ARG NH2	B	35 GLU OE1	5.76
C	59 ARG NH1	B	23 GLU OE1	5.96

PROTEIN DATABANK FILES

Protein DataBank accession codes 6GDU (L0), 6GDV (L1), 6GDW (L2) and 6GDY (L3).

ACKNOWLEDGEMENTS

We thank F.J. Mojica, Universidad de Alicante for the Keio collection *E. coli* strain, Nadine Gougeard (CIBERER-IBV) for technical support and the IBV-CSIC crystallographic facility for crystal growth. This work was supported by grants from the Spanish Government: BFU2014-58229-P and BFU2017-84264-P; and BioStruct-X (grant agreement N°283570, proposal 7687) for synchrotron access. We are grateful to Diamond (Oxfordshire, UK) and Alba (Barcelona, Spain) synchrotrons for access and for staff support. The authors declare no conflict of interest.

REFERENCES

- Armon, A., Graur, D. and Ben-Tal, N. (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* 307, 447–463.
- Arnesano, F., Banci, L., Benvenuti, M., Bertini, I., Calderone, V., Mangani, S. and Viezzoli, M.S. (2003) The evolutionarily conserved trimeric structure of CutA1 proteins suggests a role in signal transduction. *J Biol Chem* 278, 45999–46006.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2, 2006.0008.
- Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., Buchanan, S.G., Buchanan, M.D., Batiyenko, Y., Christopher, J.A., et al. (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* 60, 787–796.

- Bagautdinov, B. (2014) The structures of the CutA1 proteins from *Thermus thermophilus* and *Pyrococcus horikoshii*: characterization of metal-binding sites and metal-induced assembly. *Acta Crystallogr F Struct Biol* 70, 404–413.
- Bagautdinov, B., Matsuura, Y., Bagautdinova, S., Kunishima, N. y Yutani, K. (2008) Structure of putative CutA1 from *Homo sapiens* determined at 2.05 Å resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64, 351-357.
- Bagautdinov, B., Matsuura, Y., Yamamoto, H., Sawano, M., Ogasahara, K., Takehira, M., Kunishima, N., Katoh, E. and Yutani, K. (2015) Thermodynamic analysis of unusually thermostable CutA1 protein from human brain and its protease susceptibility. *J Biochem* 157, 169–176.
- Baker, N.A., Sept, D., Joseph S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98, 10037–10041
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248–254.
- Buchko, G.W., Abendroth, J., Clifton, M.C., Robinson, H., Zhang, Y., Hewitt, S.N., Staker, B.L., Edwards, T.E., Van Voorhis, W.C. y Myler, P.J. (2015) Structure of a CutA1 divalent-cation tolerance protein from *Cryptosporidium parvum*, the protozoal parasite responsible for cryptosporidiosis. *Acta Crystallogr F Struct Biol* 71, 522–530.
- Burkhead, J.L., Abdel-Ghany, S.E., Morrill, J.M., Pilon-Smits, E.A.H. and Pilon, M. (2003) The *Arabidopsis thaliana* CUTA gene encodes an evolutionarily conserved copper binding chloroplast protein. *Plant J* 34, 856–867.

Chellamuthu, V. R., Ermilova, E., Lapina, T., Lüddecke, J., Minaeva, E., Herrmann, C., et al. (2014). A widespread glutamine-sensing mechanism in the plant kingdom. *Cell* 159, 1188–1199.

Clerico, E.M., Ditty, J.L. and Golden, S.S. (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362, 155–171.

Elbing, K. and Brent, R. (2002) Growth in liquid media. *Current Protocols in Molecular Biology* Chapter 1, Unit 1.2.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst Sect. D, Biol Cryst* 66, 486–501.

Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Cryst Sect. D, Biol Cryst* 69, 1204–1214.

Fiedler, S. and Wirth, R. (1988) Transformation of bacteria with plasmid DNA by electroporation. *Anal Biochem* 170, 38–44.

Fong, S.T., Camakaris, J. and Lee, B.T. (1995) Molecular genetics of a chromosomal locus involved in copper tolerance in *Escherichia coli* K-12. *Mol Microbiol* 15, 1127–1137.

Forcada-Nadal, A., Llácer, J.L., Contreras, A., Marco-Marín, C. and Rubio V. (2018) The PII-NAGK-PipX-NtcA regulatory axis of cyanobacteria: a tale of changing partners, allosteric effectors and non-covalent interactions. *Front Mol Biosci.* 5, 91

Forchhammer, K. and Lüddecke, J. (2016) Sensory properties of the PII signalling protein family. *FEBS J* 283, 425–437.

Goldenberg, O., Erez, E., Nimrod, G. and Ben-Tal, N. (2009) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res* 37, D323–7.

Green, M. and Sambrook, J. (2012) Chapter 3: Cloning and transformation with plasmid vectors. In: *Molecular Cloning*. 4th ed. New York: Cold Spring Harbor Laboratory Press.

Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat Rev Genet*. 3, 838–849.

Holtman, C.K., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M.S., Thomas, T.L., Youderian, P. and Golden, S.S. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* 12, 103–115.

Joosten, R.P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A.-C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A.L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I.J. and Vriend, G. (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr*. 42, 376–384.

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

Kinch, L.N. and Grishin, N.V. (2002) Expanding the nitrogen regulatory protein superfamily: Homology detection at below random sequence identity. *Proteins* 48, 75–84.

Labella, J.I., Cantos, R., Espinosa, J., Forcada-Nadal, A., Rubio, V. and Contreras, A. (2017) PipY, a member of the conserved COG0325 family of PLP-binding

proteins, expands the cyanobacterial nitrogen regulatory network. *Front Microbiol* 8, 1244.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.

Laskowski, R.A., Watson, J.D. and Thornton, J.M. (2003) From protein structure to biochemical function? *J Struct Funct Genomics* 4, 167–177.

Llácer, J.L., Contreras, A., Forchhammer, K., Marco-Marín, C., Gil-Ortiz, F., Maldonado, R., Fita, I. and Rubio, V. (2007) The crystal structure of the complex of PII and acetylglutamate kinase reveals how PII controls the storage of nitrogen as arginine. *Proc Natl Acad Sci USA* 104, 17644–17649.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003) Structure validation by C_{alpha} geometry: phi,psi and C_{beta} deviation. *Proteins* 50, 437–450.

Matsuura, Y., Ota, M., Tanaka, T., Takehira, M., Ogasahara, K., Bagautdinov, B., Kunishima, N. and Yutani, K. (2010) Remarkable improvement in the heat stability of CutA1 from *Escherichia coli* by rational protein design. *J Biochem* 148, 449–458.

Matsuura, Y., Takehira, M., Makhatadze, G.I., Joti, Y., Naitow, H., Kunishima, N., Yutani, K. (2018) Strategy for Stabilization of CutA1 Proteins Due to Ion-Ion Interactions at Temperatures of over 100 °C. *Biochemistry*. 57, 2649-2656.

Matsuura, Y., Takehira, M., Sawano, M., Ogasahara, K., Tanaka, T., Yamamoto, H., Kunishima, N., Katoh, E. and Yutani, K. (2012) Role of charged residues in stabilization of *Pyrococcus horikoshii* CutA1, which has a denaturation temperature of nearly 150 °C. *FEBS J* 279, 78–90.

McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Cryst Sect. D, Biol Cryst* 63, 32–41.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4), pp. 658–674.

Metheringham, R., Tyson, K.L., Crooke, H., Missiakas, D., Raina, S. and Cole, J.A. (1996) Effects of mutations in genes for proteins involved in disulphide bond formation in the periplasm on the activities of anaerobically induced electron transfer chains in *Escherichia coli* K12. *Mol Gen Genet* 253, 95–102.

Missiakas, D., Schwager, F. and Raina, S. (1995) Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli*. *EMBO J* 14, 3415–3424.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst Sect. D, Biol Cryst* 67, 355–367.

Perrier, A.L., Cousin, X., Boschetti, N., Haas, R., Chatel, J.M., Bon, S., Roberts, W.L., Pickett, S.R., Massoulié, J., Rosenberry, T.L. and Krejci, E. (2000) Two distinct proteins are associated with tetrameric acetylcholinesterase on the cell surface. *J Biol Chem* 275, 34260–34265.

Rubin, B.E., Wetmore, K.M., Price, M.N., Diamond, S., Shultzaberger, R.K., Lowe, L.C., Curtin, G., Arkin, A.P., Deutschbauer, A. and Golden, S.S. (2015) The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci USA* 112, E6634–43.

Sato, A., Yokotani, S., Tadokoro, T., Tanaka, S., Angkawidjaja, C., Koga, Y., Takano, K. and Kanaya, S. (2011) Crystal structure of stable protein CutA1 from psychrotrophic bacterium *Shewanella* sp. SIB1. *J Synchrotron Radiat* 18, 6–10.

Savchenko, A., Skarina, T., Evdokimova, E., Watson, J.D., Laskowski, R., Arrowsmith, C.H., Edwards, A.M., Joachimiak, A. and Zhang, R. (2004) X-ray crystal structure of CutA from *Thermotoga maritima* at 1.4 Å resolution. *Proteins* 54, 162–165.

Sawano, M., Yamamoto, H., Ogasahara, K., Kidokoro, S., Katoh, S., Ohnuma, T., Katoh, E., Yokoyama, S. and Yutani, K. (2008) Thermodynamic basis for the stabilities of three CutA1s from *Pyrococcus horikoshii*, *Thermus thermophilus*, and *Oryza sativa*, with unusually high denaturation temperatures. *Biochemistry* 47, 721–730.

Selim, K.A., Haase, F., Hartmann, M.D., Hagemann, M. and Forchhammer K. (2018) PII-like signaling protein SbtB links cAMP sensing with cyanobacterial inorganic carbon response. *Proc Natl Acad Sci U S A* 115, E4861–E4869.

Stadtman, E.R. (2001) The story of glutamine synthetase regulation. *J Biol Chem* 276, 44357–44364.

Tanaka, T., Sawano, M., Ogasahara, K., Sakaguchi, Y., Bagautdinov, B., Katoh, E., Kuroishi, C., Shinkai, A., Yokoyama, S. and Yutani, K. (2006) Hyperthermostability of CutA1 protein, with a denaturation temperature of nearly 150 degrees. *FEBS Lett* 580, 4224–4230.

Tanaka, Y., Tsumoto, K., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I. and Kumagai, I. (2004a) Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of *Pyrococcus horikoshii* CutA1. *FEBS Lett* 556, 167–174.

Tanaka, Y., Tsumoto, K., Umetsu, M., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I., Arakawa, T. and Kumagai, I. (2004b) Structural evidence for guanidine-protein side chain interactions: crystal structure of CutA from *Pyrococcus horikoshii* in 3M guanidine hydrochloride. *Biochem Biophys Res Commun* 323, 185–191.

Thomas, G., Coutts, G. and Merrick, M. (2000) The *glnKamtB* operon. A conserved gene pair in prokaryotes. *Trends Genet* 16, 11–14.

Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H., Hui, R., Marsden, B.D., Nordlund, P., Sundstrom, M., Weigelt, J. and Edwards, A.M. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci USA* 103, 15835–15840.

Winter, G., Lobley, C.M.C. and Prince, S.M. (2013) Decision making in Xia2. *Acta Cryst Sect. D, Biol Cryst* 69, 1260–1273.

Xu, Y., Carr, P.D., Clancy, P., Garcia-Dominguez, M., Forchhammer, K., Florencio, F., Vasudevan, S.G., Tandeau de Marsac, N. and Ollis, D.L. (2003) The structures of the PII proteins from the cyanobacteria *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803. *Acta Cryst Sect. D, Biol Cryst* 59, 2183–2190.

SUPPORTING INFORMATION (Capítulo 3)

Is the near-universal CutA protein really a divalent metal cation tolerance factor? Functional and structural studies in bacteria.

Lorena Tremiño¹, Javier Espinosa², Asunción Contreras² and Vicente Rubio^{1*}

¹ Instituto de Biomedicina de Valencia (IBV-CSIC), and CIBER de Enfermedades Raras (CIBERER-ISCIII), Valencia, Spain.

² Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain.

Contains:
Supplementary Table 1

Supplementary Table S1. Structures of CutA proteins reported and/or deposited in the Protein DataBank and superimposition of the trimers of those determined at highest resolution (for each species) with the trimer of SeCutA found in the L3 crystal

Organism	PDB	Resol. (Å)	Ligands and other notes ¹	Ref. ²	Rmsd/atoms (Å/n)
<i>Pyrococcus horikoshii</i>	1UKU*	1.45	Cu ²⁺	1	1.03/300
	1UMJ*	1.60	Guanidinium	2	0.94/300
	4NYO	1.80	Cl ⁻ , MES, Na ⁺ , SO ₄ ⁻ , CSD	3	0.94/298
<i>Archaeoglobus fulgidus</i>	1P1L*	2.00	-	-	1.10/303
<i>Escherichia coli</i>	1NAQ	1.70	p-hydroxymercuribenzoate	4	1.13/302
	3AA9	2.30	E61V mutant	5	1.18/302
<i>Thermotoga maritima</i>	1KR4*	1.40	SeMet	6	0.90/315
	1VHF	1.54	SeMet	7	0.93/300
<i>Thermus thermophilus</i>	1NZA*	1.70	Cl ⁻ , Na ⁺ , SO ₄ ⁻ , GOL	3	1.39/304
<i>Xylella fastidiosa</i>	2NUH*	1.39	Imidazole	-	1.20/305
<i>Yersinia pestis</i>	3GSD	2.05	Na ⁺ , HEPES, EDO, PEG	-	1.06/310
<i>Shewanella sp.</i>	3AHP	2.70	-	8	1.28/300
<i>Salmonella enterica</i>	3OPK	1.90	Na ⁺ , Mg ²⁺ , ACT	-	1.30/307
<i>Ehrlichia chaffeensis</i>	4IYQ	2.55	Ca ²⁺	-	1.21/303
<i>Cryptosporidium parvum</i>	4E98	2.00	Cl ⁻	9	1.27/312
<i>Oryza sativa</i>	2ZOM	3.02	SO ₄ ⁻ , GOL	-	1.21/304
<i>Rattus norvegicus</i>	1OSC	2.15	-	4	1.43/315
<i>Homo sapiens</i>	2ZFH	2.05	-	10	1.38/312

Blue, green and red shadowing: archaea, bacteria and eukaria, respectively.

*Trimer generated by application of the crystal symmetry.

¹MES: 2-(N-morfolino)-ethanesulfonic acid. GOL: glycerol, ACT: acetate, EDO: ethyleneglycol. PEG: polyethylene glycol. CSD: cysteine converted to 3-sulfinoalanina. Muts: mutations.

²1: Tanaka et al., 2004a. 2: Tanaka et al., 2004b. 3: Bagautdinov 2014. 4: Arnesano et al., 2003. 5: Matsuura et al., 2010. 6: Savchenko et al., 2004. 7: Badger et al., 2005. 8: Sato et al., 2011. 9: Buchko et al., 2015. 10: Bagautdinov et al., 2008.

SECCIÓN 3

Discusión general

Sección 3. Discusión general

NOTA: La numeración de las figuras continúa la de la "Introducción general".

A. Proteínas de la familia COG0325: PipY y PLPHP

A.1. Antigüedad y distribución de la familia COG0325.

La conservación del gen *COG0325* en especies tan diversas como cianobacterias y el ser humano debe indicar que su producto proteico desempeña una función que aporta una clara ventaja evolutiva, función no bien caracterizada pero aparentemente universal (Ito et al., 2013; Darin et al., 2013). Su presencia en cianobacterias evidencia su antigüedad, ya que estos procariontes capaces de llevar a cabo la fotosíntesis oxigénica, habitaban ya la Tierra hace unos 2.500 millones de años, cuando fueron organismos clave para la generación de la atmósfera oxigénica terrestre (Giordano et al., 2005). La amplia distribución del gen *COG0325* en todos los dominios de la vida (corroborado por nosotros mediante BLAST para arqueas, datos no presentados) avala la presencia de este gen en LUCA, el último antepasado común universal.

Dada esta antigüedad, quizá no es sorprendente que se corresponda con una proteína con plegamiento básicamente de tipo barril TIM, pues este plegamiento estructural es muy común entre las proteínas depositadas en el PDB (Hegyí y Gerstein, 1999), presente en al menos a 15 funciones enzimáticas diferentes (revisado en Wierenga, 2001) y por tanto heredero de un único plegamiento ancestral común (Copley y Bork, 2000). Tampoco resulta sorprendente que el cofactor de las proteínas de esta familia sea también muy antiguo. La denominación de piridoxina (PN) dada al primer vitámero descrito de la vitamina B₆ obedece a la similitud de su estructura con la pirimidina (Stanton et al., 1939 y Flexner y Chassin, 1941), molécula considerada clave en

“el mundo del RNA” que en el pasado remotísimo precedió al mundo proteico en el curso hacia la vida como existe hoy (Powner y Sutherland, 2011; Stairs et al., 2017 y Roberts et al., 2018). Además, el piridoxal y piridoxal fosfato pudieron producirse prebióticamente (Austin y Waddell, 1999; Aylward y Bofinger, 2006), y se ha sugerido que la síntesis de piridoxal o de piridoxal fosfato pudo ser la reacción más antigua del metabolismo aerobio (Kim et al., 2012). Quizá por ello, y por la enorme versatilidad del PLP como catalizador (Jansenius, 1998), pues se ha implicado en transaminaciones, decarboxilaciones, racemizaciones, escisiones aldólicas, y beta y gamma eliminaciones y reemplazos, el piridoxal fosfato ha tenido tanto éxito evolutivo: se ha distribuido en gran número de enzimas (más de 150) a medida que iban apareciendo, y así las PLP-enzimas pertenecen hoy a todos los grupos troncales de la *Enzyme Commission* (excepto el de las ligasas), representando una fracción importante (sobre el 4 %) de todas las proteínas (Perрудani y Peracchi, 2003 y 2009; <http://bioinformatics.unipr.it/cgi-bin/bioinformatics/B6db/home.pl>). Así, tanto la estructura de las proteínas COG0325, con plegamientos de barril TIM, como el cofactor que utilizan, PLP, respaldan la presencia de estas proteínas a lo largo de toda la escala evolutiva.

Es interesante que el PLP no siempre usa el grupo piridoxal como agente activo en la catálisis, ya que en la glucógeno fosforilasa, el enzima que es el miembro único del grupo V de los enzimas portadores de PLP, no parece ser el grupo piridoxal, sino el grupo fosfato del PLP el que participa en la catálisis de la reacción enzimática (Palm et al., 1990; Stirtan y Withers, 1996), siendo el grupo piridoxal utilizado únicamente para anclar covalentemente el cofactor en el centro activo del enzima. En este sentido las proteínas COG0325 podrían usar también la formación del enlace covalente piridoxal-lisina como forma de fijar el cofactor para su transporte como parte de la función homeostática de PLP

que se propone para las proteínas COG0325 (Ito et al., 2003; Prunetti et al. 2016; Darin et al., 2016; Labella et al., 2017; Plecko et al., 2017; Tremiño et al., 2017 y 2018).

A.2. Principales diferencias entre PipY y PLPHP.

La producción de PipY y PLPHP recombinantes ha permitido caracterizar ambas proteínas. En el caso de PipY, la expresión de una proteína cianobacteriana en un sistema bacteriano (*E. coli*) requiere la realización de un protocolo bastante usual en los laboratorios de biología molecular. Sin embargo, la producción de PLPHP humana en *E. coli* suponía, *a priori*, posibles problemas asociados a la producción de proteínas humanas en sistemas bacterianos (diferentes usos de codones, formación de cuerpos de inclusión, etc), pudiendo requerir para tener éxito el uso de sistemas alternativos, como son los utilizados por otros miembros de nuestro laboratorio basados en baculovirus y células de insecto (Diez-Fernandez et al., 2013; Panza et al., 2016). A pesar de todo ello, ambas proteínas, PipY y PLPHP, pudieron ser obtenidas de forma análoga.

Una diferencia de posible importancia entre PipY y PLPHP se refleja en sus distintos comportamientos en la cromatografía de exclusión molecular, en la que PipY se comporta como un monómero, mientras que PLPHP se comporta como monómero mayoritario con una cierta proporción de forma dimérica. Se ha señalado (Marianayagam et al., 2004) que el estado de oligomerización de las proteínas afecta a las propiedades de las mismas, confiriéndoles con frecuencia ciertas ventajas como un aumento de estabilidad, actividad o regulación de la actividad. Así, sería concebible la posible existencia de regulación de la PLPHP por dimerización que fuera exclusiva de eucariotas multicelulares, no dándose en bacterias, o incluso la existencia de funciones nuevas de PLPHP no desempeñadas por la proteína cianobacteriana, que

podrían requerir del concurso de la dimerización, como se ha postulado en otros contextos (Marianayagam et al., 2004).

Como se ha mencionado en el Capítulo 2 de la Sección 2, dicho aumento de dimerización en la forma humana parece ser dependiente de la mayor riqueza de cisteínas de PLPHP que de PipY, siendo mediada por la formación de puentes disulfuro. Sería deseable conocer el estado oligomérico de PLPHP en la célula, pues es bastante probable que la dimerización observada sea consecuencia de la formación de puentes disulfuro en el entorno oxidativo de fuera de la célula, que es el prevalente para nuestras preparaciones purificadas. Naturalmente, también sería interesante establecer qué grupos sulfhidrilo son responsables de la dimerización, una cuestión que para ser desentrañada requerirá el uso de mutagénesis dirigida y/o de aproximaciones estructurales.

Otras diferencias detectadas entre las dos proteínas producidas *in vitro* es el “envejecimiento” de PipY, asociado a la liberación de PLP en condiciones de almacenamiento a 4°C durante un periodo de tiempo de unos meses, que no se detectó en PLPHP. Este fenómeno determinó que fuese posible estimar la constante de disociación PipY-PLP mediante la incubación de proteína PipY “envejecida” con concentraciones crecientes de PLP. A falta de la estructura de PLPHP desconocemos las razones para esta diferencia, aunque quizá la diferente longitud de las hélices N-terminales y/o cambios en la región C-terminal de anclaje del grupo fosfato puedan ser factores involucrados en la determinación de una mayor o menor facilidad de pérdida de PLP.

A.3. Dificultades para la cristalización de PLPHP: diferencias estructurales entre PipY y PLPHP.

Como se avanzaba en la Sección 1, tras la resolución de la estructura cristalina de PipY se trató de cristalizar la proteína PLPHP, que ha encontrado problemas no resueltos por ahora, por lo que no ha habido más remedio que

recurrir al modelizado de la estructura de PLPHP, habiendo utilizado el servidor I-TASSER (*Iterative Threading Assembly Refinement*) (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>), que dio un plegamiento claro y único como barril TIM modificado muy similar a PipY, pero que dio tanto para la estructura secundaria como para la disposición espacial varias alternativas plausibles para el extremo C-terminal de PLPHP (Fig. 16), aunque en todos ellos se apunta a la posible formación de una hélice adicional $\alpha 9$.

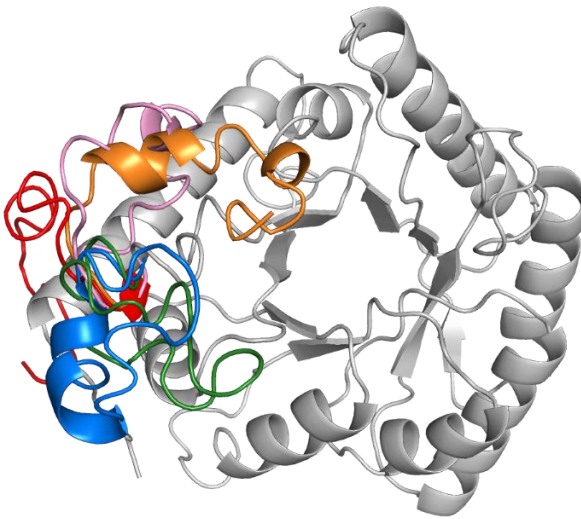


Figura 16. Representación en cintas de los 5 modelos de PLPHP obtenidos con I-TASSER. Las zonas comunes en los 5 modelos se muestran en gris y las zonas variables, correspondientes al extremo C-terminal, en diferentes colores (rojo, azul, verde, naranja o rosa).

La comparación de la estructura cristalográfica de PipY con el modelo de PLPHP revela dos diferencias sustanciales entre ambas proteínas: la longitud de la hélice $\alpha 1$, que presenta menor longitud en PipY de lo que se predice para PLPHP (Fig. 17); y la presencia de las varias predicciones alternativas para el extremo C-terminal de PLPHP (Figs. 16 y 17A). No parecen existir razones por las que las posibles funciones apuntadas aquí para las hélices $\alpha 1$ y $\alpha 9$ de PipY (interacción con otras proteínas y mecanismo de regulación de la fuerza de la unión de PLP, respectivamente, ver Capítulo 1 de la sección 2) no puedan ser aplicadas también a PLPHP, si bien podrían existir diferencias entre PipY y PLPHP justificables en base a la evolución de la proteína desde el grupo de las

cianobacterias hasta el de los humanos y la adaptación de su estructura al reconocimiento de diferentes proteínas o desempeño de nuevas funciones. Así,

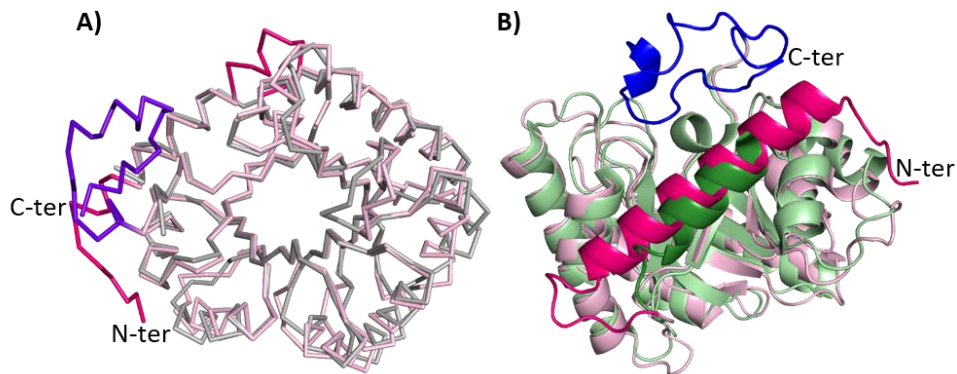


Figura 17. Superposición de la estructura de PipY con el modelo de PLPHP. Los extremos N y C-terminales de la proteína se indican como N-ter y C-ter, respectivamente. Nótese que sólo se ha representado uno de los 5 modelos de PLPHP obtenidos con I-TASSER. A) Representación del esqueleto de PipY (gris) y PLPHP (rosa). Los segmentos de la hélice $\alpha 1$ de PLPHP ausentes en PipY se muestran en magenta y el extremo C-terminal de PLPHP ausente en PipY se muestra en violeta. B) Representación lateral en modelo de cintas de PipY (verde) y PLPHP (rosa). La hélice $\alpha 1$ de ambas proteínas se colorea del mismo color en tono más intenso al del resto de la proteína, y el extremo C-terminal de PLPHP en azul.

se habría evolucionado desde una proteína más sencilla (PipY) a una versión más sofisticada (PLPHP) con una hélice $\alpha 1$ de mayor longitud que podría interactuar con más o diferentes proteínas y un extremo C-terminal más largo con funciones reguladoras y/o también de interacción. Como corolario, las diferencias estructurales entre el modelo estructural de PLPHP y la estructura experimental de PipY sugieren una posible estrategia a seguir que debería explotar el laboratorio para tratar de obtener cristales de PLPHP, consistente en la producción de formas con truncaciones en la porción C-terminal de la proteína, donde se eliminase todo o parte de dicho extremo C-terminal, dado que dicha porción C-terminal podría presentar movilidad que dificulte o impida la cristalización de la proteína.

A.4. Epilepsia dependiente de vitamina B₆ provocada por mutaciones en *PLPBP*.

La reciente identificación de mutaciones en *PLPBP* como causa de un tipo de epilepsia dependiente de vitamina B₆ en tres individuos de una familia consanguínea de origen sirio y en cuatro pacientes adicionales (Darin et al., 2016), junto a la posterior documentación prácticamente inmediata de otros cuatro pacientes (Plecko et al., 2017), sitúa a *PLPBP* entre los genes responsables de esta patología. Hasta la fecha han sido 11 los pacientes identificados y publicados, con un total de 6 mutaciones de cambio de sentido, 2 mutaciones sin sentido que dan lugar a un producto proteico truncado y 2 mutaciones ubicadas en intrones que afectan al procesamiento del transcrito de *PLPBP*. Cabe esperar que durante los próximos años se describan nuevas mutaciones en el gen *PLPBP* implicadas en esta epilepsia dependiente de vitamina B₆. De hecho, la base de datos ExAC (<http://exac.broadinstitute.org/gene/ENSG00000147471>) recoge 294 variantes del gen ya registradas, 81 de las cuales son de cambio de sentido, es decir, dan lugar a una variante de la proteína con la secuencia de aminoácidos alterada en una posición. Es posible que muchas de estas mutaciones sean inocuas, incluso si se presentaran en homocigosis o en heterocigosis compuesta, por la naturaleza y/o posición en la estructura de la proteína del aminoácido cambiado (Creighton, 1993). En principio, se puede predecir una mayor repercusión sobre la estructura y función de la proteína en los casos en que las mutaciones afecten a aminoácidos situados en torno al sitio de unión del cofactor, PLP en este caso, o en zonas con una fuerte implicación en el mantenimiento de la estructura de la proteína, o incluso en zonas de interacción con otras proteínas (si las hubiera). En todo caso, más aún dadas las facilidades actuales para la realización de análisis genético y secuenciación

completa de genomas, lo esperable es que pronto se describan nuevas mutaciones en el gen *PLPBP* asociadas a epilepsia dependiente de vitamina B₆.

Un aspecto llamativo relativo al tratamiento de la epilepsia asociada a mutaciones en *PLPBP* reside en el vitámero específico que es efectivo. Todos los pacientes descritos hasta la fecha como portadores de mutaciones de cambio de sentido fueron tratados con la forma piridoxina (PN), excepto uno de ellos que tuvo que ser tratado con PLP al no responder de forma adecuada a la PN (para más detalle, ver Tabla 1 del artículo presentado en el Capítulo 2 de la Sección 2; y además, Darin et al., 2016; Plecko et al., 2017). Este paciente es portador de la mutación Leu175Pro en homocigosis. Nótese que dicha forma de *PLPBP* no ha podido ser producida *in vitro* (Capítulo 2 de la Sección 2) lo que indica que la mutación afecta gravemente al plegamiento de la proteína. La ausencia total de la proteína *PLPBP* en el paciente requiere de un aporte de la forma PLP, mientras que el resto de los pacientes que expresan la proteína, aunque alterada, son capaces de recuperarse mediante el suministro de PN. Esta diferencia en el tratamiento podría ser indicativa de la severidad de la mutación a nivel proteico, de modo que se esperaría que pacientes con *PLPBP* semi-funcional podrían ser tratados con PN, mientras que la ausencia de actividad de *PLPBP* requiriese el aporte directo del vitámero PLP. El mecanismo molecular que explique la necesidad de uno u otro vitámero para tratar la patología todavía no está claro, pues, aunque se ha propuesto un papel en la homeostasis del PLP (Darin et al., 2016, Labella et al., 2017), todavía no se ha establecido la función de la proteína *PLPBP* a nivel molecular.

B. CutA, miembro no canónico de la superfamilia PII

B.1. Anotación incorrecta: CutA no parece implicada en tolerancia a cobre

El contexto en el que se identificó el gen codificante de CutA en *E. coli* (Rouch et al., 1989; Fong et al., 1995) llevó a atribuir a *cutA1* funciones relacionadas con la tolerancia a metales. Actualmente el análisis de la secuencia de los genes *cutA1* y *cutA2* en *E. coli* revela que existe un solapamiento de los últimos 25 pares de bases de *cutA1* y los primeros 25 pares de bases de *cutA2* (Fig. 18), si bien tal solapamiento no se detectó en el trabajo original de Fong et al. (1995). Pues estos autores situaron el codón de inicio de *cutA2* en el sitio que hoy sabemos corresponde con la segunda metionina de la secuencia de *cutA2* (Fig. 18).

CGAGGTGTTT**ATG**C**TTGATGAAAAAGTTCGAATACCGCGTCTGTGCGTGGTGC**TATGTACGGCACCAGATGAAG
CGACAGCCCAGGATT**TAGCCGCCAAAGTGC**TGGCGGAAAAACTGGCGGCC**TGCGCAGCTTGATCCCCGGCGCT**
ACCTCTCTCTATTACTGGGAAGGTAAGCTGGAGCAAGAATACGAAGTGCAGATGATTTTTAAAACTACCGTATC
TCACCAGCAGGCACTGCTGGAATGCCTGAAGTCTCATCATCCATATCAAACCCCGGA**ACTTCTGGTTTTACCTG**
TTACACCGGAGACAGATTACCTCTCATGGCTCAACGCAT**CTTTACGC**FGA**TCCTGCTACTTTGCAGCACTT**
CCGTTTTTGGCCGATTATTCGACGCGCCGGGACGTT**CACAATTTG**CCCCGCGGATCAAGCCTTTGCTTTTGGAT
TTTCAGCAAAACCAACATGACCTTAATCTGACCTGGCAGATCAAAGACGGTTACTACCTTACCGTAAACAGAT
CCGCATTACGCCGGAACACCGGAAAATGGCGAC**GTG**CAGCTGCCCAAGGCGTCTGGCATGAAGATGAGTTTT
ACGGCAAAGCGAGATTTACCGCGATCGGCTGACGCTTCCCGTCA**CCATCA**ACCAGGCGAGTGGCGGAGCGACG
TTA**ACTGTC**ACTACCAGGGCTGTGCTGATGCCGGTTCTGTTATCCGCCAGAAACCAAAACCGTTCCGTTAAG
CGAAGTGGTGGCCAAACACGACGCGCCACAGCCTGTGCTGTTCGCCAGCAAGAGCAGCCACCGGCAATTGC
CCTTTTCCGCGCTCTGGGCGTGTGGATCGGTATTGGTATCGCCTTTACGCCATCGGTGCTGCCAATGTACCCA
CTGATTTCTGGCATCGTCTGGGTGGTAAACAGCGGCTCTCCACTGCCAGAGCATTGTTGCTGACCTTTATTTA
TGTGCAGGGGATGGCGCTGACTACACGGCGCTGGGTCTGGTGGTTGCCGCCGAGGGTTACAGTTCCAGGCGG
CGTACAGCACCCATACGCTCATTTGGCCTCGCCATCGTCTTACCTTGCTGGCGATGCAATGTTGGCTTG
TTTACCTGCAACTCCCTCTTCGGTGAACACAGTCTCACGTTGATGAGCAATCGCCAACAGCGCGGCTCAC
TGGCGGTGTGTTTGGTATGGGGCGATTGCCGGACTGATCTGTTCCACATGCACCACCGCACCCTTAGCGCGA
TTCTGCTGATATCGCCAAAGCGGGAACATGTGGCTGGGCGCGGCACGCTTTATCTCTATGCGTTGGGCATG
GGCCTGCCGCTGATGCAATACCGTCTTTGGTAACCGCTTGCTGCCGAAAGCGGCCCGTGGATGGAACAAGT
CAAAACCGCGTTGGTTTTGTGATCCTCGACTGCCGGTCTTCCTGCTGGAGCGAGTGA**TTGGT**GATG**TATGGG**
GATTACGCTTGGTGGCGCTGGGTGTCGATCTTTGGCTGGGCTTTATCACCAGCCTACAGGCTAAACGC
GGCTGGATGCGTATTGTGCAATATCTGCTGGCAGCGGCA**TTGGT**TAGCGTGCGCCACTTACAGGATTGGGC
ATTTGGTGCGACGCATACCGCGAAACTCAGACGCATCTCACTTTACACAAATCAAACCGGTAGATGAGTTAA
ATCAGGCGCTCGTTGAAGCCAAAGGCAAAACCGGTGATGTTAGATCTTTATGCCGACTGGTGCCTGCCTGTAAA
GAGTTTGAGAAATACACCTTACGCGACCCGAGGTGCAAAAAGCGTTAGCAGACACGGTCTTACTTCAGGCCAA
CGTACGCGCCAAACGACGCACAAGATGTGGCGCTGTAAAGCATCTTAATGTCCTTGGCCTACCGACAATCTCT
TTTTTGACGGACAAGGCCAGGAGCATCCACAAGCAGCGCTCACGGGCTTTATGGATGCTGAAACCTTACAGCGCA
CATTTGCGCGATCGCAACCG**FGA**ACGACACTTGCAGTGGGAAAGACGGA

Figura 18. Fragmento de secuencia génica de *E.coli* K12 que incluye los genes *cutA1* (negrita) y *cutA2* (subrayado). Los sombreados verde y azul marcan los codones de inicio anotados actualmente para, respectivamente, los genes *cutA1* y *cutA2* (secuencias de los archivos correspondientes de UniProt KB, , P69488 y P36655 para CUTA y DSBD, respectivamente, ambos de *E. coli*). El sombreado amarillo señala el codón de inicio de *cutA2* propuesto por Fong et al. en 1995.

Nuestros experimentos de toxicidad por metales en *E. coli* y *S. elongatus*, presentados en el Capítulo 3 de la Sección 2 de esta tesis no han apoyado una función de CutA en la resistencia a metales, al menos en bacterias. Además, la ausencia de cobre en los cristales tampoco respalda esta función. Por ello creemos necesario revisar las anotaciones del gen *cutA* como de su producto proteico en las bases de datos, puesto que actualmente no existen datos experimentales robustos que demuestren que se trate de una “proteína de tolerancia a cationes-divalentes”. Quizás, sería más conveniente que CutA se reconociera como “proteína de función desconocida, miembro no canónico de la superfamilia PII” hasta que se identifique su papel fisiológico y su función en los seres vivos.

B.2. La proteína CutA une cobre

Aunque nuestros experimentos indican que el gen *cutA* de *E. coli* y *S. elongatus*, y por extensión, de otras bacterias, no está implicado en la tolerancia a metales, las evidencias revisadas o proporcionadas aquí en la Introducción y en el Capítulo 3 de la Sección 2 están de acuerdo en apuntar a que la proteína CutA de diversas especies sí es capaz de unir cobre. Para tratar de obtener mayor información sobre la unión de cobre a CutA se llevó a cabo la cristalización de la proteína en presencia del metal, pero no fue posible detectar cobre en ninguno de los cristales resueltos. De hecho, la revisión bibliográfica permitió detectar múltiples intentos de cristalización de proteínas CutA con cobre unido (usando técnicas de co-cristalización y/o de impregnación de los cristales en soluciones con cobre) (Arnesano et al., 2003; Tanaka et al., 2004a; Bagautdinov 2014), destacando como único caso de éxito de evidencia de unión de cobre a nivel estructural el de la proteína CutA de *P. horikoshii*, en

la que cada cobre interacciona con el residuo de Asp48 de dos trímeros vecinos, al coincidir con un eje cristalográfico binario (Tanaka et al., 2004a) (Fig. 19).

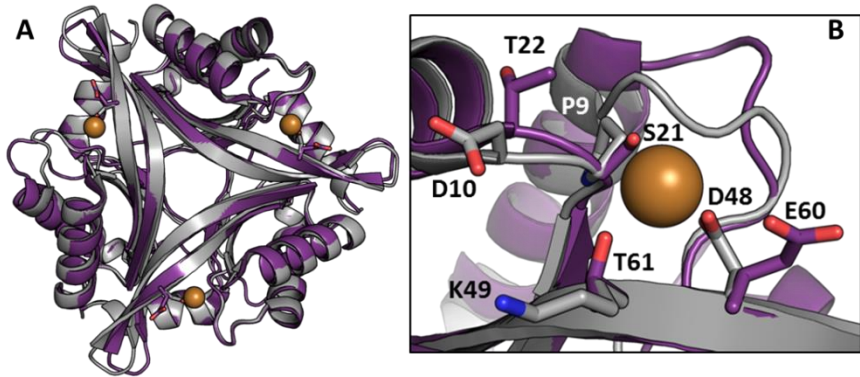


Figura 19. Sitio de unión a cobre en un trímero de CutA de *P. horikoshii*. Como el Cu^{2+} se une en el eje de simetría binaria, solo se está representando, por claridad, la mitad del sitio. A) Superposición global y B) detalle para una subunidad, de las proteínas de *S. elongatus* (en morado) y *P. horikoshii* (PDB:1UKU, en gris) con cobre unido (esfera naranja). En modelo de varillas se representa el residuo D48 de *P. horikoshi* que une cobre y el equivalente (E60) en *S. elongatus*, así como, en el detalle (B), de los principales residuos de la zona, nombrados con el código de aminoácidos de una letra.

La comparación de proteínas CutA de distintos organismos revela conservación de aspartato o glutamato en la posición correspondiente al Asp48 de *P. horikoshii* (Fig. 20), lo que podrían sugerir que el sitio de unión a cobre podría estar conservado en CutA de *S. elongatus* y de muchos otros organismos (Tanaka et al., 2004a), pero dicha hipótesis pierde fuerza al tratarse del caso especial de que el metal sea compartido por dos trímeros vecinos, algo que ni siquiera se da siempre en el caso de CutA de *P. horikoshii*, ya que se han obtenido también cristales del mismo sin Cu^{2+} unido a pesar de agregarlo a la solución de cristalización (Bagautdinov, 2014). Quizá la naturaleza y disposición espacial de los residuos de la zona correspondiente en otros organismos pueda dificultar o impedir la unión del metal (Fig. 19).

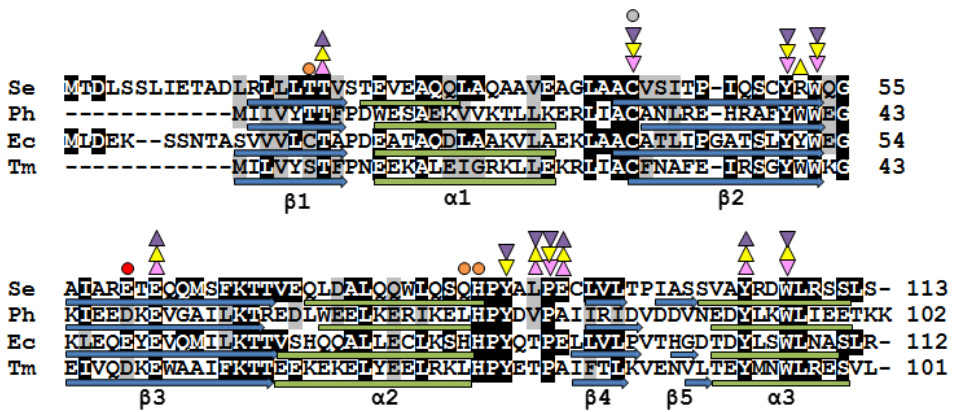


Figura 20: Alineamiento basado en estructura de las secuencias de CutA de *S. elongatus* (Se) *P. horikoshii* (Ph), *E. coli* (Ec) y *T. maritima* (Tm). Sus identificadores respectivos del PDB son 1UKU, 1NAQ y 1KR4. Usando la secuencia de CutA de *S. elongatus* como referencia, los residuos conservados con identidad aparecen con letra blanca sobre fondo negro y los residuos reemplazados de forma conservadora se indican con letra negra sobre fondo gris. Las flechas horizontales (en azul) y los rectángulos (en verde) debajo de cada secuencia indican las posiciones en la secuencia de las hebras β y las hélices α , respectivamente. Al pie del alineamiento se identifican nominalmente los elementos de estructura secundaria. En la parte superior, los triángulos hacia arriba y hacia abajo señalan los residuos que contactan con el Bis-Tris en la estructura de CutA de *S. elongatus* mediante puentes de hidrógeno o interacciones hidrofóbicas, respectivamente; los colores morado, amarillo y rosa indican si dicho contacto se da en la subunidad A, B o C respectivamente. En la parte superior, los círculos señalan los residuos que interaccionan con el cobre en CutA de *P. horikoshii* (coloreado en rojo) o han sido sugeridos como sitios implicados en la unión de cobre en CutA de *E. coli* (coloreado en naranja) y *T. maritima* (coloreado en gris).

B.3. Los bolsillos intersubunidades como elementos con potencialidad funcional.

Ya hemos comentado que la estructura de CutA de *S. elongatus* (Fig. 21A) ha revelado en cada subunidad el núcleo ferredoxina clásico de la familia de proteínas PII, nucleado por el motivo $\beta\alpha\beta\beta\alpha\beta$ y la formación de una gran hoja β

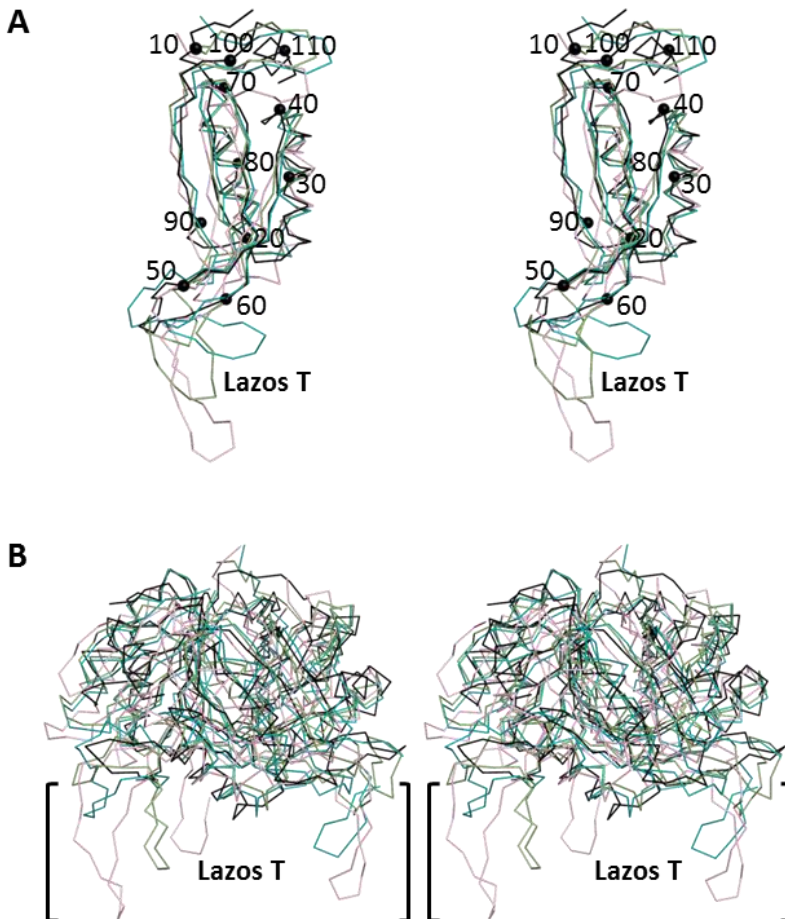


Figura 21. Representación estereoscópica de las estructuras de CutA y PII superpuestas, representadas como cadena principal de subunidad (A) y trímero (B). Se representa: en negro la estructura de CutA de *S. elongatus* donde se indica la posición de algunos residuos con esferas y el número de residuo; en verde la estructura de PII de *E. coli* (PDB: 2PII); y en azul y rosa dos formas distintas de PII de *S. elongatus* (PDB: 2V5H y PDB: 2XG8). Nótese que CutA carece de los móviles lazos T de PII.

antiparalela, seguido de una extensión C-terminal que en CutA se organiza como una hebra β_5 y una hélice α_3 , forma diferente a las de PII canónica, presentando además como nota distintiva de CutA de *S. elongatus* una hebra β_0 al comienzo del polipéptido.

Nuestras estructuras han confirmado el ensamblaje de 3 subunidades en una estructura homotrimérica con hojas beta mixtas que crean un continuo beta en toda la extensión del homotrímero (Fig. 21B), lo que sin duda es clave en conferir a CutA su importantísima termoestabilidad. Igualmente han revelado la formación de bolsillos accesibles desde el solvente, situados en las superficies de interacción entre cada dos subunidades (Fig. 22). La

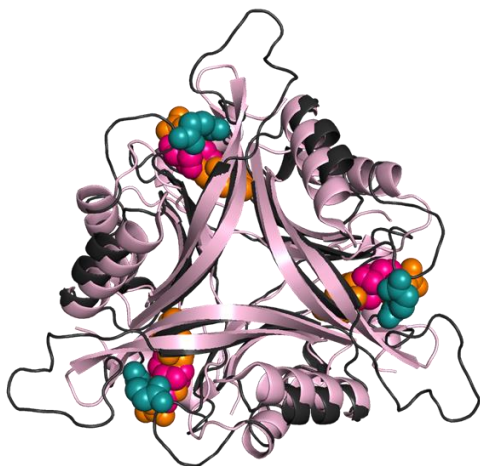


Figura 22. Superposición de CutA (rosa) y PII (PDB:2XZ, negro) de *S. elongatus* representadas en modelo de cintas. Las moléculas presentes en las cavidades equivalentes formadas entre las subunidades que conforman los trímeros de CutA y PII se representan en modelo de esferas: Bis-Tris (magenta), ATP (naranja) y 2-OG (verde).

identificación de Bis-Tris unido de forma inespecífica y ligeramente variable entre las cavidades de CutA (Fig. 20), las cuales pueden considerarse equivalentes a los sitios canónicos de unión de efectores en proteínas PII (Fig. 22), refuerza la posibilidad de que la unión específica de determinada molécula en la cavidad podría modular la actividad de CutA, al igual que ocurre con las proteínas PII canónicas.

Basándose en la similitud estructural con la proteína PII y en la presencia de cavidades equivalentes formadas entre las subunidades del trímero (Fig. 22), se propuso para CutA un papel en transducción de señales (Arnesano et al., 2003), aunque parece militar contra esta propuesta la ausencia en CutA de los característicos lazos T de PII (Figs. 21 y 22), donde se ubican los residuos modificables postraduccionalmente que regulan las funciones de PII permitiendo o no la interacción con sus proteínas diana, así como el pequeño

desarrollo en CutA del lazo B, también involucrado en PII en sus interacciones proteína-proteína. Hay que decir que los bolsillos de CutA en los que hemos identificado moléculas de Bis-Tris (Fig. 23A), no sería capaz de albergar ATP ni

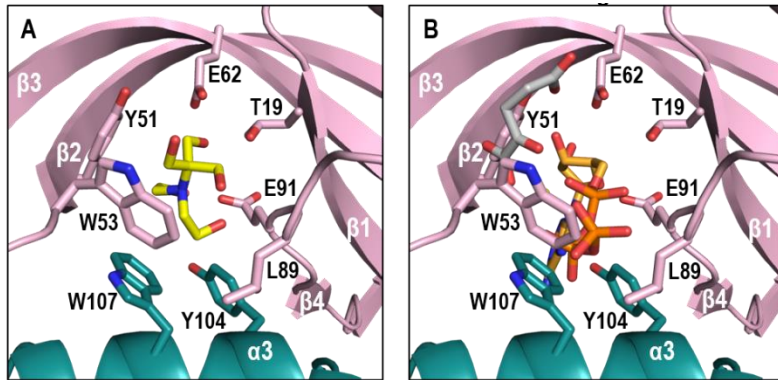


Figura 23. Las cavidades de CutA albergan Bis-Tris (A), pero no son adecuadas para albergar ATP y 2-OG (B). Representación de una de las cavidades de CutA. Se dan colores diferentes a las dos subunidades que la forman. Se han representado cadenas laterales de residuos (código de una letra) que limitan la cavidad. El Bis-Tris (A) ATP y 2-OG (B) se representan con sus átomos de carbono en amarillo, naranja y gris, respectivamente (N y O en azul y rojo, respectivamente).

2-OG por impedimentos de carga, así como por impedimentos estéricos, particularmente con los residuos Y51 y W53 (Fig. 23B).

El alineamiento estructural de la Fig. 20 revela que la mayoría de los residuos de CutA de *S. elongatus* implicados en la unión de Bis-Tris están conservados, al menos en secuencias de CutA de bacterias y arqueas. Un alineamiento de proteínas CutA de un número más alto de especies, incluyendo animales y vegetales (Fig. 24A), muestra que la mayoría de residuos invariantes se ubican principalmente en las cavidades formadas entre los homotrimeros de CutA (Fig. 24B), apoyando la importancia de estas cavidades (Laskowski et al., 2003), quizá en la unión de bio-ligandos naturales no identificados todavía, y que, como se ha comentado en el capítulo 3 de la Sección 2, podrían quizá regular la función de la proteína. Probablemente, el carácter hidrofóbico con

A)

```

PH  LPPDWESAELVVKTLLEKRLIACANLRE-HRAFYVWEGKLEEDKVEGAILKREDLWEEFKERIKELFFDVPFAITRIDVDDVNEDLLKRL
TM  LPPNEKALELGRKLEKRLIACFNAFE-IRSGYVWEGETVQDRFWAAIRKTEEEKKELYEIRKLEHVEYTFAITIKVENVLTGMNKL
HS  LCPNEKVAKEELARAVVEKRLAACVNLIFQITSIYEWGKLEEDSEVLLMKIQSSLVPATDFVRSVHPEVEVAEVIALPVEQGNFPLQV
RN  LCPNEKVAKEELARAVVEKRLAACVNLIFQITSIYEWGKLEEDSEVLLMKIQSSLVPATDFVRSVHPEVEVAEVIALPVEQGNFPLQV
MM  LCPNEKVAKEELARAVVEKRLAACVNLIFQITSIYEWGKLEEDSEVLLMKIQSSLVPATDFVRSVHPEVEVAEVIALPVEQGNFPLQV
RG  LTPSLQVAKEELASSLEARLIACANLVEQVSMYRWGKWNIDEVLLMKIQTSLFPSTQAVLRSKHPVKVEIISNVQDSLPAARDV
SE  LVPSTVEVAQQLAQAAVEAGLAACVSIYTF-IQSCYRWGKALARETEQQMSFKITVEQQLDAQVWQSQHPEVALPECLVITPFIASSVANRDV
PA  LAPDDACAQRILANSLLEKRLAACVNLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
YP  LAPDEASAQNLAQVLEKRLAACVNLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
EC  LAPDEATAQDLAAKVLAEKLAACATLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
SD  LAPDEATAQDLAAKVLAEKLAACATLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
ST  LAPDEATAQDLAAKVLAEKLAACATLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
CK  LAPDEATAQDLAAKVLAEKLAACATLIFGATSLYVWEGKLEQYEVQMLIKQTVSHQQALDLECLKSHHPEYQTPPELLVLPVTHGDDLLSNL
AT  LVPNREAGKILANSIVQEKLAACVNIIVEGIESVYVWEGKIQSDSEELLIKRQSLLEPTEHVNANHEBDVPEVIALPITGGSDKLLKEL
ZM  LVPNKEAGKILAGSILSEKLAACVNIIVEGIESVYVWEGKIQSDSEELLIKRQSLLEPTEHVNANHEBDVPEVIALPITGGSDKLLKEL
    
```

B)

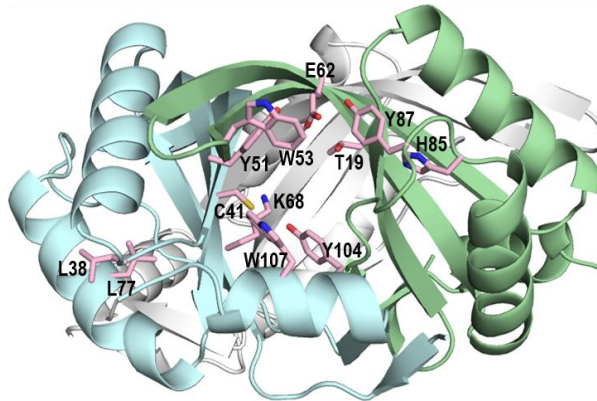


Figura 24. Conservación de residuos en la secuencia de CutA. A) Alineamiento de la región central de las secuencias de CutA donde se señalan los residuos idénticos (letra blanca sobre fondo negro) y conservados (letra negra sobre fondo gris) únicamente cuando la conservación se da en todas las secuencias alineadas. El organismo al que pertenece cada secuencia de CutA se indica en código de dos letras: PH, *P. horikoshii* (arquea); TM, *Thermotoga maritima* (bacteria); HS, *Homo sapiens* (humano); RN, *Rattus norvegicus* (rata); MM, *Mus musculus* (ratón); RG, *Rhizoclostratium globosum* (hongo); SE, *S. elongatus* (cianobacteria); PA, *Pectobacterium atrosepticum* (bacteria); YP, *Yersinia pestis* (bacteria); EC, *E.coli* (bacteria); SD, *Shigella dysenteriae* (bacteria); ST, *Salmonella typhimurium* (bacteria); CK, *Citrobacter koseri* (bacteria); AT, *Arabidopsis thaliana* (vegetal) y ZM, *Zea mays* (arroz). B) Representación en modelo de cintas de CutA donde cada una de las cadenas se colorea de un color (azul, verde y gris) y las cadenas laterales de los residuos conservados se representan en modelo de barras, donde C, N y O se colorean de rosa, azul y rojo, respectivamente.

elementos polares (N en el grupo indol del triptófano, grupo OH fenólico en las tirosinas) de la mayoría de estos residuos estén relacionados con la naturaleza

del ligando fisiológico, cuya unión podría inducir el movimiento de los lazos β_2 - β_3 y α_2 - β_4 , un efecto no observado con la unión del artificial Bis-Tris.

B.4. Consideraciones adicionales sobre el papel de la estructura de CutA en la elevada estabilidad térmica de la proteína

La elevada estabilidad térmica de CutA es un fenómeno bien registrado en la bibliografía (Tanaka et al., 2006; Sawano et al., 2008; Sato et al., 2011; Matsuura et al., 2012; 2015; Bagautdinov et al., 2015) que ha sido tratado en la Introducción general y en el Capítulo 3 de la Sección 2 de esta tesis doctoral, y que nos ha llevado a la caracterización de la estabilidad térmica de SeCutA en ensayos de termofluor, revelando una elevada termoestabilidad que hemos discutido en base a la existencia de redes de pares iónicos intra e intersubunidad como posibles elementos estabilizadores. Aquí resaltaremos que las interacciones hidrofóbicas también son importantes determinantes de termoestabilidad (revisado en Matsuura et al., 2018) y que, en este último ámbito, vale la pena destacar que la Leu38 y Leu77 (numeración de SeCutA) son invariantes y no forman parte de los bolsillos intersubunidades, localizándose al comienzo del lazo $\alpha_1\beta_2$ y al comienzo de la hélice α_2 , respectivamente (Figs. 20 y 24). Su mutua contigüidad espacial (Fig. 24), formando contactos de van der Waals entre ellos, y su ubicación en la estructura de la proteína en un núcleo hidrofóbico (no mostrado, pero su ámbito espacial se señala con dos grandes arcos de paréntesis en la Fig. 25) de elevada densidad de residuos que incluye, además de las cadenas laterales de Leu38 y Leu77, las de la Phe67 y la Leu81 (ambos residuos constantemente hidrofóbicos y grandes, Fig. 24A), con las que definen el centro del núcleo hidrofóbico, a cuyo alrededor se sitúan en SeCutA las cadenas laterales hidrofóbicas de Ala36, Val42, Ile44, Met65 y Trp 80 (el subrayado indica conservación de cadena lateral hidrofóbica, grande o pequeña, Fig. 24A), todos ellos de la misma subunidad, sugieren que este

importante núcleo hidrofóbico (ubicado en Fig. 25 con grandes paréntesis) tiene una gran importancia estructural, conectando y pegando mutuamente las dos hélices del plegamiento ferredoxina a la hoja beta de dicho plegamiento sobre la que se apoyan y a la que recubren dichas hélices, seguramente contribuyendo fuertemente a la estabilidad térmica y estructural de CutA.

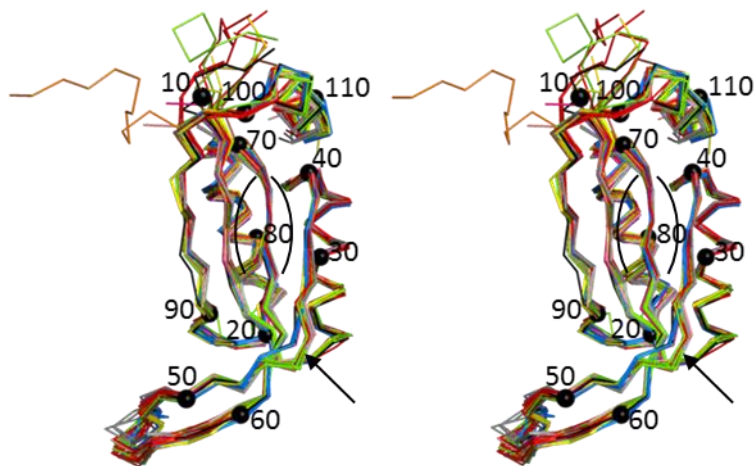


Figura 25: Representación estereoscópica de las estructuras superpuestas de las proteínas CutA depositadas en el PDB. Se representa la cadena principal de la subunidad A de todas las estructuras, excepto en el caso del PDB: 3OPK en el que se representa la cadena B debido a que es más completa que la A en el extremo N-terminal. La estructura de CutA de *S. elongatus* se colorea en negro y mediante esferas se representa el Ca de algunos de sus residuos cuya numeración se indica. Los colores asignados a las estructuras del PDB son: azul (*P. horikoshii*, PDB: 1UKU, 1J2V, 1UMJ, 2E66, 4NYP y 4NYO), morado (*A. fulgidus*, PDB:1P1L), amarillo (*T. maritima*, PDB: 1KR4, 1O5J y 1VHF), verde (*E. coli*, PDB:1NAQ, 3AA9, 3AA8, 3AH6, 3X3U, 4Y65, 4Y6I), gris (*T. thermophilus*, PDB: 1NZA, 1V6H, 4ZK7), rosa (*X. fastidiosa*, PDB:2NUH), fucsia (*Y. pestis*, PDB:3GSD), salmón (*Shewanella* sp., PDB: 3AHP), naranja (*S. enterica*, PDB:3OPK), marrón (*E. chaffeensis*, PDB:4IYQ), rojo (eucariotas, PDB: 4E98, 2ZOM, 1OSC, 1XK8 y 2ZFH). Los grandes paréntesis negros indican la región del núcleo hidrofóbico y la flecha indica el cambio estructural en la hebra beta.

Puede verse en la Fig. 25 que esta región es la que experimenta menos cambios estructurales en las subunidades de CutA, superpuesta para todas las estructuras de CutA de distintas especies existentes en el PDB. De hecho, esta figura demuestra escasa variabilidad estructural entre las subunidades de CutA de diferentes organismos en la mayoría de la cadena polipeptídica, quizá con la limitada y modesta variabilidad, que debe reflejar meramente movilidad (Capítulo 3 de la Sección 2), en el extremo de la horquilla $\beta 2$ - $\beta 3$. Existe genuina variabilidad estructural en el extremo N-terminal de la proteína, como hemos detectado nosotros dicho Capítulo 3 para SeCutA. Hay que mencionar también la existencia de cierta variación estructural en la mitad distal de $\beta 2$ (Fig. 25; señalado con una flecha), con dos conformaciones principales, una de ellas observada en las proteínas CutA de *P. horikoshii*, *A. fulgidus* (ambos organismos del filo Archaea), *T. marítima* y *S. elongatus*; mientras que la otra conformación de $\beta 2$ corresponde al resto de las proteínas CutA de estructura conocida. Estas diferentes conformaciones de $\beta 2$ fueron descritas por primera vez sobre la estructura de CutA de *Shewanella sp.* (PDB: 3AHP), donde se asoció la pérdida de un residuo (generalmente una prolina) a un aumento de estabilidad de la proteína (Sato et al., 2011). Los alineamientos de secuencia (no mostrados) prueban que dicha microdelección se da también en secuencias de proteínas CutA de otras especies de arqueas y bacterias (proteobacterias, termotogales y cianobacterias) (Arnesano et al., 2003). Quizá la proteína CutA ancestral podría corresponder a una secuencia similar a la de estos grupos. Es probable que en repetidas ocasiones en la evolución se produjera una inserción que permitiese una mayor flexibilidad a la proteína o mejorase su todavía desconocida función. Para confirmar dicha hipótesis se precisan nuevos estudios estructurales que generalicen esta observación, así como estudios funcionales que permitan resolver la incógnita de la función de esta proteína. En este sentido hay que

advertir que quizá la gran estabilidad estructural es una característica clave de las proteínas CutA, que podría estar vinculada directamente a su desconocida función, ya que en un organismo psicrotrofo (estos organismos tienen temperaturas bajas de crecimiento, no soportando temperaturas superiores a unos 20°C) se encontró (Sato et al., 2011) una termoestabilidad muy elevada de CutA, observación que hemos corroborado con la proteína CutA de un psicrófilo (*Pseudoalteromonas haloplanktis* TC125; organismo que crece óptimamente a unos 4°C) que hemos expresado, purificado y cuya estructura hemos determinado, pero que no hemos incluido en esta memoria de Tesis Doctoral por razones de tiempo.

C. Nota final

El nitrógeno tiene un papel fundamental en las moléculas que forman la materia viva, por lo que es importante mantener un buen aporte de los niveles de sus formas metabolizables en la célula para garantizar la supervivencia y el crecimiento de los organismos. Asimismo, en numerosas ocasiones, moléculas nitrogenadas o sus precursores no nitrogenados (glutamina, α -glutamaro) actúan como señalizadores, permitiendo procesos de regulación del metabolismo del nitrógeno, particularmente en microorganismos y plantas, aunque, por ejemplo, una molécula similar, el N-acetil-L-glutamato, actúa en peces, anfibios y mamíferos como interruptor clave del proceso de conversión de amonio en urea (de Cima et al., 2015; Díez-Fernández et al., 2013).

Durante los últimos años, las investigaciones llevadas a cabo por nuestro laboratorio han contribuido a esclarecer los procesos de regulación para el metabolismo de nitrógeno, así como también a la caracterización estructural de diversas proteínas implicadas en el proceso. Partiendo de la ruta

de señalización por nitrógeno en cianobacterias en la que PII y PipX se pueden considerar como elementos principales, que ya había sido utilizada como modelo en estudios anteriores en el laboratorio, en esta tesis se han caracterizado principalmente dos elementos:

- La proteína CutA; por su pertenencia a la superfamilia de proteínas PII debido a la conservación de su plegamiento estructural.

- La proteína PipY de *S. elongatus*; por la ubicación de su gen subyacente al de *pipX* en este organismo y su conservación en otras cianobacterias.

Además, dada la implicación y experiencia del grupo en patologías causadas por errores congénitos del metabolismo y la descripción de epilepsia para mutaciones en la proteína ortóloga humana, PLPHP, también la he caracterizado.

Además, como se avanzaba en la Introducción de esta tesis, a pesar de no estar incluido en esta memoria, con mi trabajo en el laboratorio también he contribuido en los estudios iniciales de la caracterización de otro elemento relacionados con el sistema: el factor de transcripción PlmA de *S. elongatus*; tras haber sido detectado como proteína de interacción con el complejo PII-PipX en experimentos de triple híbrido de levadura por miembros del laboratorio de la Dra. Contreras de la Universidad de Alicante.

La caracterización estructural de los componentes de la red de regulación del nitrógeno presenta un gran interés intrínseco, pero, además, parte del trabajo reflejado en esta memoria de tesis doctoral demuestra que este conocimiento puede ofrecer información útil para la salud humana, y no se descarta que estudios futuros en este mismo campo resulten clave para la modificación de procesos biotecnológicos y puedan llegar a tener importantes aplicaciones medioambientales.

SECCIÓN 4

Conclusiones

Sección 4. Conclusiones

1. Hemos producido en *Escherichia coli* y purificado el nuevo miembro de función poco clara de la red de señalización de PII y PipX de *Synechococcus elongatus*, PipY (COG0325), desvelando que es un monómero que contiene piridoxal fosfato (PLP) unido, cofactor que puede disociar por envejecimiento de la proteína y por reacción con el antibiótico D-cicloserina. La constante de disociación de PipY para D-cicloserina no apoya que esta proteína sea una D/L-alanina racemasa.

2. Mediante cristalografía de rayos X demostramos que PipY es un monómero organizado como un barril TIM con una hélice α extra N-terminal y con una molécula de piridoxal fosfato (PLP) expuesta y unida a un residuo de lisina. La estructura descarta una función racemasa o descarboxilasa. La pérdida de PLP se asocia a un movimiento de hasta 4.5 Å en una pequeña hélice C-terminal que podría estar involucrada en una función de PipY en homeostasis del PLP, como inductora de transferencia del cofactor.

3. La proteína ortóloga humana de PipY, PROSC (ahora re-denominada PLPHP), implicada en epilepsia genética dependiente de vitamina B₆, se produjo en *E. coli* y se purificó y caracterizó. PROSC es principalmente un monómero que contiene PLP unido mediante una base de Schiff. Usando primero PipY como modelo de PROSC, y luego PROSC humana, se caracterizaron los efectos de seis mutaciones de cambio de sentido encontradas en pacientes con epilepsia vitamina B₆-dependiente. p.Leu175Pro causó un mal plegamiento extremo, y p.Arg241Gln y p.Pro87Leu disminuyeron la solubilidad y el rendimiento de PROSC pura. En su fracción purificable, estos dos últimos mutantes, así como los mutantes p.Pro40Leu, p.Tyr69Cys y p.Arg205Gln, estaban bien plegados, a

juzgar por sus espectros de dicroísmo circular. p.Arg241Gln, p.Pro40Leu y p.Arg205Gln disminuyeron la estabilidad térmica de PROSC, monitorizada en ensayos de termofluór. Las formas mutantes p.Arg241Gln y p.Tyr69Cys respectivamente carecen de PLP o contienen una cantidad disminuida de este cofactor. p.Tyr69Cys es principalmente dimérica por formación de un puente disulfuro, posiblemente a través de su nueva cisteína, y exhibe una reducción en la accesibilidad del sitio para PLP, a juzgar por el cambio en la reactividad con D-cicloserina. Como 5 de las 6 mutaciones tienen efectos negativos sobre la estabilidad o el correcto plegamiento (en la fase de producción) de PROSC, es concebible la eficacia del uso terapéutico de chaperonas farmacológicas en epilepsia por mutaciones puntuales en PROSC.

4. Hemos probado que la proteína CutA bacteriana, miembro no canónico de la superfamilia de P_{II}, no tiene en *E. coli* y en la cianobacteria *Synechococcus elongatus* las funciones protectoras frente a metales divalentes propuestas en sus anotaciones usuales en las bases de datos, pues la transformación con el gen que la codifica en un mutante de *E.coli* sin dicho gen no mejoró la resistencia al Cu²⁺ de esta bacteria, y cuando se eliminó el gen ortólogo de *S. elongatus*, la sensibilidad de esta cianobacteria por Cu²⁺, Zn²⁺ y Co²⁺ no aumentó.

5. Probamos que CutA de *S. elongatus* producida en *E. coli* y purificada es homotrimérica en solución e interacciona con Cu²⁺, resultando en agregación, que monitorizada turbidimétricamente indica que el Cu²⁺ se une a un único tipo de sitio con un valor de K_D de aproximadamente 60 μM, un valor considerablemente más alto que las concentraciones tóxicas de este metal para *S. elongatus*.

6. La determinación de la estructura cristalina de CutA de *S. elongatus* ha revelado una estructura homotrimérica similar a la de P_{II}, con un plegamiento de cada subunidad de tipo ferredoxina ($\beta\alpha\beta$)₂ al que sigue en CutA una proyección terminal $\beta_5\alpha_3$. Esta proyección, la falta (por fijación) de un largo y flexible lazo T y la minimización del lazo B son las diferencias estructurales más importantes respecto a PII canónica. Aún los cristales preparados en presencia de Cu²⁺ carecieron de Cu²⁺ unido.

7. Las estructuras de CutA objetivan un bolsillo entre cada dos subunidades formado por residuos altamente conservados, de carácter polar y potencial eléctrico negativo, observándose la ocupación de 1 a 3 bolsillos del trímero por una molécula de Bis-Tris, sin inducción de cambios conformacionales importantes. El análisis de la unión revela la importancia de que la molécula que se une tenga carga positiva o sea neutra y tenga la capacidad de formar múltiples puentes de hidrógeno con átomos de oxígeno. Se propone la posible existencia de biomoléculas endógenas que se unan en estos bolsillos, y se especula sobre posibles mecanismos reguladores de estas uniones.

8. CutA de *S. elongatus* presenta la elevada estabilidad térmica característica de otras proteínas CutA. El análisis de los pares iónicos entre subunidades y dentro de cada subunidad y su comparación con otras proteínas CutA ya caracterizadas sugiere que los pares iónicos entre subunidades tienen un papel importante en conferir dicha elevada estabilidad térmica de CutA.

SECCIÓN 5

Referencias correspondientes a las Secciones 1 y 3

Sección 5. Referencias correspondientes a las Secciones 1 y 3

Abed, R.M.M., Dobretsov, S. y Sudesh, K. (2009) Applications of cyanobacteria in biotechnology. *J Appl Microbiol* **106**, 1–12.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. y Walter, P. (2016) 2. Química Celular y biosíntesis. En: *Biología molecular de la célula*. 6th ed. Ediciones Omega, pp. 41–106.

Arcondéguy, T., Jack, R. y Merrick, M. (2001) PII signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* **65**, 80–105.

Arnesano, F., Banci, L., Benvenuti, M., Bertini, I., Calderone, V., Mangani, S. y Viezzoli, M.S. (2003) The evolutionarily conserved trimeric structure of CutA1 proteins suggests a role in signal transduction. *J Biol Chem* **278**, 45999–46006.

Austin, S.M. y Waddell, T.G. (1999) Prebiotic synthesis of vitamin B6-type compounds. *Orig Life Evol Biosph* **29**, 287-296.

Aylward, N. y Bofinger, N. (2006) A plausible prebiotic synthesis of pyridoxal phosphate: vitamin B6 - a computational study. *Biophys Chem* **123**, 113-121.

Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., Buchanan, S.G., Buchanan, M.D., Batiyenko, Y., Christopher, J.A., Emtage, S., Eroshkina, A., Feil, I., Furlong, E.B., Gajiwala, K.S., Gao, X., He, D., Hendle, J., Huber, A., Hoda, K., Kearins, P., Kissinger, C., Laubert, B., Lewis, H.A., Lin, J., Loomis, K., Lorimer, D., Louie, G., Maletic, M., Marsh, C.D., Miller, I., Molinari, J., Muller-Dieckmann, H.J., Newman, J.M., Noland, B.W., Pagarigan, B., Park, F., Peat, T.S., Post, K.W., Radojicic, S., Ramos, A., Romero, R., Rutter, M.E., Sanderson, W.E., Schwinn, K.D., Tresser, J., Winhoven, J., Wright, T.A., Wu, L.,

Xu, J. y Harris, T.J. (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* **60**, 787-796.

Bagautdinov, B. (2014) The structures of the CutA1 proteins from *Thermus thermophilus* and *Pyrococcus horikoshii*: characterization of metal-binding sites and metal-induced assembly. *Acta Crystallogr F Struct Biol* **70**, 404–413.

Bagautdinov, B., Matsuura, Y., Bagautdinova, S., Kunishima, N. y Yutani, K. (2008) Structure of putative CutA1 from *Homo sapiens* determined at 2.05 Å resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 351-357.

Bagautdinov, B., Matsuura, Y., Yamamoto, H., Sawano, M., Ogasahara, K., Takehira, M., Kunishima, N., Katoh, E. y Yutani, K. (2015) Thermodynamic analysis of unusually thermostable CutA1 protein from human brain and its protease susceptibility. *J Biochem* **157**, 169–176.

Bale, J.B., Park, R.U., Liu, Y., Gonen, S., Gonen, T., Cascio, D., King, N.P., Yeates, T.O. y Baker, D. (2015) Structure of a designed tetrahedral protein assembly variant engineered to have improved soluble expression. *Protein Sci* **24**, 1695–1701.

Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D., Waley, S.G. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* **255**, 609-614.

Beck, C., Knoop, H., Axmann, I.M. y Steuer, R. (2012) The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics* **13**, 56.

- Berkovic, S.F. y Scheffer, I.E. (2001) Genetics of the epilepsies. *Epilepsia*. **42**, 16-23.
- Bertini, I., Jiménez, B., Pierattelli, R., Wedd, A.G. y Xiao, Z. (2008) Protonless ¹³C direct detection NMR: characterization of the 37 kDa trimeric protein CutA1. *Proteins* **70**, 1196–1205.
- Bláha, L., Babica, P. y Maršálek, B. (2009) Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdiscip Toxicol*. **2**, 36-41.
- Bowling, F.G. (2011) Pyridoxine supply in human development. *Semin Cell Dev Biol* **22**, 611–618.
- Brown, N.L., Camakaris, J., Lee, B.T., Williams, T., Morby, A.P., Parkhill, J. y Rouch, D.A. (1991) Bacterial resistances to mercury and copper. *J Cell Biochem* **46**, 106–114.
- Buchko, G.W., Abendroth, J., Clifton, M.C., Robinson, H., Zhang, Y., Hewitt, S.N., Staker, B.L., Edwards, T.E., Van Voorhis, W.C. y Myler, P.J. (2015) Structure of a CutA1 divalent-cation tolerance protein from *Cryptosporidium parvum*, the protozoal parasite responsible for cryptosporidiosis. *Acta Crystallogr F Struct Biol* **71**, 522–530.
- Burillo, S., Luque, I., Fuentes, I. y Contreras, A. (2004) Interactions between the nitrogen signal transduction protein PII and N-acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J Bacteriol* **186**, 3346–3354.
- Burkhead, J.L., Abdel-Ghany, S.E., Morrill, J.M., Pilon-Smits, E.A.H. y Pilon, M. (2003) The *Arabidopsis thaliana* CUTA gene encodes an evolutionarily conserved copper binding chloroplast protein. *Plant J* **34**, 856–867.

Burris, R.H. (1991) Nitrogenases. *J Biol Chem* **266**, 9339–9342.

Campistol, J. (2016) Epilepsy in inborn errors of metabolism with therapeutic options. *Semin Pediatr Neurol.* **23**, 321-331.

Cantos, R., Labella, J.I., Espinosa, J. y Contreras, A. (2018) The nitrogen regulator PipX acts in cis to prevent operon polarity. *Environ Microbiol Rep.* en prensa.

Cheah, E., Carr, P.D., Suffolk, P.M., Vasudevan, S.G., Dixon, N.E. y Ollis, D.L. (1994) Structure of the *Escherichia coli* signal transducing protein PII. *Structure* **2**, 981-990.

Clayton, P.T. (2006) B₆-responsive disorders: a model of vitamin dependency. *J Inherit Metab Dis.* **29**, 317-326.

Coburn, S.P. (2015) Vitamin B₆ metabolism and interactions with TNAP. *Subcell Biochem.* **76**, 207-238.

Conroy, M.J., Durand, A., Lupo, D., Li, X.-D., Bullough, P.A., Winkler, F.K. y Merrick, M. (2007) The crystal structure of the *Escherichia coli* AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc Natl Acad Sci USA* **104**, 1213–1218.

Copley, R.R. y Bork, P. (2000) Homology among $\beta\alpha_8$ barrels: implications for the evolution of metabolic pathways. *J Mol Biol.* **303**,627-641.

Coursin, D.B. (1954) Convulsive seizures in infants with pyridoxine-deficient diet. *J Am Med Assoc.* **154**, 406-408.

Creighton, T.E. (1993) *Proteins: structures and molecular properties*. W.H. Freeman and Company, Nueva York.

Crooke, H. y Cole, J. (1995) The biogenesis of c-type cytochromes in *Escherichia coli* requires a membrane-bound protein, DipZ, with a protein disulphide isomerase-like domain. *Mol Microbiol* **15**, 1139–1150.

Darin, N., Reid, E., Prunetti, L., Samuelsson, L., Husain, R.A., Wilson, M., El Yacoubi, B., Footitt, E., Chong, W.K., Wilson, L.C., Prunty, H., Pope, S., Heales, S., Lascelles, K., Champion, M., Wassmer, E., Veggiotti, P., de Crécy-Lagard, V., Mills, P.B. y Clayton, P.T. (2016) Mutations in PROSC disrupt cellular pyridoxal phosphate homeostasis and cause vitamin-B₆-dependent epilepsy. *Am J Hum Genet* **99**, 1325–1337.

De Cima, S., Polo, L.M., Díez-Fernández, C., Martínez, A.I., Cervera, J., Fita, I. y Rubio, V. (2015) Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Sci Rep* **5**, 16950.

Diez-Fernandez, C., Martínez, A.I., Pekkala, S., Barcelona, B., Pérez-Arellano, I., Guadalajara, A.M., Summar, M., Cervera, J. y Rubio, V. (2013) Molecular characterization of carbamoyl-phosphate synthetase (CPS1) deficiency using human recombinant CPS1 as a key tool. *Hum Mutat* **34**, 1149–1159.

Eliot, A.C. y Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem.* **73**, 383-415. Review.

Espaillet, A., Carrasco-López, C., Bernardo-García, N., Pietrosevoli, N., Otero, L.H., Álvarez, L., de Pedro, M.A., Pazos, F., Davis, B.M., Waldor, M.K., Hermoso,

J.A. y Cava, F. (2014) Structural basis for the broad specificity of a new family of amino-acid racemases. *Acta Crystallogr D Biol Crystallogr.* **70**, 79-90.

Espinosa, J., Forchhammer, K., Burillo, S. y Contreras, A. (2006) Interaction network in cyanobacterial nitrogen regulation: PipX, a protein that interacts in a 2-oxoglutarate dependent manner with PII and NtcA. *Mol Microbiol* **61**, 457–469.

Eswaramoorthy, S., Gerchman, S., Graziano, V., Kycia, H., Studier, F.W. y Swaminathan, S. (2003) Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta Crystallogr D Biol Crystallogr* **59**, 127–135.

Falsaperla, R. y Corsello, G. (2017) Pyridoxine dependent epilepsies: new therapeutical point of view. *Ital J Pediatr* **43**, 68.

Farrant, R.D., Walker, V., Mills, G.A., Mellor, J.M. y Langley, G.J. (2001) Pyridoxal phosphate de-activation by pyrroline-5-carboxylic acid. Increased risk of vitamin B₆ deficiency and seizures in hyperprolinemia type II. *J Biol Chem.* **276**, 15107-15116.

Flexner J, Chassin MR. (1941) Clinical studies on pyridoxine (vitamin B₆). *J Clin Invest* **20**, 313-316

Fong, S.T., Camakaris, J. y Lee, B.T. (1995) Molecular genetics of a chromosomal locus involved in copper tolerance in *Escherichia coli* K-12. *Mol Microbiol* **15**, 1127–1137.

Forcada-Nadal, A., Forchhammer, K. y Rubio, V. (2014) SPR analysis of promoter binding of *Synechocystis* PCC6803 transcription factors NtcA and CRP suggests

cross-talk and sheds light on regulation by effector molecules. *FEBS Lett.* **588**, 2270-2276.

Forcada-Nadal, A., Llácer, J.L., Contreras, A., Marco-Marín, C. y Rubio V. (2018) The PII-NAGK-PipX-NtcA regulatory axis of cyanobacteria: a tale of changing partners, allosteric effectors and non-covalent interactions. *Front Mol Biosci* **5**, 92.

Forcada-Nadal, A., Palomino-Schätzlein, M., Neira, J.L., Pineda-Lucena, A. y Rubio, V. (2017) The PipX protein, when not bound to its targets, has its signaling C-terminal helix in a flexed conformation. *Biochemistry* **56**, 3211–3224.

Forchhammer, K. y Lüddecke, J. (2016) Sensory properties of the PII signalling protein family. *FEBS J* **283**, 425–437.

Forchhammer, K. y Tandeau de Marsac, N. (1995) Phosphorylation of the PII protein (*glnB* gene product) in the cyanobacterium *Synechococcus* sp. strain PCC 7942: analysis of *in vitro* kinase activity. *J Bacteriol.* **177**, 5812-5817.

Giordano, M., Beardall, J. y Raven, J.A. (2005) CO₂ concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol* **56**, 99–131.

Gruswitz, F., O'Connell, J. y Stroud, R.M. (2007) Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 Å. *Proc Natl Acad Sci USA* **104**, 42–47.

Guan, D., Ramirez, M., Shao, L., Jacobsen, D., Barrera, I., Lutkenhaus, J. y Chen, Z. (2013) Two-component protein hydrogels assembled using an engineered disulfide-forming protein-ligand pair. *Biomacromolecules* **14**, 2909–2916.

Gundlach, J., Dickmanns, A., Schröder-Tittmann, K., Neumann, P., Kaesler, J., Kampf, J., Herzberg, C., Hammer, E., Schwede, F., Kaever, V., Tittmann, K., Stülke, J. y Ficner, R. (2015) Identification, characterization, and structure analysis of the cyclic di-AMP-binding PII-like signal transduction protein DarA. *J Biol Chem* **290**, 3069–3080.

Hegyí, H. y Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol.* **288**, 147-164.

Heinrich, A., Maheswaran, M., Ruppert, U. y Forchhammer, K. (2004) The *Synechococcus elongatus* PII signal transduction protein controls arginine synthesis by complex formation with N-acetyl-L-glutamate kinase. *Mol Microbiol* **52**, 1303–1314.

Hoffman, B.M., Lukoyanov, D., Yang, Z.-Y., Dean, D.R. y Seefeldt, L.C. (2014) Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem Rev* **114**, 4041–4062.

Hou, P., Liu, G., Zhao, Y., Shi, Z., Zheng, Q., Bu, G., Xu, H. y Zhang, Y. (2015) Role of copper and the copper-related protein CUTA in mediating APP processing and A β generation. *Neurobiol Aging* **36**, 1310–1315.

Hunt, A.D., Stokes, J., McCrory, W.W. y Stroud, H.H. (1954) Pyridoxine dependency: report of a case of intractable convulsions in an infant controlled by pyridoxine. *Pediatrics* **13**, 140–145.

- Ito, F., Usui, K., Kawahara, D., Suenaga, A., Maki, T., Kidoaki, S., Suzuki, H., Taiji, M., Itoh, M., Hayashizaki, Y. y Matsuda, T. (2010) Reversible hydrogel formation driven by protein-peptide-specific interaction and chondrocyte entrapment. *Biomaterials* **31**, 58–66.
- Ito, T., Iimori, J., Takayama, S., Moriyama, A., Yamauchi, A., Hemmi, H. y Yoshimura, T. (2013) Conserved pyridoxal protein that regulates Ile and Val metabolism. *J Bacteriol* **195**, 5439–5449.
- Ito, T., Yamauchi, A., Hemmi, H. y Yoshimura, T. (2016) Ophthalmic acid accumulation in an *Escherichia coli* mutant lacking the conserved pyridoxal 5'-phosphate-binding protein YggS. *J Biosci Bioeng* **122**, 689–693.
- Janitz, K., Wild, A., Beck, S., Savasta, S., Beluffi, G., Ziegler, A. y Volz, A. (1999) Genomic organization of the *HSET* locus and the possible association of HLA-linked genes with immotile cilia syndrome (ICS). *Immunogenetics* **49**, 644–652.
- Jansonius, J.N. (1998) Structure, evolution and action of vitamin B₆-dependent enzymes. *Curr Opin Struct Biol* **8**, 759–769.
- Jiang, P., Mayo, A.E. y Ninfa, A.J. (2007) *Escherichia coli* glutamine synthetase adenylyltransferase (ATase, EC 2.7.7.49): kinetic characterization of regulation by PII, PII-UMP, glutamine, and alpha-ketoglutarate. *Biochemistry* **46**, 4133–4146.
- Kim, K.M., Qin, T., Jiang, Y.Y., Chen, L.L., Xiong, M., Caetano-Anollés, D., Zhang, H.Y. y Caetano-Anollés, G. (2012) Protein domain structure uncovers the origin of aerobic metabolism and the rise of planetary oxygen. *Structure* **20**, 67-76.

Kinch, L.N. y Grishin, N.V. (2002) Expanding the nitrogen regulatory protein superfamily: Homology detection at below random sequence identity. *Proteins* **48**, 75–84.

King, N.P., Bale, J.B., Sheffler, W., McNamara, D.E., Gonen, S., Gonen, T., Yeates, T.O. y Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108.

Kurlemann, G., Ziegler, R., Grüneberg, M., Bömelburg, T., Ullrich, K. y Palm, D.G. (1992) Disturbance of GABA metabolism in pyridoxine-dependent seizures. *Neuropediatrics* **23**, 257-259.

Labella, J.I., Cantos, R., Espinosa, J., Forcada-Nadal, A., Rubio, V. y Contreras, A. (2017) PipY, a member of the conserved COG0325 family of PLP-binding proteins, expands the cyanobacterial nitrogen regulatory network. *Front Microbiol* **8**, 1244.

Labella, J.I., Obrebska, A., Espinosa, J., Salinas, P., Forcada-Nadal, A., Tremiño, L., Rubio, V. y Contreras, A. (2016) Expanding the cyanobacterial nitrogen regulatory network: the GntR-like regulator PlmA interacts with the PII-PipX complex. *Front Microbiol* **7**, 1677.

Laskowski, R.A., Watson, J.D. y Thornton, J.M. (2003) From protein structure to biochemical function? *J Struct Funct Genomics* **4**, 167–177.

Leklem, J.E. (1990) Vitamin B₆: a status report. *J Nutr* **120**, 1503-1507.

Lee, C.C., Ribbe, M.W. y Hu, Y. (2014) Cleaving the N-N triple bond: the transformation of dinitrogen to ammonia by nitrogenases. *Met Ions Life Sci* **14**, 147–176.

Lemire, J.A., Harrison, J.J. y Turner, R.J. (2013) Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nat Rev Microbiol* **11**, 371–384.

Liang, D., Carvalho, S., Bon, S. y Massoulié, J. (2009a) Unusual transfer of CutA into the secretory pathway, evidenced by fusion proteins with acetylcholinesterase. *FEBS J* **276**, 4473–4482.

Liang, D., Nunes-Tavares, N., Xie, H.Q., Carvalho, S., Bon, S. y Massoulié, J. (2009b) Protein CutA undergoes an unusual transfer into the secretory pathway and affects the folding, oligomerization, and secretion of acetylcholinesterase. *J Biol Chem* **284**, 5195–5207.

Llácer, J.L., Contreras, A., Forchhammer, K., Marco-Marín, C., Gil-Ortiz, F., Maldonado, R., Fita, I. y Rubio, V. (2007) The crystal structure of the complex of PII and acetylglutamate kinase reveals how PII controls the storage of nitrogen as arginine. *Proc Natl Acad Sci USA* **104**, 17644–17649.

Llácer, J.L., Espinosa, J., Castells, M.A., Contreras, A., Forchhammer, K. y Rubio, V. (2010) Structural basis for the regulation of NtcA-dependent transcription by proteins PipX and PII. *Proc Natl Acad Sci USA* **107**, 15397–15402.

Llácer, J.L., Fita, I. y Rubio, V. (2008) Arginine and nitrogen storage. *Curr Opin Struct Biol.* **18**, 673-681.

Margulis, L. y Schwartz, K. (1982) *Five kingdoms. An illustrated guide to the phyla of life on earth.* WH Freeman & Company. San Francisco, CA, USA.

Marianayagam, N.J., Sunde, M. y Matthews, J.M. (2004) The power of two: protein dimerization in biology. *Trends Biochem Sci* **29**, 618–625.

Marina, A., Uriarte, M., Barcelona, B., Fresquet, V., Cervera, J. y Rubio, V. (1998) Carbamate kinase from *Enterococcus faecalis* and *Enterococcus faecium*--cloning of the genes, studies on the enzyme expressed in *Escherichia coli*, and sequence similarity with N-acetyl-L-glutamate kinase. *Eur J Biochem* **253**,280–291.

Matsuura, Y., Ota, M., Tanaka, T., Takehira, M., Ogasahara, K., Bagautdinov, B., Kunishima, N. y Yutani, K. (2010) Remarkable improvement in the heat stability of CutA1 from *Escherichia coli* by rational protein design. *J Biochem* **148**, 449–458.

Matsuura, Y., Takehira, M., Joti, Y., Ogasahara, K., Tanaka, T., Ono, N., Kunishima, N. y Yutani, K. (2015) Thermodynamics of protein denaturation at temperatures over 100 °C: CutA1 mutant proteins substituted with hydrophobic and charged residues. *Sci Rep.* **5**, 15545.

Matsuura, Y., Takehira, M., Makhatadze, G.I., Joti, Y., Naitow, H., Kunishima, N., Yutani, K. (2018) Strategy for stabilization of CutA1 proteins due to ion-ion interactions at temperatures of over 100 °C. *Biochemistry.* **57**, 2649-2656.

Matsuura, Y., Takehira, M., Sawano, M., Ogasahara, K., Tanaka, T., Yamamoto, H., Kunishima, N., Katoh, E. y Yutani, K. (2012) Role of charged residues in stabilization of *Pyrococcus horikoshii* CutA1, which has a denaturation temperature of nearly 150 °C. *FEBS J* **279**, 78–90.

Merrick, M. (2015) Post-translational modification of PII signal transduction proteins. *Front Microbiol* **5**, 763.

Merrick, M.J. y Edwards, R.A. (1995) Nitrogen control in bacteria. *Microbiol Rev* **59**, 604–622.

Metheringham, R., Tyson, K.L., Crooke, H., Missiakas, D., Raina, S. y Cole, J.A. (1996) Effects of mutations in genes for proteins involved in disulphide bond formation in the periplasm on the activities of anaerobically induced electron transfer chains in *Escherichia coli* K12. *Molecular Gen Genet* **253**, 95–102.

Mills, P.B., Struys, E., Jakobs, C., Plecko, B., Baxter, P., Baumgartner, M., Willemsen, M.A.A.P., Omran, H., Tacke, U., Uhlenberg, B., Weschke, B. y Clayton, P.T. (2006) Mutations in antiquitin in individuals with pyridoxine-dependent seizures. *Nat Med* **12**, 307–309.

Mills, P.B., Surtees, R.A.H., Champion, M.P., Beesley, C.E., Dalton, N., Scambler, P.J., Heales, S.J.R., Briddon, A., Scheimberg, I., Hoffmann, G.F., Zschocke, J. y Clayton, P.T. (2005) Neonatal epileptic encephalopathy caused by mutations in the PNPO gene encoding pyridox(am)ine 5'-phosphate oxidase. *Hum Mol Genet* **14**, 1077–1086.

Missiakas, D., Schwager, F. y Raina, S. (1995) Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli*. *EMBO J* **14**, 3415–3424.

Myers, C.T. y Mefford, H.C. (2015) Advancing epilepsy genetics in the genomic era. *Genome Med.* **7**, 91.

Navaratnam, D.S., Fernando, F.S., Priddle, J.D., Giles, K., Clegg, S.M., Pappin, D.J., Craig, I. y Smith, A.D. (2000) Hydrophobic protein that copurifies with human brain acetylcholinesterase: amino acid sequence, genomic organization, and chromosomal localization. *J Neurochem* **74**, 2146–2153.

Naylor, M.J., Oakes, S.R., Gardiner-Garden, M., Harris, J., Blazek, K., Ho, T.W., Li, F.C., Wynick, D., Walker, A.M. y Ormandy, C.J. (2005) Transcriptional changes

underlying the secretory activation phase of mammary gland development. *Mol Endocrinol.* **19**, 1868-1883.

Nelson, D.L. y Cox, M.M. (2005) Chapter 22. Biosynthesis of amino acids, nucleotides, and related molecules. En: *Lehninger. Principles of Biochemistry*. 4th ed. WH Freeman & Company, Nueva York, pp. 833–880.

Ohashi, Y., Shi, W., Takatani, N., Aichi, M., Maeda, S., Watanabe, S., Yoshikawa, H. y Omata, T. (2011) Regulation of nitrate assimilation in cyanobacteria. *J Exp Bot* **62**, 1411–1424.

Palanca, C., Pedro-Roig, L., Llácer, J.L., Camacho, M., Bonete, M.J. y Rubio, V. (2014) The structure of a PII signaling protein from a halophilic archaeon reveals novel traits and high-salt adaptations. *FEBS J* **281**, 3299–3314.

Palanca, C. y Rubio, V. (2016) Structure of AmtR, the global nitrogen regulator of *Corynebacterium glutamicum*, in free and DNA-bound forms. *FEBS J* **283**, 1039–1059.

Palanca, C. y Rubio, V. (2017) Effects of T-loop modification on the PII-signalling protein: structure of uridylylated *Escherichia coli* GlnB bound to ATP. *Environ Microbiol Rep* **9**, 290–299.

Palm, D., Klein, H.W., Schinzel, R., Buehner, M. y Helmreich, E.J. (1990) The role of pyridoxal 5'-phosphate in glycogen phosphorylase catalysis. *Biochemistry* **29**, 1099–1107.

Panza, E., Escamilla-Honrubia, J.M., Marco-Marín, C., Gougeard, N., De Michele, G., Morra, V.B., Liguori, R., Salviati, L., Donati, M.A., Cusano, R., Pippucci, T., Ravazzolo, R., Németh, A.H., Smithson, S., Davies, S., Hurst, J.A., Bordo, D.,

Rubio, V. y Seri, M. (2016) *ALDH18A1* gene mutations cause dominant spastic paraplegia SPG9: loss of function effect and plausibility of a dominant negative mechanism. *Brain* **139**, e3.

Percudani, R. y Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* **4**, 850-854.

Percudani, R. y Peracchi, A. (2009) The B₆ database: a tool for the description and classification of vitamin B₆-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics.* **10**, 273.

Perrier, A.L., Cousin, X., Boschetti, N., Haas, R., Chatel, J.M., Bon, S., Roberts, W.L., Pickett, S.R., Massoulié, J., Rosenberry, T.L. y Krejci, E. (2000) Two distinct proteins are associated with tetrameric acetylcholinesterase on the cell surface. *J Biol Chem* **275**, 34260–34265.

Perrier, A.L., Massoulié, J. y Krejci, E. (2002) PRiMA: the membrane anchor of acetylcholinesterase in the brain. *Neuron* **33**, 275–285.

Plecko, B. (2013) Pyridoxine and pyridoxal phosphate-dependent epilepsies. *Handb Clin Neurol* **113**, 1811–1817.

Plecko, B., Zweier, M., Begemann, A., Mathis, D., Schmitt, B., Striano, P., Baethmann, M., Vari, M.S., Beccaria, F., Zara, F., Crowther, L.M., Joset, P., Sticht, H., Papuc, S.M. y Rauch, A. (2017) Confirmation of mutations in PROSC as a novel cause of vitamin B₆-dependent epilepsy. *J Med Genet* **54**, 809-814.

Powner MW, Sutherland JD. (2011) Prebiotic chemistry: a new modus operandi. *Philos Trans R Soc Lond B Biol Sci.* **366**, 2870-2877.

Prunetti, L., El Yacoubi, B., Schiavon, C.R., Kirkpatrick, E., Huang, L., Bailly, M., ElBadawi-Sidhu, M., Harrison, K., Gregory, J.F., Fiehn, O., Hanson, A.D. y de Crécy-Lagard, V. (2016) Evidence that COG0325 proteins are involved in PLP homeostasis. *Microbiology* **162**, 694–706.

Radchenko, M.V., Thornton, J. y Merrick, M. (2014) Association and dissociation of the GlnK-AmtB complex in response to cellular nitrogen status can occur in the absence of GlnK post-translational modification. *Front Microbiol* **5**, 731.

Ramón-Maiques, S., Fernández-Murga, M.L., Gil-Ortiz, F., Vagin, A., Fita, I. y Rubio, V. (2006) Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J Mol Biol* **356**, 695–713.

Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. y Rubio, V. (2002) Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure* **10**, 329–342.

Rasmussen, H. (1983) Hypophosphatasia. En: J. B. Stanbury, J. Wyngaarden, D. S. Fredrickson, J. L. Goldstein, y M. S. Brown (Eds), *The Metabolic Basis of Inherited Disease*. 5th ed. McGraw-Hill, Inc., Nueva York, pp-1497-1507.

Reitzer, L. (2003) Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu Rev Microbiol* **57**, 155–176.

Richardson, D.J. (2001) Introduction: nitrate reduction and the nitrogen cycle. *Cell Mol Life Sci* **58**, 163–164.

Roberts SJ, Szabla R, Todd ZR, Stairs S, Bučar DK, Šponer J, Sasselov DD, Powner MW. (2018) Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat Commun.* **9**, 4073.

Rouch, D., Camarakis, J. y Lee, B.T.O. (1989) Copper transport in *Escherichia coli*. En: D.H. Hamer y D.R. Winge (Eds), *Metal ion homeostasis: molecular biology and chemistry*. Wiley-Liss, Nueva York, pp. 469–477.

Sant'Anna, F.H., Trentini, D.B., de Souto Weber, S., Cecagno, R., da Silva, S.C. y Schrank, I.S. (2009) The PII superfamily revised: a novel group and evolutionary insights. *J Mol Evol* **68**, 322–336.

Sato, A., Yokotani, S., Tadokoro, T., Tanaka, S., Angkawidjaja, C., Koga, Y., Takano, K. y Kanaya, S. (2011) Crystal structure of stable protein CutA1 from psychrotrophic bacterium *Shewanella* sp. SIB1. *J Synchrotron Radiat* **18**, 6–10.

Savchenko, A., Skarina, T., Evdokimova, E., Watson, J.D., Laskowski, R., Arrowsmith, C.H., Edwards, A.M., Joachimiak, A. y Zhang, R. (2004) X-ray crystal structure of CutA from *Thermotoga maritima* at 1.4 Å resolution. *Proteins* **54**, 162–165.

Sawano, M., Yamamoto, H., Ogasahara, K., Kidokoro, S., Katoh, S., Ohnuma, T., Katoh, E., Yokoyama, S. y Yutani, K. (2008) Thermodynamic basis for the stabilities of three CutA1s from *Pyrococcus horikoshii*, *Thermus thermophilus*, and *Oryza sativa*, with unusually high denaturation temperatures. *Biochemistry* **47**, 721–730.

Sharma, S. y Prasad, A.N. (2017) Inborn errors of metabolism and epilepsy: current understanding, diagnosis, and treatment approaches. *Int J Mol Sci* **18**.

Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., Kaplan, A. y Ogawa, T. (2002) Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis. *J Biol Chem* **277**, 18658–18664.

Stadtman, E.R. (2001) The story of glutamine synthetase regulation. *J Biol Chem* **276**, 44357–44364.

Stairs S, Nikmal A, Bučar DK, Zheng SL, Szostak JW, Powner MW. (2017) Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat Commun.* **8**,15270.

Stanton A. Harris, Eric T. Stiller, Karl. Folkers (1939) Structure of Vitamin B₆. *J. Am. Chem. Soc.* **61**, 1242–1244.

Stirtan, W.G. y Withers, S.G. (1996) Phosphonate and alpha-fluorophosphonate analogue probes of the ionization state of pyridoxal 5'-phosphate (PLP) in glycogen phosphorylase. *Biochemistry* **35**, 15057–15064.

Tamás, M.J., Sharma, S.K., Ibstedt, S., Jacobson, T. y Christen P. (2014) Heavy metals and metalloids as a cause for protein misfolding and aggregation. *Biomolecules* **4**, 252-267.

Tanaka, T., Sawano, M., Ogasahara, K., Sakaguchi, Y., Bagautdinov, B., Katoh, E., Kuroishi, C., Shinkai, A., Yokoyama, S. y Yutani, K. (2006) Hyper-thermostability of CutA1 protein, with a denaturation temperature of nearly 150 degrees. *FEBS Lett* **580**, 4224–4230.

Tanaka, Y., Tsumoto, K., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I. y Kumagai, I. (2004a) Structural implications for heavy metal-induced reversible

assembly and aggregation of a protein: the case of *Pyrococcus horikoshii* CutA1. *FEBS Lett* **556**, 167–174.

Tanaka, Y., Tsumoto, K., Umetsu, M., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I., Arakawa, T. y Kumagai, I. (2004b) Structural evidence for guanidine-protein side chain interactions: crystal structure of CutA from *Pyrococcus horikoshii* in 3M guanidine hydrochloride. *Biochem Biophys Res Commun* **323**, 185–191.

Thomas, G., Coutts, G. y Merrick, M. (2000) The *glnK-amtB* operon. A conserved gene pair in prokaryotes. *Trends Genet* **16**, 11–14.

Tremiño, L., Forcada-Nadal, A., Contreras, A. y Rubio, V. (2017) Studies on cyanobacterial protein PipY shed light on structure, potential functions, and vitamin B₆-dependent epilepsy. *FEBS Lett* **591**, 3431-3442.

Tremiño, L., Forcada-Nadal, A. y Rubio V. (2018) Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations. *Hum Mutat* **39**, 1002-1013.

Wang, H.-S. y Kuo, M.-F. (2007) Vitamin B₆ related epilepsy during childhood. *Chang Gung Med J* **30**, 396–401.

Watson, S.B., Zastepa, A., Boyer, G.L. y Matthews, E. (2017) Algal bloom response and risk management: on-site response tools. *Toxicon* **129**, 144–152.

Whittaker, J.W. (2016) Intracellular trafficking of the pyridoxal cofactor. Implications for health and metabolic disease. *Arch Biochem Biophys* **592**, 20–26.

Wierenga RK. (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193-198.

Writzl, K. y Knecht, A.C. (2013) 6p21.3 microdeletion involving the SYNGAP1 gene in a patient with intellectual disability, seizures, and severe speech impairment. *Am J Med Genet A* **161A**, 1682-1685.

Xu, Y., Cheah, E., Carr, P.D., van Heeswijk, W.C., Westerhoff, H.V., Vasudevan, S.G. y Ollis, D.L. (1998) GlnK, a PII-homologue: structure reveals ATP binding site and indicates how the T-loops may be involved in molecular recognition. *J Mol Biol* **282**, 149–165.

Yang, J., Yang, H., Yan, L., Yang, L. y Yu, L. (2009) Characterization of the human CUTA isoform2 present in the stably transfected HeLa cells. *Mol Biol Rep.* **36**, 63-69.

Zhang, Y., Yao, Y., Jiang, S., Lu, Y., Liu, Y., Tao, D., Zhang, S. y Ma, Y. (2015) Identification of proteins interacting with the circadian clock protein PER1 in tumors using bacterial two-hybrid system technique. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **32**, 192-197.

Zhao, Y., Wang, Y., Hu, J., Zhang, X. y Zhang, Y.-W. (2012) CutA divalent cation tolerance homolog (*Escherichia coli*) (CUTA) regulates β -cleavage of β -amyloid precursor protein (APP) through interacting with β -site APP cleaving protein 1 (BACE1). *J Biol Chem* **287**, 11141–11150.

Zollino, M., Gurrieri, F., Orteschi, D., Marangi, G., Leuzzi, V. y Neri, G. (2011) Integrated analysis of clinical signs and literature data for the diagnosis and therapy of a previously undescribed 6p21.3 deletion syndrome. *Eur J Hum Genet* **19**, 239–242.

