



UNIVERSIDAD MICHOACANA
DE SAN NICOLÁS DE HIDALGO

INSTITUTO DE INVESTIGACIONES
QUÍMICO-BIOLÓGICAS

“Filogenia y evaluación de
aminoácidos conservados en la función
de la proteína transportadora de
cromato ChrA”

TESIS

Para obtener el Grado de:

MAESTRO EN CIENCIAS

Especialidad en:
Biología Experimental

Presenta:

IBQ. César Díaz Pérez

Asesor: Dr. Carlos Cervantes Vega
Coasesor: Dr. Héctor Riveros Rosas

Por la persona que más ha influido en mi vida,
la que siempre me apoyo en todo momento y en toda circunstancia,
aquella que en los mejores momentos esta a mi lado,
ella que en los peores instantes de mi existencia ha venido a mi
para darme el apoyo y el aliento necesario.

Madre sin ti, nada se hubiese dado...

*¡Oh, cuan lejos están aquellos días
en que cantando alegre y placentera,
jugando con mi negra cabellera,
en tu blando regazo me dormías!*

*¡Con que grato embeleso recogías
la balbuciente frase pasajera
que, por ser de mis labios la primera
con maternal orgullo repetías!*

*Hoy que de la vejez en el quebranto,
mi barba se desata en blanco armiño,
y contemplo la vida sin encanto,*

*al recordar tu celestial cariño,
e="escrito_sep04-Converted.enl" path="D:¡que, pensando en ti, me siento niño*

*un golpe di con temblorosa mano
sobre su tumba venerada y triste;
y nadie respondió... Llamé en vano
porque ¡la madre de mi amor no existe!*

*Volví a llamar, y del imperio frío
se alzo una voz que dijo: ¡Si existe!
Las madres, nunca mueren... Hijo mío
desde la tumba te vigilo triste...*

*¡Las madres, nunca mueren!
Si dejan la envoltura terrenal,
suben a Dios, en espiral de nubes...
¡La madre, es inmortal!*

(Vicente Riva Palacio)



AGRADECIMIENTOS

Gracias a mi familia, por todo el apoyo que me da, por ellos he podido seguir dentro de este apasionante mundo. Nunca tendré las palabras necesarias para expresar todo lo que son para mí.

A mi asesor, el Dr. Carlos Cervantes Vega, por tener la confianza en mi y aceptarme dentro de este excelente grupo. No pudo haberme dado mejor consejo, orientación y ejemplo.

A mi coasesor, el Dr. Héctor Riveros Rosas, por su extraordinaria guía y apoyo. Este trabajo nunca sería lo que es sin su consejo.

A mis sinodales, el Dr. Carlos Sosa y Dr. Jesús Campos García, por su observaciones y ayuda en la realización de esta tesis. Quiero agradecer en especial al Dr. Campos cuya amistad y apego al trabajo son un ejemplo para mí.

A mis grandes amigos de Irapuato, Lenin, Gloria, March, Iván, Luis David, Gina, Nidia, Alberto. Ellos que siempre han creído en mí a pesar de todo; ellos que marcaron una de la mejores etapas de mi vida; gracias a ellos que me han dado una manera distinta de ver la vida.

A mis compañeros del laboratorio, Eréndira, Esther, Rafa, Lupita, Sonia, Ricarda, Martita, Amada, por todas las ayudas, discusiones, preguntas, regaños, bromas, fiestas, en fin ... por todo. Este excelente ambiente de compañerismo se ha visto reflejado en la calidad de esta tesis.

A mis amigos de la degeneración, con los cuales hemos compartido horas de estudio, risas y lamentos, un sin fin de hermosas vivencias que me llevo, por ustedes Pat, Luz, Areli, León, Lupita, Arnold, Soni, Ross, Yess, Ido, Alí. Pronto seguirán ustedes.

A mis amigos, Gaby, Daniel, Ángel, Brenda, Diana, Bto, Lupita, que me acompañaron y me siguen acompañando en muchos momentos inolvidables.

A Miros y Cess, por esas interminables platicas y esos geniales momentos.

A Migue, Toño, José Luis, Juan Javier, Toño, Andrés; mis amigos de toda una vida, los que me han acompañado en lo bueno, lo malo, lo peor, y que espero que por muchos años más sigamos en este juego.



A. RESUMEN

La proteína ChrA, codificada en un plásmido de *Pseudomonas aeruginosa*, confiere resistencia al ión tóxico cromato (CrO_4^{2-}) mediante su expulsión del citoplasma. Dentro de las bases de datos de familias proteicas, ChrA se encuentra formando parte de la familia denominada CHR. El objetivo de este trabajo fue elaborar un análisis filogenético de la familia CHR y verificar la importancia de algunos aminoácidos conservados en la función de la proteína ChrA de *P. aeruginosa*. Se realizó una búsqueda exhaustiva de proteínas homólogas a ChrA en las bases de datos Swiss-Prot, NCBI, COG, Pfam y SYSTERS, mediante los programas Blastp y PSI-Blast. Se obtuvieron 135 secuencias homólogas a ChrA de *P. aeruginosa*: 77 secuencias de cadena larga, o bidominio (LCHR), y 58 secuencias de cadena corta, o monodominio (SCHR). Las secuencias LCHR se dividieron en los dominios amino y carboxilo y se alinearon con las proteínas SCHR, formando un conjunto de 212 dominios en total. Se construyó un árbol filogenético aplicando cuatro métodos diferentes, obteniéndose resultados muy similares. La distribución por separado de los dominios de las LCHR y las SCHR, sugiere que las proteínas SCHR forman un grupo parálogo dentro de la familia LCHR, y que sólo las proteínas LCHR tienen la capacidad de actuar como transportadores de cromato. Se encontró evidencia que sugiere que las proteínas LCHR surgieron a partir de la fusión de dos genes codificantes de proteínas SCHR; dicho evento ocurrió hace mucho tiempo. Se seleccionaron cinco aminoácidos altamente conservados (80-90% de identidad), Gly44, Gly45, Trp63, Gly92 y Gly100. Para verificar su importancia al conferir resistencia a cromato, se realizó mutagénesis dirigida y a las mutantes obtenidas se les hicieron pruebas de susceptibilidad a cromato. Se observó que en las mutaciones G44A y G100A la resistencia a cromato no se pierde totalmente, mientras que en las mutantes G44L, G44S, G45A, G45L, G45S, W63F, W63Y, G92A, G92L, G92S y G100S se abatió la resistencia a cromato completamente, lo cual confirma que estos residuos son esenciales para que ChrA confiera resistencia a cromato.



B. ÍNDICE

A. RESUMEN-----	I
B. ÍNDICE-----	II
C. ÍNDICE DE FIGURAS -----	IV
B. ÍNDICE DE TABLAS -----	V
 I. INTRODUCCIÓN-----	 1
I.1. Interacciones bacterianas con el cromato-----	1
I.1.A. Química del cromo -----	1
I.1.B. Interacciones dañinas del cromato sobre las bacterias-----	1
I.2. Estrategias bacterianas de resistencia a cromato -----	3
I.3. Proteínas ChrA-----	5
I.3.A. ChrA de <i>Pseudomonas aeruginosa</i> -----	5
I.3.B. ChrA de <i>Alcaligenes eutrophus</i> -----	9
I.4. Sistemas de clasificación y filogenia molecular-----	9
I.5. Filogenia bacteriana -----	12
I.6. Filogenia de proteínas-----	13
I.7. Antecedentes-----	14
I.7.A. El caso de la proteína ChrA-----	14
I.8. Justificación-----	14
 II. HIPÓTESIS Y OBJETIVOS-----	 16
II.1. Hipótesis -----	16
II.2. Objetivo general-----	16
II.3. Objetivos específicos-----	16
 III. MATERIALES Y MÉTODOS-----	 17
III.1. Filogenia molecular -----	17
III.1.A. Obtención de secuencias-----	17
III.1.B. Depuración de las secuencias-----	17
III.1.C. Análisis de dominios-----	18
III.1.D. Alineamiento múltiple-----	18
III.1.E. Construcción de los árboles filogenéticos-----	18
III.1.F. Identificación de aminoácidos conservados-----	19
III.2. Métodos generales-----	19
III.2.A. Reactivos y medios de cultivo-----	19
III.2.B. Cepas empleadas-----	20
III.2.C. Plásmidos empleados-----	20
III.2.D. Aislamiento y manipulación de DNA-----	20
III.2.D.i. Aislamiento de DNA plasmídico-----	20
III.2.D.ii. Electroforesis en geles de agarosa-----	22
III.2.D.iii. Aislamiento del DNA de geles de agarosa-----	23
III.2.D.iv. Tratamientos enzimáticos del DNA-----	23
III.2.D.iv.1. Restricción con endonucleasas-----	23
III.2.D.iv.2. Ligaciones-----	23
III.2.E. Métodos de transformación Bacteriana-----	24
III.2.E.i. Transformación de <i>Escherichia coli</i> por electroporación-----	24



III.2.E.i.1. Preparación de células competentes-----	24
III.2.E.i.2. Transformación-----	24
III.2.E.ii. Transformación de <i>Escherichia coli</i> por choque térmico-----	24
III.2.E.ii.1. Preparación de células competentes-----	24
III.2.E.ii.2. Transformación-----	24
III.2.E.iii. Electroporación de <i>P. aeruginosa</i> o <i>E. coli</i> sembradas recientemente en placa-----	25
III.3. Mutagénesis dirigida-----	25
III.3.A. Mutagénesis dirigida basada en PCR y digestión con <i>DpnI</i> -----	25
III.3.A.i. Diseño de oligonucleótidos mutagénicos-----	27
III.3.A.ii. Síntesis de la cadena mutada-----	27
III.3.A.iii. Digestión con <i>DpnI</i> -----	29
III.3.A.iv. Transformación-----	29
III.4. Secuenciación de los genes mutados-----	29
III.4.A. Limpieza del DNA con fenol:cloroformo y bromuro de etidio-----	29
III.4.B. Secuenciación-----	29
III.5. Transferencia de los genes mutados a <i>P. aeruginosa</i> -----	30
III.5.A. Escisión y recuperación de los genes mutados-----	30
III.5.B. Ligación y transferencia de los genes mutados a <i>P. aeruginosa</i> -----	30
III.6. Pruebas de susceptibilidad a cromato-----	30
IV. RESULTADOS-----	32
IV.1. Filogenia molecular-----	32
IV.1.A. Obtención e identificación de secuencias de la familia CHR-----	32
IV.1.B. Análisis de dominios-----	34
IV.1.C. Partición de dominios-----	39
IV.1.D. Alineamiento múltiple-----	40
IV.1.E. Reconstrucción del árbol filogenético-----	40
IV.1.E.i. Reconstrucción de la filogenia con base en los dominios independientes de CHR-----	42
IV.1.E.ii. Reconstrucción de la filogenia de las proteínas CHR (de cadena larga)-----	46
IV.1.F. Identificación de aminoácidos conservados-----	52
IV.2. Evaluación de los aminoácidos conservados en la familia CHR-----	54
IV.2.A. Obtención de las mutantes-----	54
IV.2.B. Subclonación de los genes mutantes en pUCP20-----	57
IV.2.C. Pruebas de susceptibilidad a cromato-----	62
IV.2.C.i. Glicina 44-----	62
IV.2.C.ii. Glicina 45-----	62
IV.2.C.iii. Triptófano 63-----	62
IV.2.C.iv. Glicina 92-----	66
IV.2.C.v. Glicina 100-----	66
V. DISCUSIÓN-----	69
V.1. Filogenia molecular-----	69
V.1.A. La torre de babel: un viejo problema en la actualidad-----	70
V.1.B. Filogenia de la familia CHR-----	71
V.1.B.i. Reconstrucción de la filogenia-----	72
V.1.C. Historia evolutiva de la familia CHR-----	75



V.1.D. La macrofamilia CHR-----	78
V.2. Importancia de los residuos conservados de la macrofamilia CHR-----	80
V.2.A. Mutaciones en residuos de triptófano-----	80
V.2.A.i. Triptófano 63-----	80
V.2.B. Mutaciones en residuos de glicina-----	83
V.2.B.i. Glicinas 44 y 45-----	83
V.2.B.ii. Glicinas 92 y 100 -----	86
V.2.C. Modelo estructural de los STM I, II y III de ChrA de <i>P. aeruginosa</i> -----	88
VI. CONCLUSIONES -----	91
VII. ANEXOS -----	93
VII.1. Anexo I – Grupo de datos de secuencias de proteínas usadas-----	93
VII.2. Anexo II – No todo lo que brilla es oro-----	97
VII.3. Anexo III – Código fuente de los scripts-----	98
VIII. BIBLIOGRAFÍA -----	105

B. ÍNDICE DE FIGURAS

Figura 1.- Mecanismos de transporte y toxicidad del cromato en bacterias-----	2
Figura 2.- Mecanismos bacterianos de resistencia a cromato codificados en el cromosoma -----	4
Figura 3.- Mecanismos bacterianos de resistencia a cromato codificados por plásmidos	6
Figura 4.- Proteína ChrA de <i>P. aeruginosa</i> -----	8
Figura 5.- Árbol universal basado en los estudios de las secuencias del RNA ribosomal	11
Figura 6.- Esquema de los plásmidos usados en este estudio-----	21
Figura 7 – Diagrama de la mutagénesis sitio dirigida basada en PCR y digestión con <i>DpnI</i> -----	26
Figura 8 – Esquema de la subclonación de <i>chrAHis</i> en pUCP20-----	31
Figura 9.- Dendrograma preliminar de las 135 secuencias del grupo de trabajo-----	35
Figura 10.- Distribución del tamaño de las 135 proteínas del juego de datos-----	37
Figura 11- Dendrograma preliminar de las secuencias de los 212 dominios del grupo de trabajo-----	41
Figura 12 – Topología del árbol filogenético de la familia de CHR-----	43
Figura 13 – Grupos de ChrA-----	44
Figura 14 – Topología y miembros de los grupos la familia de CHR-----	45
Figura 15 – Distribución taxonómica de las secuencias perteneciente a la familia CHR--	47
Figura 16 – Clasificación Gram de los organismos de la familia CHR-----	48
Figura 17 – Análisis Bootstrap de la rama donde se agrupa ChrA de <i>Pseudomonas aeruginosa</i> -----	49
Figura 18 – Árbol filogenético de proteínas mayores de la familia CHR-----	51
Figura 19 – Aminoácidos conservados en las 212 secuencias proteicas de la familia CHR -----	53
Figura 20 – Aminoácidos seleccionados para la mutagénesis sitio-dirigida-----	55
Figura 21 – Cambios propuestos para la evaluación de los aminoácido conservados en la familia CHR-----	56
Figura 22 – Caracterización de los plásmidos pGEMTChrAHis mutagenizados-----	58
Figura 23 – Identificación de las mutaciones puntuales del gen <i>chrAHis</i> por secuenciación-----	59



Figura 24 – Caracterización de los plásmidos pUChrA mutagenizados-----	60
Figura 25 – Caracterización de los plásmidos pUChrA mutagenizados y transferidos a <i>P. aeruginosa</i> PAO1-----	61
Figura 26 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 44 de la proteína ChrA de <i>P. aeruginosa</i> -----	63
Figura 27 – Figura 26 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 45 de la proteína ChrA de <i>P. aeruginosa</i> -----	64
Figura 28 – Pruebas de susceptibilidad a cromato de las mutantes del triptófano 63 de la proteína ChrA de <i>P. aeruginosa</i> -----	65
Figura 29 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 92 de la proteína ChrA de <i>P. aeruginosa</i> -----	67
Figura 30 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 100 de la proteína ChrA de <i>P. aeruginosa</i> -----	68
Figura 31 - Árbol filogenético de la familia CHR-----	74
Figura 32 – Posibles eventos de duplicación, fusión y reemplazo de dominios que dan origen a las diversas proteínas que forman la familia CHR-----	77
Figura 33 – Taxonomía de la familia CHR-----	79
Figura 34 – Posible papel del triptófano 63 en la proteína ChrA de <i>P. aeruginosa</i> -----	81
Figura 35 – Posible papel de las glicinas 44 y 45 en la proteína ChrA de <i>P. aeruginosa</i> -----	84
Figura 36– Posible papel de las glicinas 92 y 100 en la proteína ChrA de <i>P. aeruginosa</i> -----	87
Figura 37 – Modelo propuesto de la interacción de los segmentos transmembranales I, II y III de la proteína ChrA de <i>P. aeruginosa</i> -----	90

C. ÍNDICE DE TABLAS

Tabla 1.– Plásmidos que proporcionan resistencia a cromato-----	7
Tabla 2.– Oligos diseñados para la mutagénesis dirigida de ChrA-----	28
Tabla 3.– Resultados de la búsqueda de secuencias de transportadores de cromato y proteínas similares a ChrA de <i>P. aeruginosa</i> -----	33
Tabla 4.– Secuencias de proteínas usadas para reconstruir la filogenia de la proteína ChrA de <i>P. aeruginosa</i> -----	93



I. INTRODUCCIÓN

I.1. Interacciones bacterianas con el cromato

1.1.A. Química del cromo

El cromo es un metal que se encuentra en la naturaleza en distintos estados de oxidación, los cuales van desde -2 hasta +6 (Mertz, 1969), aunque los estados de oxidación más comunes son el cromo 0, cromo trivalente -Cr(III)- y el cromo hexavalente -Cr(VI)- (ATSDR, 2000). El Cr(III) es mucho menos soluble en agua que el Cr(VI) y por lo tanto menos móvil, su solubilidad es baja a pH 4, precipitándose a un pH mayor a 5.5 (McGrath y Smith, 1990). El Cr(VI) es un oxidante muy fuerte, el cual se reduce a Cr(III) en presencia de materia orgánica (Gaberell y cols., 2003). El Cr(VI) se encuentra de manera predominante en un pH mayor a 6, es hidrosoluble y presenta una estructura atómica tetraédrica. Los compuestos de Cr(VI) son 100 veces más permeables a las membranas biológicas que los compuestos de Cr(III) (Alcedo y Wetterhahn, 1990). El Cr(VI), es considerado la forma más tóxica de este elemento, existe como un anión en un equilibrio dependiente del pH entre distintas formas como HCrO_4^- , dicromato (Cr_2O_7^-) y cromato (CrO_4^{2-}) (McGrath y Smith, 1990).

I.1.B. Interacciones dañinas del cromato sobre las bacterias

El cromato penetra a las células por su parecido al sulfato (SO_4^{2-}) (Pardee y cols., 1966; Ballatori, 2002), un anión esencial que provee el azufre para la síntesis de proteínas (Figura 1A). La similitud del cromato con el sulfato radica en que tienen la misma estructura molecular tetraédrica y que sus distancias de enlace son muy similares, 1.64 Å y 1.54 Å, respectivamente (Pardee y cols., 1966).

Se ha estudiado este fenómeno llamado mimetismo en *Escherichia coli* (Sirko y cols., 1990) y en mayor medida en *Salmonella typhimurium*, donde el sistema captador de sulfato *cysPTWA*, el cual funciona conjuntamente con la ruta de síntesis de cisteína (Hryniewicz y cols., 1990), transporta también al cromato al interior de la célula mediante la hidrólisis de ATP (Dreyfuss, 1964; Pardee y cols., 1966). La expresión de los genes del sistema captador de sulfato de *Bacillus subtilis*, *cysA*, *cysT* y *cysP*, en *E. coli* confiere sensibilidad a cromato; más aún, en las cepas *cysP*⁺ de *B. subtilis* se inhibe el crecimiento en presencia de cromato cuando se utiliza glutatión como única fuente de azufre, lo que significa que el producto del gen *cysP* media el transporte de cromato en presencia de compuestos azufrados (Mansilla y de Mendoza, 2000). Igualmente, el cromato es un



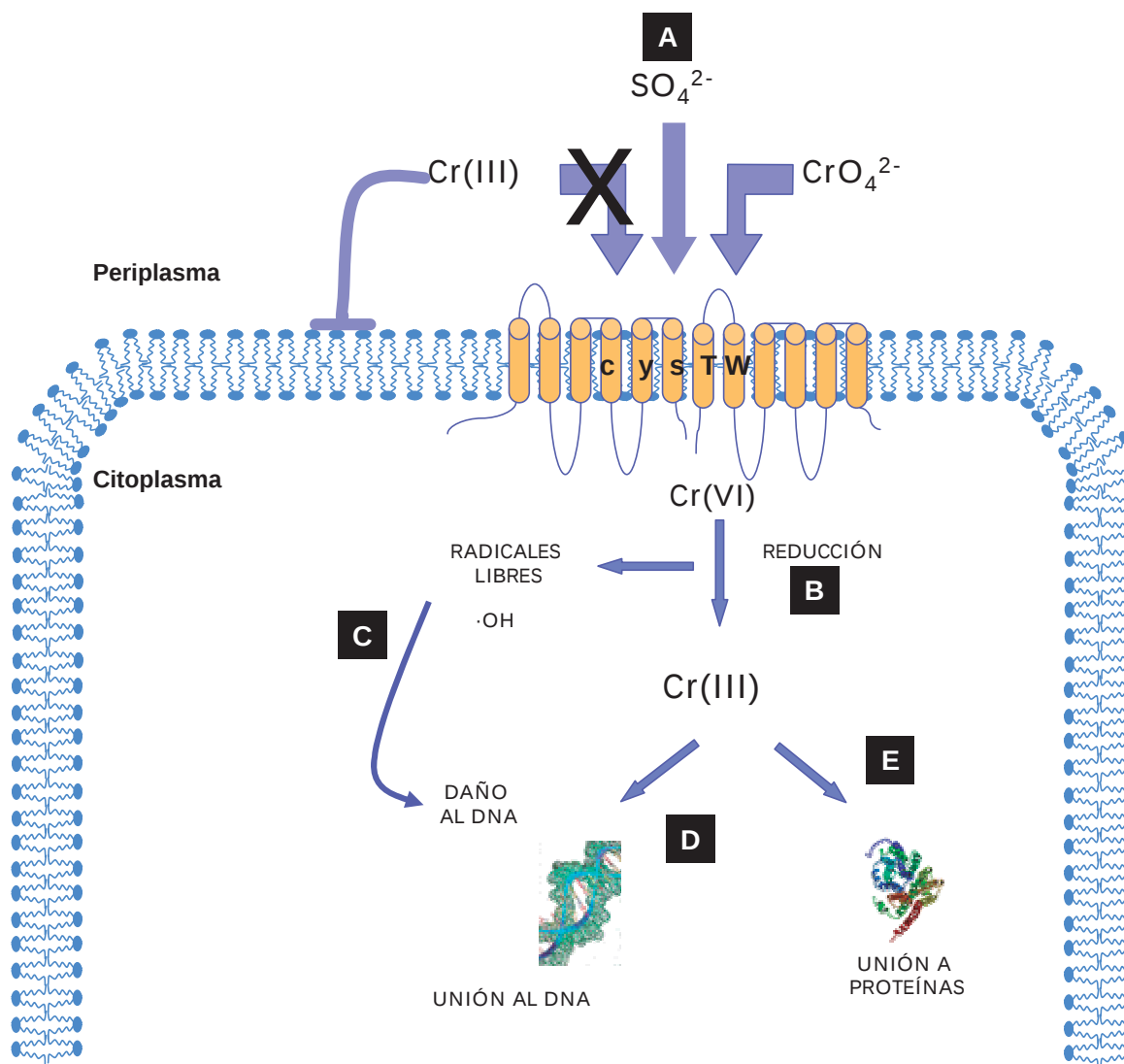


Figura 1 – Mecanismos de transporte y toxicidad del cromato en bacterias. A) El CrO_4^{2-} [Cr(VI)] ingresa a la célula a través de la vía de entrada del sulfato; las membranas biológicas son impermeables al Cr(III) , por lo que éste resulta prácticamente inocuo extracelularmente. B) Dentro de la célula el Cr(VI) puede reducirse a Cr(III) ; C) en este proceso se pueden formar radicales libres y productos intermedios como Cr(V) que interactúan con el DNA. D) El Cr(III) también puede interactuar con el DNA e incluso E) con las proteínas. Adaptado de Miranda (2003).

inhibidor competitivo del transporte de sulfato en *Pseudomonas fluorescens* (Ohtake y cols., 1987). El Cr(III) no es reconocido por el sistema de transporte de sulfato, además no penetra la membrana plasmática, ya que esta es impermeable a esta forma del cromo (Alcedo y Wetterhahn, 1990).

Los efectos tóxicos del cromato van desde la inhibición del crecimiento hasta la mutagénesis, dependiendo de la concentración utilizada. En *Shewanella spp*, una de las bacteria reductoras de metales más estudiadas, su crecimiento se ve inhibido cuando se agrega cromato al medio de crecimiento en bajas concentraciones (0.015 mM) (Viamajala y cols., 2004). Estudios en *S. typhimurium* demuestran el carácter mutagénico del cromato, ya que causa errores del marco de lectura y sustituciones de bases en el DNA (Petrilli y De Flora, 1977). Además, las sales de cromo potencian el efecto de otros agentes mutagénicos en *E. coli* y *S. typhimurium* (LaVelle, 1986). En *E. coli* se ha visto que el cromato es extremadamente genotóxico, aunque sólo causa un bajo grado de mutagénesis; más aún, se han observado aductos formados en el DNA por el Cr(III) en unión con ascorbato, los cuales bloquean la replicación (Quievryn y cols., 2003). En lodos activados, el cromato y el Cr(III) inhiben el crecimiento de los consorcios bacterianos (Vankova y cols., 1999).

1.2. Estrategias bacterianas de resistencia a cromato

Las bacterias a lo largo de su existencia se han convertido en especialistas de la supervivencia, con lo cual no es extraño que hayan desarrollado diversas estrategias de resistencia a cromato; dichas estrategias pueden ser clasificadas como cromosómicas y plasmídicas. Dentro de las estrategias cromosómicas de resistencia a cromato (Figura 2), podemos mencionar: la resistencia mediante mutaciones (Silver y Walderhaug, 1992); la reducción ya sea periplásmica o interna (Campos-García y cols., 1997), los mecanismos de inactivación de radicales libres (Fournier y cols., 2004) y los sistemas de reparación del DNA (Kada y cols., 1972; Miranda, 2003; González, 2004). En la resistencia plasmídica, los sistemas más estudiados son los de expulsión mediados por proteínas de membrana (Cervantes y cols., 1990; Nies y cols., 1990) y en menor medida los mecanismos de inactivación de radicales libres (Juhnke y cols., 2002).

Dado que el cromato entra a la célula por el sistema de transporte de sulfato, la resistencia cromosómica se puede dar mediante mutaciones en los genes de este sistema



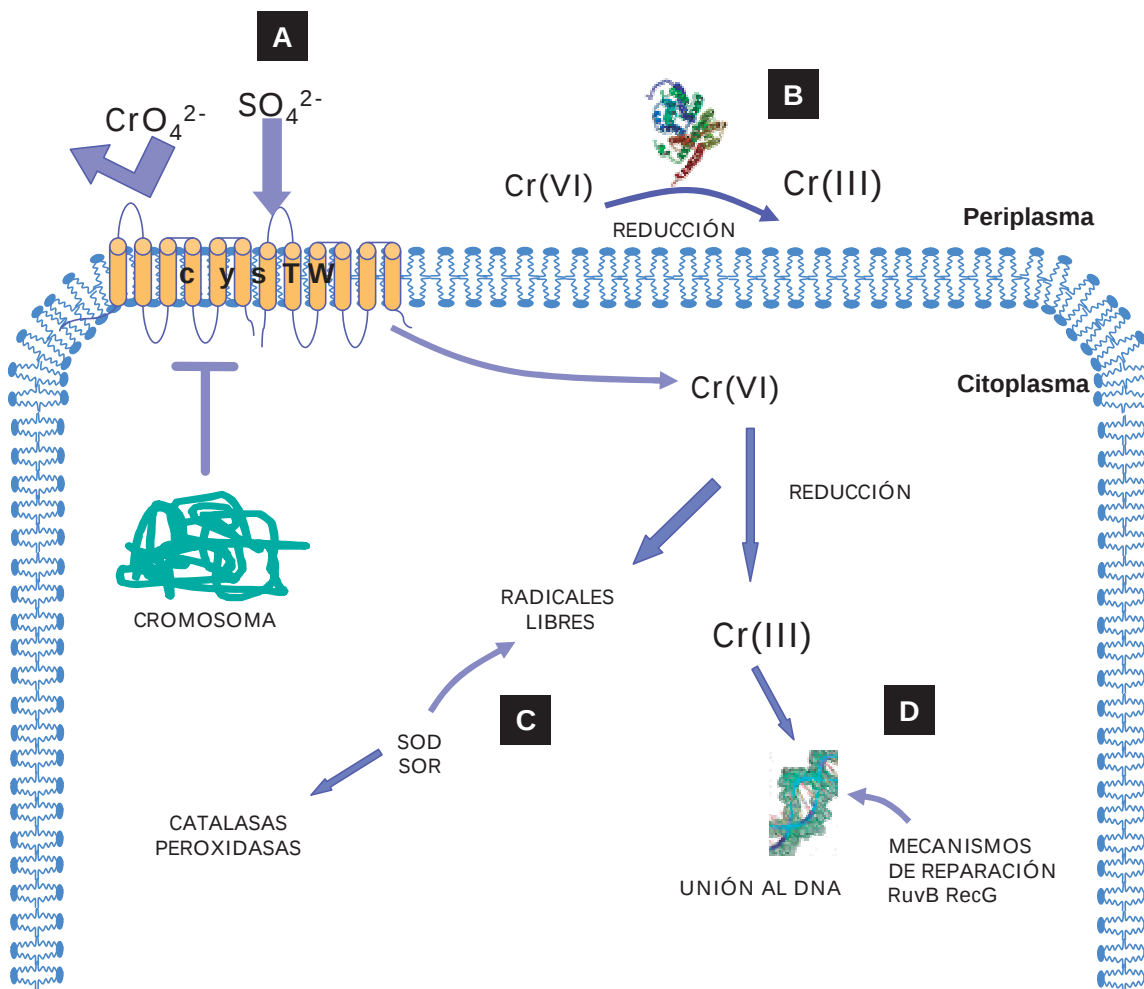


Figura 2 – Mecanismos bacterianos de resistencia a cromato codificados en el cromosoma. A) El mecanismo mediado por mutaciones en el sistema de transporte de sulfato altera la entrada de cromato. B) Reducción periplásmica que transforma el Cr(VI) a Cr(III), al cual la célula es impermeable. C) Las enzimas superóxido dismutasa (SOD), superóxido reductasa (SOR) y, subsecuentemente, catalasas y peroxidasas inactivan los radicales libres disminuyendo el efecto tóxico del cromato. D) Los daños en el DNA ocasionados por los compuestos del cromato puedan activar sistemas de reparación del DNA. Adaptado de Miranda (2003).

que originan bacterias resistentes al cromato (Figura 2A) (Pardee y cols., 1966; Cervantes y cols., 2001)

La reducción del cromato a la especie menos permeable Cr(III) (Figura 2B), es llevada a cabo por distintas especies bacterianas, entre estas se encuentran *Enterobacter cloacae* HO1, la bacteria reductora de cromato mejor descrita (Ohtake y cols., 1990; Wang y cols., 1990; Cervantes y cols., 2001), *Pseudomonas ambigua* (Suzuki y cols., 1992), *Bacillus* sp. QC1-2 (Campos-García y cols., 1997), *Desulfotomaculum reducens*, *Pantoea agglomerans*, *Desulfovibrio vulgaris*, *E. coli*, *Shewanella oneidensis* (Arias y Tebo, 2003), *Pseudomonas putida* y *E. coli* (Ackerley y cols., 2004). En el género *Pseudomonas* existen dos sistemas separados de reducción, uno unido a membrana y otro citoplasmático (McLean y Beveridge, 2001).

La destoxificación con enzimas como la superóxido dismutasa (SOD) y la superóxido reductasa (SOR) se ha visto en *D. vulgaris* (Figura 2C) (Fournier y cols., 2004). La reparación del DNA es otro mecanismo de resistencia, en *P. aeruginosa* mutantes insercionales en los genes de helicasas del sistema de reparación de DNA por recombinación, *recG* y *ruvB*, son más susceptibles a cromato que las cepas silvestres (Figura 2D) (Miranda, 2003; González, 2004).

El otro tipo de estrategia de resistencia reside en la expulsión de cromato mediada por proteínas transportadoras membranales, las cuales están codificadas en plásmidos (Figura 3). Se conocen un gran número de determinantes plasmídicos que confieren resistencia a cromato (Tabla 1), sin embargo, solamente en dos sistemas se ha probado la expulsión de cromato, además de que se han caracterizado a nivel molecular, ChrA de *P. aeruginosa* (Cervantes y cols., 1990) y ChrA de *Alcaligenes eutrophus* (también conocida como *Ralstonia metallidurans* o ahora *Cupriavidus necator*) (Nies y cols., 1990).

I.3. Proteínas ChrA

I.3.A. ChrA de *Pseudomonas aeruginosa*

ChrA es una proteína de 416 aminoácidos (no. de acceso P14285), aislada de una cepa clínica de *P. aeruginosa* que está codificada por el gen *chrA* del plásmido pUM505 (Cervantes y cols., 1990). Esta proteína confiere resistencia al cromato mediante un sistema de expulsión (Pimentel y cols., 2002) mediado por la energía que otorgan los sistemas generadores de potencial de membrana (Álvarez y cols., 1999) (Figura 4A). El contenido y distribución de los aminoácidos hidrofóbicos de ChrA, así como su perfil



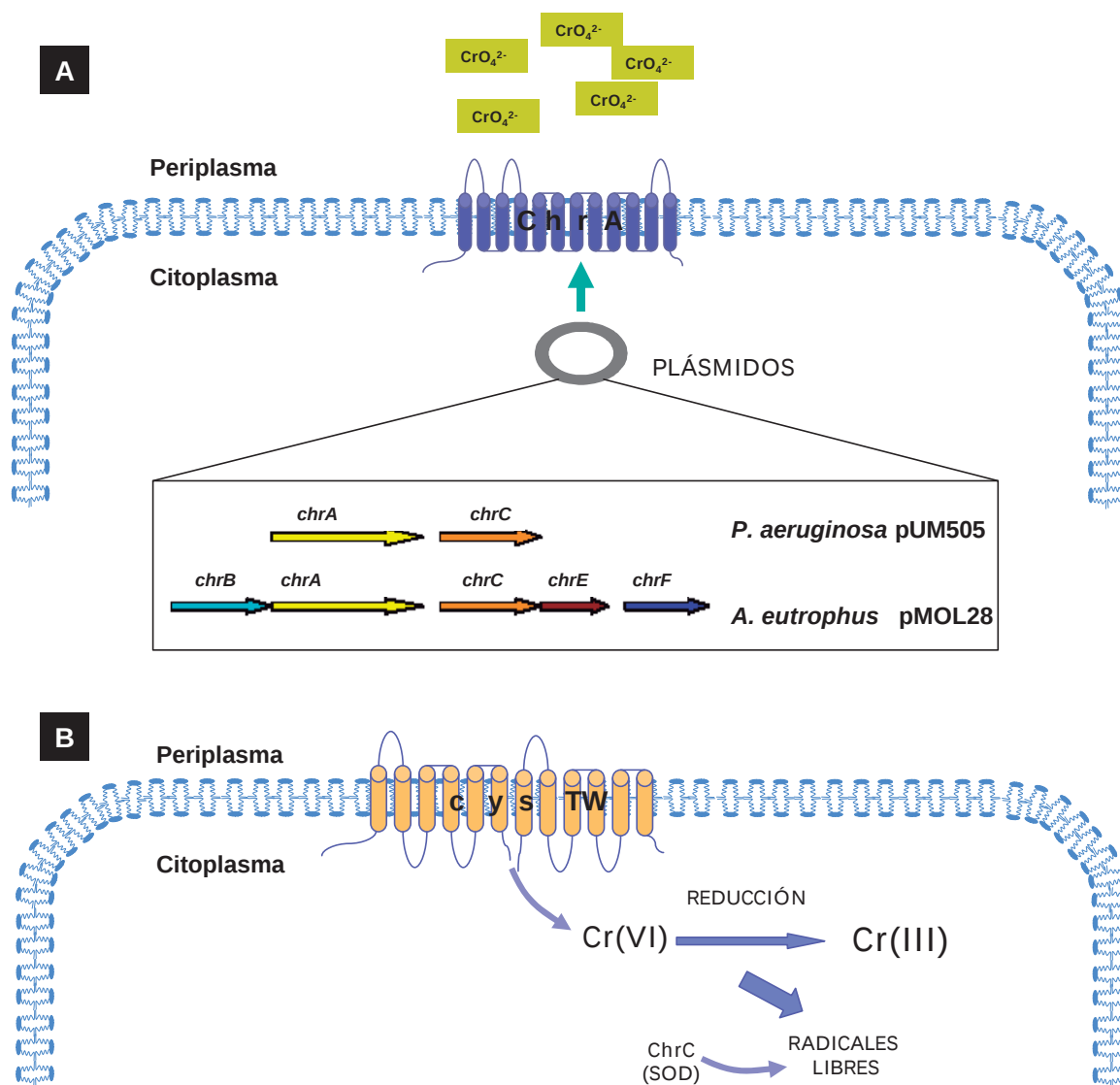


Figura 3 – Mecanismos bacterianos de resistencia a cromato codificados por plásmidos. A) Se muestra la región y los genes identificados en los plásmidos de *P. aeruginosa* y *A. eutrophus* que confieren resistencia a cromato en las cuales ChrA es el transportador de cromato. B) Sistema de reducción propuesto en *A. eutrophus* donde la proteína ChrC posee actividad de superóxido dismutasa.

Tabla 1.- Plásmidos que proporcionan resistencia a cromato.

BACTERIAS	PLÁSMIDO	ORIGEN	REFERENCIA
<i>Streptococcus lactis</i>	pLM3001	Fermentación láctea	(Efsthathiou y McKay, 1977)
<i>Pseudomonas aeruginosa</i>	pMG6	Clínico	(Summers y Jacoby, 1978)
<i>Pseudomonas aeruginosa</i>	pUM505	Clínico	(Cervantes y cols., 1990)
<i>Alcaligenes eutrophus</i>	pMOL28	Suelo contaminado	(Nies y cols., 1990)
<i>Pseudomonas mendocina</i>	pARI180	Suelo contaminado	(Dhakephalkar y cols., 1996)
<i>Pseudomonas fluorescens</i>	pLHB1	Sedimento de río	(Harrison, 2001)



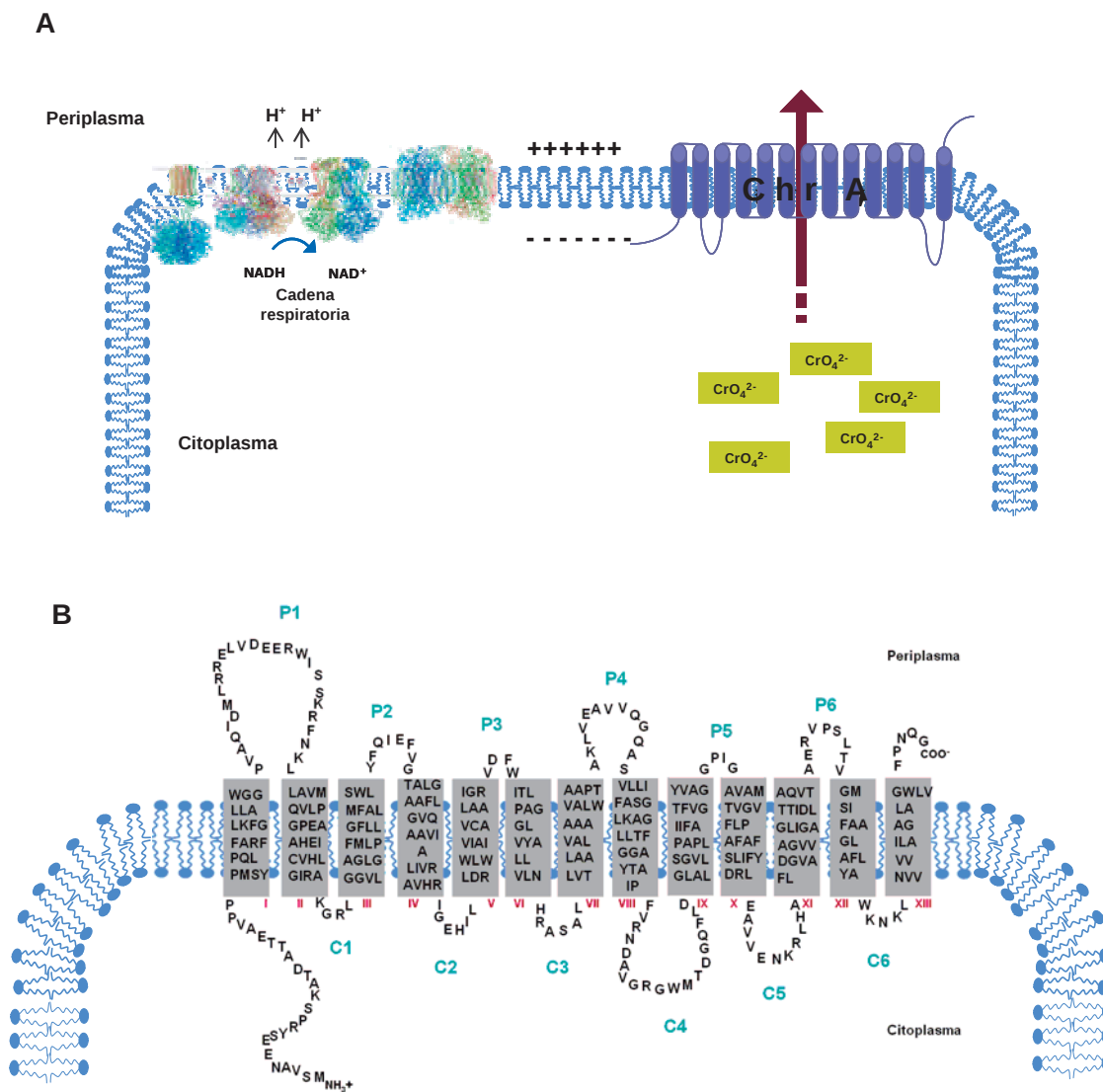


Figura 4 – Proteína ChrA de *P. aeruginosa*. A) Mecanismo de resistencia a cromato mediado por ChrA. El cromato es expulsado por la proteína ChrA, la cual utiliza el potencial de membrana producido por los componentes de la cadena respiratoria (Álvarez y cols., 1999). B) Secuencia y probable estructura secundaria. ChrA consta de 13 segmentos transmembranales, 6 asas periplásmicas y 6 asas citoplasmáticas. P se refiere a las asas periplásmicas y C a las asas citoplásmicas; en número romano (I-XIII) se marcan los segmentos transmembranales.

hidropático, parecen indicar que es una proteína transmembranal (Cervantes y Silver, 1992). Además, las predicciones de la estructura secundaria indican que se trata de una proteína de 13 segmentos transmembranales (Figura 4B).

I.3.B. ChrA de *Alcaligenes eutrophus*

La proteína ChrA de *A. eutrophus* confiere resistencia cromato tal vez mediante un sistema de expulsión (Nies y cols., 1990). Esta proteína hidrofóbica, codificada por el gen *chrA* del plásmido pMOL28, es de 401 aminoácidos y su perfil de hidrofobicidad, así como el análisis mediante fusiones traduccionales con β -galactosidasa y fosfatasa alcalina, revelan que posiblemente contiene 10 segmentos transmembranales (Nies y cols., 1998). Este sistema cuenta con un segundo gen, *chrB*, el cual se encuentra río arriba del gen *chrA*. La ausencia de la proteína ChrB provoca una superacumulación de cromato dentro de la célula e hipersensibilidad al mismo, aunado a que, aparentemente, ChrB no presenta segmentos transmembranales, es probable que esta proteína sea un regulador del sistema de expulsión de cromato (Nies y cols., 1990). Cabe hacer mención que, aunque aún no se ha comprobado su función, *A. eutrophus* presenta en su cromosoma el gen *chrA*₂ del operón *chr*₂, el cual presenta una gran similitud con los sistemas plasmídicos de expulsión de cromato (Juhnke y cols., 2002).

I.4. Sistemas de clasificación y filogenia molecular

Constantemente los naturalistas se han esforzado en construir un esquema de clasificación de los organismos vivos, siguiendo el postulado de Dobzhansky “nada en biología tiene sentido excepto a la luz de la evolución” (Dobzhansky, 1973). Este sistema de clasificación debe de ser significativo, que pueda recrear, idealmente, el “orden natural”, y las relaciones “únicas y reales” entre los organismos vivos (Baptiste y Brochier, 2004).

Desde antes de C. Darwin, se creía que las plantas y los animales eran las divisiones primarias de la vida, por lo que se propuso la existencia de dos grandes reinos, *Plantae* y *Animalia*; tiempo después, en 1866, Haeckel propuso la introducción de un tercer reino al cual llamo *Protista*. El reino *Monera* o *Bacteria* fue añadido en 1938 por Copelnad; y el reino *Fungi*, el cual completó la idea general de los cinco reinos, fue propuesto por Whittaker en 1959. Aunque las ideas de separar los organismos en otro tipo de clasificaciones han seguido, por ejemplo, Chatton, Stainer y van Niel postularon clasificar los organismos vivos en dos grupos, los procariotas y los eucariotas (Brown, 2001).



La idea de representar la evolución de los seres vivos como un árbol, con las ramas identificando a las especies existentes y los nodos como antecesores extintos, viene de Darwin (Wolf y cols., 2002). Al principio los árboles fueron solamente árboles de especiación, hasta que los trabajos de Woese trajeron consigo el mejor entendimiento de la historia de la vida (Wolf y cols., 2002), y actualmente la idea de la filogenia siempre va acompañada de esta representación gráfica y han sido integrados a los estudios evolutivos (Penny, 2004).

La última clasificación de los organismos vivos aceptada, aunque con controversias, es la clasificación hecha por Woese, Kandler y Wheelis, la cual no contradice la idea de los cinco reinos de Whittaker, se basa en la existencia de tres suprarreinios o dominios: *Bacteria* (antes Eubacterias), *Archaea* (antes Archaeobacterias) y *Eukarya* (antes Eucariotes) (Woese y cols., 1990). Esta idea, que se visualiza como el "Árbol de la vida" (Figura 5), se basa en estudios hechos con las secuencias de nucleótidos del RNA ribosomal, una molécula ubicua, cuya función es universal y que cambia muy poco con el tiempo (Woese y Fox, 1977), por la cual es posible no sólo hacer grandes filogenias, sino además se puede ir más profundo en las ramas del árbol, hasta el punto de saber dónde los tres dominios pudieron estar unificados (Woese, 2000). Aunado a esto, los postulados de Woese redefinen las ideas sobre la forma del antecesor común (Woese, 1998). Esta clasificación de la vida es el principal y más grande ejemplo de la filogenia molecular.

La filogenia molecular, o sistemática filogenética, consiste en la obtención de información filogenética a partir del estudio de secuencias moleculares, es decir, secuencias de nucleótidos (DNA, RNA) o de aminoácidos (proteínas) (Baldauf, 2003). La sistemática filogenética, referida como "cladística", "cladismo", "Hennigismo" o "Método Hennigiano", fue introducida entre los biólogos por el entomólogo alemán Willi Hennig en 1950. En su sentido original, como Hennig lo definió, la Sistemática Filogenética descansa en el postulado de una descendencia ha tenido lugar como modificación o evolución en el sentido moderno. Por lo tanto, los caracteres observados en cualquier organismo pueden mostrar un estado primitivo o uno avanzado (derivado). El fundamento de la Sistemática filogenética de Hennig es que las similitudes de los organismos no siempre tienen el mismo significado en las relaciones entre éstos. La sistemática filogenética abarca el campo de la biología comparativa, la cual trata sobre la diversidad de los organismos, la



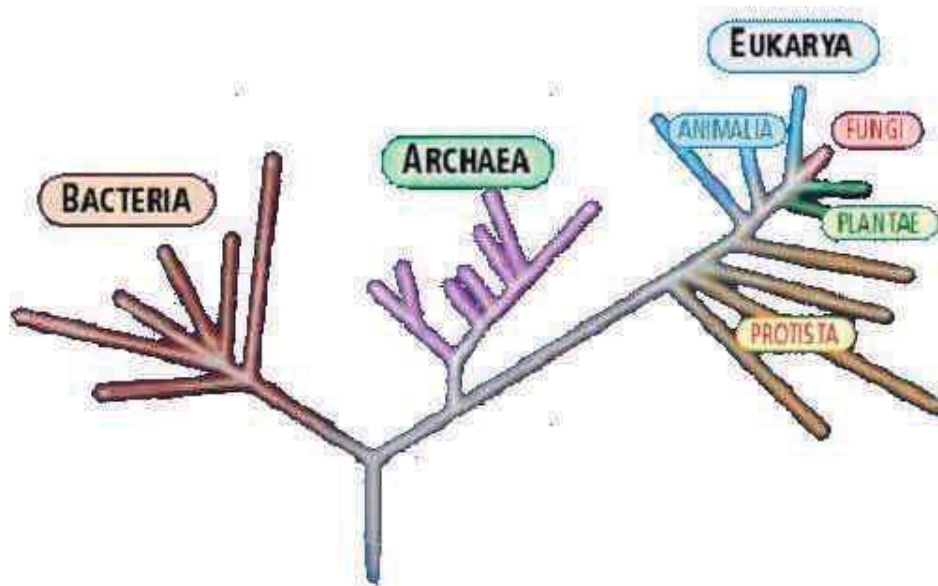


Figura 5 – Árbol universal basado en los estudios de las secuencias del RNA ribosomal. Este árbol, propuesto por Woese y cols. (1990), agrupa a los seres vivos en tres grandes dominios, las Bacterias, las Arqueas y los Eucariotas, estableciendo la raíz del árbol entre las bacterias y las arqueas.

distribución de los caracteres entre ellos y el posible significado de este patrón, es decir, cómo son, cómo están relacionadas y busca una posible explicación de la diversidad (Telford y Budd, 2003).

La evolución requiere de variación génica dentro de las especies, dicha variación dentro de la población determinará, en parte, el potencial evolutivo. Aquellos individuos más aptos para sobrevivir y reproducirse en el ambiente cambiante pasarán su composición genética a las generaciones posteriores (Harrison, 2001). Esta idea aplica desde los organismos más avanzados hasta los organismos inferiores, por lo que las bacterias también se rigen por el principio de adaptación.

I.5. Filogenia bacteriana

El registro fósil sugiere que las bacterias han existido desde hace aproximadamente 3,500 millones de años, aunque es casi imposible distinguir entre los microfósiles de los primeros procariotes y los primeros eucariotes unicelulares. Además, las arqueas y las bacterias son casi indistinguibles unas de otras en términos de morfología general, por lo tanto, a veces la información de las secuencias moleculares es la única forma disponible de determinar las relaciones evolutivas entre los tres dominios de la vida (Brown, 2001).

Sin embargo, el dominio bacteriano siempre ha sido un grupo difícil al hacer comparaciones, esto es principalmente debido a la transferencia lateral de genes, la cual ocurre frecuentemente en bacterias y ha sido catalogada como rampante para toda clase de genes, resultando potencialmente en una diversidad de historias filogenéticas, complicando o incluso deteniendo los intentos por reconstruir la historia de la evolución bacteriana en los eventos más antiguos o en los más recientes (Philippe y Douady, 2003).

La revolución de la secuenciación de genes y la disponibilidad de secuencias de genomas completos han hecho accesible una gran cantidad de información histórica contenida en los genes, por lo que con el tiempo la taxonomía ha tenido que basar la búsqueda de un sistema de clasificación natural a partir de los datos moleculares (Woese y cols., 1990). Una de las oportunidades inmediatas para realizar este tipo de estudios son las bacterias, ya que sus genomas son pequeños y presentan, relativamente, poca complejidad, además de que los datos de secuenciación de varios grupos de genomas relacionados están disponibles (Lerat y cols., 2003).



I.6. Filogenia de proteínas

A nivel molecular se pueden llevar a cabo dos tipos de análisis filogenéticos, de ácidos nucleicos o de proteínas. El uso de cada tipo de moléculas tiene sus propias y distintas ventajas y desventajas. El DNA presenta las siguientes ventajas: es más fácil de secuenciar, por lo que es más rápida la obtención de información; es mejor para inferir relaciones evolutivas recientes, los programas computacionales para reconstruir filogenias de nucleótidos son más rápidos y requieren de menos memoria (Hall, 2001). Por otra parte, algunas de las ventajas de usar proteínas son: se puede obtener la secuencia proteica a partir de la secuencia de DNA; se ha visto, en general, un grado de evolución molecular constante en todos los linajes evolutivos, es decir un reloj molecular, por lo cual se puede estimar la divergencia de las especies y reconstruir las relaciones evolutivas (Zuckermandl y Pauling, 1965); las proteínas están mejor conservadas que el DNA, por lo que son mejores para inferir relaciones más antiguas de genes y especies; los modelos matemáticos son más simples que para el DNA (Nei y Kumar, 2000).

En algunos casos, debido a la divergencia, se pueden hacer alineamientos de aminoácidos y no de nucleótidos (Hall, 2001); además, una de las principales virtudes del empleo de las proteínas es la posibilidad de comparar de manera directa y rápida los aminoácidos más conservados para dilucidar los sitios activos, sitios de interacción, aminoácidos de transporte, aminoácidos estructurales, etc. El análisis detallado de estas ventajas y desventajas, hace que el uso de filogenias a partir de secuencias proteicas sea en general el análisis elegido (Hall, 2001).

Ahora bien, la relación evolutiva de las proteínas, o los genes, puede ser clasificada como ortóloga, si las proteínas poseen de un origen común, presentan la misma función y surgen de manera natural durante los procesos de especiación, o bien pueden ser parálogas, si surgen de un evento de duplicación del gen que las codifica dentro de una misma especie, por lo que generalmente adquieren una función distinta durante el transcurso de la evolución (Baldauf, 2003). Para llevar a cabo un análisis filogenético es necesario la identificación de las proteínas ortólogas y parálogas, ya que son la base para entender los aspectos funcionales y evolutivos, además de ayudar a dilucidar los escenarios evolutivos, encontrar los eventos de pérdida de los genes codificantes de las proteínas y localizar los eventos de transferencia lateral (Koonin y cols., 2004), es por esto que las bases de datos de familias proteicas juegan un papel muy importante dentro de la filogenia.



I.7. ANTECEDENTES

I.7.A. El caso de la proteína ChrA

Se han descrito proteínas similares a ChrA, lo que ha causado que se agrupen estas secuencias en una familia de proteínas, la familia CHR, definida como una familia de antiportadores de cromato/sulfato (Nies y cols., 1998). El sistema de clasificación TC (Transport Classification) (Saier, 2000) ubica a las proteínas transportadoras del ión cromato en la familia 2.A.51; dichas proteínas son transportadores que requieren de un potencial electroquímico para su funcionalidad, tienen un tamaño de alrededor de 400 aminoácidos, con alrededor de 10 segmentos transmembranales. Esta familia de proteínas se identificó como tal con base en 10 miembros reportados (Saier, 2000). En un análisis posterior dentro de los genomas bacterianos secuenciados, en 2003, Nies encontró 21 genes homólogos a ChrA (Nies, 2003).

Las proteínas ChrA se clasifican dentro de la familia CHR, que a su vez son miembros cercanos de la superfamilia de transportadores MFS (Major Facilitator Superfamily), también llamada familia de uniportadores-simportadores-antiportadores- (Pao y cols., 1998), y al igual que esta familia, los miembros de la familia CHR se caracterizan por ser proteínas de un solo péptido, que transportan solutos pequeños en respuesta a gradientes quimiostáticos (Pao y cols., 1998); además, de forma similar a algunas de las permeasas de la MFS, se cree que la familia CHR posiblemente surgió a partir de una duplicación intergénica en tándem (Rubin y cols., 1990).

En estos momentos se tiene a la mano una gran cantidad de información, de libre acceso a la comunidad, a la espera de ser analizada a mayor detalle, por lo que algunos de los grandes esfuerzos en el campo de la genómica y bioinformática se han enfocado a la creación de bases de datos de familias relacionadas, las cuales agrupan proteínas cuyo parecido a nivel de la secuencia de aminoácidos parece indicar que se trata de productos de genes ortólogos. Dentro de este tipo de iniciativas se encuentran las bases de datos COG (Clusters of Orthologous Groups of proteins) del NCBI (National Center for Biotechnology Information) (Tatusov y cols., 2003), Pfam (Protein families) del Instituto Sanger (Bateman y cols., 2004) y SYSTERS del Instituto Max Planck (Meinel y cols., 2005). Estas tres bases de datos hacen mención a la familia de proteínas CHR.

I.8. JUSTIFICACIÓN

La filogenia puede usarse de diversas maneras para inferir los efectos de la selección natural en la función molecular (Chang y Donoghue, 2000). Uno de los posibles



usos es la utilización de los residuos conservados para la inferencia de posibles sitios de unión a sustratos o cofactores dentro de una proteína; más aún, el rastreo de los cambios en las proteínas a lo largo de las ramas de los árboles filogenéticos puede proveer información importante dentro de las funciones moleculares y el papel de la selección al moldear la relación entre la estructura y la función (Chang y Donoghue, 2000).

La gran cantidad de información disponible en bases de datos de libre acceso, nos da la oportunidad de realizar análisis genómicos a gran escala de manera fácil, rápida y con pocos recursos, por ejemplo, en el caso de la familia CHR, la cual está reportada en tres importantes bases de datos de familias proteicas.

Actualmente se han secuenciado 185 genomas de bacterias, 20 de arqueas y 15 de eucariotes, además están en proceso más de 200 proyectos de secuenciación (NCBI), por lo que la información referente a familias proteicas crece día a día. Sin embargo, esta información se encuentra muy mal curada, es decir el grado de análisis está en una etapa inicial, como lo demuestra el grado de redundancia dentro de las bases de datos del NCBI, problema que se extiende a otras fuentes de información.

Dentro de los estudios para la caracterización de la proteína ChrA de *P. aeruginosa* se han efectuado análisis de mutagénesis al azar, con la consiguiente obtención de sitios esenciales para la función de ChrA (Aguilera y cols., 2004), dado que con esta estrategia no es posible controlar el cambio del aminoácido que se muta, un complemento a esta estrategia puede ser la mutagénesis dirigida a sitios específicos, provocando cambios sutiles para poder tener más información sobre la importancia del aminoácido original. Esta estrategia se ha llevado a cabo para algunos de los aminoácidos que se han mutado al azar en ChrA de *P. aeruginosa* (Aguilar, 2005).

Una alternativa al laborioso proceso de la mutagénesis al azar es la utilización de un abordaje bioinformático, mediante el uso de técnicas filogenéticas, para la localización de residuos conservados. Mediante la implementación de procesos bioinformáticos se puede hacer más eficiente y eficaz la búsqueda y caracterización de genes ortólogos de la familia CHR, todo esto, complementado con el procedimiento de mutagénesis sitio-dirigida a aminoácidos específicos, nos podrá ayudar a probar la importancia de la conservación de los residuos dentro de esta familia.



II. HIPÓTESIS Y OBJETIVOS

II.1. HIPÓTESIS

Los residuos conservados en la familia de proteínas transportadoras CHR son esenciales para la correcta función de la proteína.

II.2. OBJETIVO GENERAL

Determinar los residuos conservados evolutivamente en la familia de proteínas CHR y evaluar experimentalmente la importancia de algunos de éstos para la función de la proteína ChrA de *P. aeruginosa*.

II.3 OBJETIVOS ESPECÍFICOS

1. Se hará una búsqueda exhaustiva de las proteínas homólogas a ChrA dentro de las bases de datos de acceso público.
2. Se efectuará un análisis filogenético para la búsqueda de residuos conservados en ChrA, seleccionando un pequeño conjunto de ellos para su análisis.
3. Se hará mutagénesis dirigida de algunos de los residuos conservados de ChrA seleccionados.
4. Se determinará la susceptibilidad a cromato de las clonas transformantes con los genes mutados de ChrA.



III. MATERIALES Y MÉTODOS

III.1 FILOGENIA MOLECULAR

Se utilizó una metodología similar a la reportada por Riveros-Rosas y cols. (2003), la cual consiste en una búsqueda exhaustiva de las proteínas relacionadas a la proteína modelo, seguida de un proceso de depuración, para así quedarse con los miembros homólogos de la proteína modelo. A este grupo se le reconstruye la filogenia mediante el uso de cuatro algoritmos filogenéticos.

III.1.A. Obtención de secuencias

Para la obtención de un conjunto de secuencias de proteínas homólogas a ChrA de *P. aeruginosa*, se siguieron dos diferentes vías: búsqueda por palabra y por similitud. Para el primer caso se efectuó una búsqueda exhaustiva en las bases de datos de proteínas anotadas en el Swiss-Prot, NCBI, COG (NCBI), Pfam (Inst. Sanger) y SYSTER (Inst. Max Planck), buscando todas las proteínas que tuvieran en su descripción alguna referencia hacia su funcionalidad como transportador de cromato.

En el caso del análisis por similitud se hizo una búsqueda de homólogos, mediante los programas BLASTP y PSI-BLAST (Altschul y cols., 1997), utilizando la secuencia de aminoácidos de ChrA de *P. aeruginosa* (no. de acceso P14285). En el caso de Blastp se utilizó la base de datos nr (proteínas no redundantes), no se usó filtro de baja complejidad, y se filtraron las secuencias con un número E menor a 1×10^{-5} ; los parámetros de penalización de huecos (gaps) y la matriz usada fueron los de uso por omisión. Cuando se buscó con PSI-Blastp, se utilizaron los mismos parámetros, con el complemento de iterar hasta convergencia, lo cual sucedió en la quinta iteración.

III.1.B. Depuración de las secuencias

Para este fin se hizo un alineamiento múltiple de todas las proteínas obtenidas, utilizando el programa clustalW (Thompson y cols., 1994) con los parámetros por omisión, primeramente por cada grupo de donde se obtuvieron y posteriormente de manera general. El alineamiento genera un árbol guía donde fácilmente se pueden ver las proteínas redundantes y los fragmentos; se utilizó TreeView (Page, 1996) e HyperTree¹ para la visualización de estos árboles. Una vez que se seleccionaron las proteínas que posiblemente fueran redundantes y las que fueran fragmentos, se procedió a hacer alineamientos por pares de cada grupo de proteínas sospechosas individualmente, para

¹ Programa basado en Java. Para obtenerlo visitar: <http://kinase.com/tools/HyperTree.html>



corroborar que efectivamente se tratara de proteínas repetidas y fragmentos. Estos pasos se repitieron cada vez que se eliminó alguna secuencia, hasta que no hubo sospecha de tener más proteínas redundantes y fragmentos. Posteriormente se formó un solo grupo final de proteínas, agrupando todas las secuencias limpias de las fuentes. Se siguió un último paso de limpieza para estar completamente seguros de no tener fragmentos ni proteínas redundantes.

III.1.C. Análisis de dominios

Para el análisis de dominios, se procedió a hacer un alineamiento múltiple preliminar para observar la agrupación de las proteínas menores y mayores y hacer grupos preliminares. Para dicho fin se utilizó el programa clustalW (ver 1.83). El análisis de los grupos preliminares se realizó corriendo Blastp local para cada una de las secuencias identificadas como miembros de la familia CHR, utilizando las bases de datos: chra_135 (135 secuencias), chra_77 (conteniendo las secuencias de las proteínas de cadena larga) y chra_58 (que contiene las secuencias de las proteínas de cadena corta), que se generaron como resultado de este mismo trabajo.

III.1.D. Alineamiento múltiple

Se utilizó el programa clustalW. El refinamiento del alineamiento se realizó manualmente, usando como base los resultados de los alineamientos obtenidos con Blastp y comparando secuencia por secuencia. Se usó como editor de secuencias el programa BioEdit (Hall, 1999).

III.1.E. Construcción de los árboles filogenéticos

Para la construcción del árbol filogenético se utilizó el programa MEGA2 (Kumar y cols., 2001). Se siguieron cuatro algoritmos distintos, abarcando los llamados métodos de matriz de peso (MMP) y métodos basados en características (MBC), los cuales fueron:

1. Unweighted Pair Group Method With Arithmetic Mean (UPGMA)
2. Vecino Más Cercano (Nj -- Neighbor-Joining)
3. Mínima Evolución (ME)
4. Máxima Parsimonia (MP)

En la construcción de los árboles, se utilizó la corrección de Poisson como modelo evolutivo y delección pareada para el alineamiento. Como prueba estadística se efectuó un



análisis de Bootstrap de 1000 repeticiones para cada árbol. Se verificó la taxonomía del organismo de donde provenía cada secuencia para corroborar los datos filogenéticos.

Para colocar la raíz del árbol se utilizó como grupo exterior a las proteínas de *Burkholderia cepacia* (ZP-00219994) y de *Burkholderia fungorum* (ZP-00278165), pertenecientes a la familia de intercambiadores de disulfuro/tiol (COG0526). Estas proteínas se encontraron en la búsqueda con Psi-Blastp, aunque con una similitud baja (10 %), y por lo tanto un número E alto ($E > 0.01$), por lo que es razonable usarlas como grupo externo (outgroup).

III.1.F. Identificación de aminoácidos conservados

Se siguieron tres rutas para la identificación de aminoácidos conservados. La primera y más estricta fue seleccionando aquellos aminoácidos que tuvieran un porcentaje de conservación elevado (mayor del 80%) en todas las secuencias. La segunda ruta fue mediante la selección de los aminoácidos más conservados en el grupo filogenético de ChrA de *P. aeruginosa*. Para este objetivo, se dibujó el árbol consenso de los cuatro métodos y se obtuvo el grupo de secuencias donde se agrupa ChrA; se hizo un alineamiento múltiple y se obtuvieron los aminoácidos más conservados. La tercera vía fue mediante el alineamiento y obtención de los aminoácidos conservados dentro de las proteínas mayores sin cortar.

III.2 MÉTODOS GENERALES

III.2.A. Reactivos y medios de cultivo

Los medios de cultivo o sus componentes se adquirieron de los laboratorios Bioxon de México S. A., Merck-México S.A. o Sigma Chemical Corporation y fueron:

- Caldo Luria Bertani (LB): NaCl 1%, peptona de caseína 1% y extracto de levadura 0.5%.
- Agar Luria Bertani (AL): Al caldo Luria se le agregó 1.5% de agar bacteriológico.
- Caldo nutritivo (CN).
- Medio SOC: Triptona 2%, extracto de levadura 0.5%, NaCl 0.0584%, glucosa 1%, $MgCl_2$ 1% y $MgSO_4$ 1%.
- K_2CrO_4 y $MgSO_4$ de los laboratorios Merck.



III.2.B. Cepas empleadas

Las cepas de *E. coli* empleadas como receptoras en la transformación fueron las siguientes:

- JM101 *supE*, *thi-1* ? (*lac-proAB*) [*F'*, *traD36*, *proAB*, *laq1^qZ*?M15] (Yanisch-Perron y cols., 1985).
- XL1-Blue *end A1*, *gyrA96*, *thi-1*, *hsdR17* *supE44* *recA1* *lac* [*F'* *proAB* *laq1^qZ*?M15 Tn10 (Tc)] (Stratagene).

La cepa PAO1 de *P. aeruginosa* fue empleada como receptora de transformación (Holloway, 1969).

III.2.C. Plásmidos empleados

pGEMTChrAHis: Este plásmido surge de la clonación del fragmento pTRC-*chrAHis* en el vector pGEMT (Promega) (Figura 6A), por lo que cuenta con las características de este vector. Además tiene varios sitios de restricción para la enzima *DpnI*.

pUCP20: Este plásmido es un vector binario que cuenta con un origen de replicación para *E. coli* y otro para *Pseudomonas* (West y cols., 1994) (Figura 6B). Tiene además un gen de resistencia a ampicilina y un segmento de DNA que codifica la región amino terminal de la β -galactosidasa (*LacZ*), que permite complementar la forma inactiva de *LacZ* de las cepas estándar de *E. coli* (sistema de α -complementación). De esta manera, las bacterias que tengan el plásmido con el gen de interés clonado desarrollaran colonias blancas *Lac⁻*, contra colonias azules *Lac⁺*, en presencia de un sustrato cromogénico (Xgal) cuando el vector se religue. Se empleó para la clonación del gen *chrAHis* silvestre y los genes de las diferentes mutantes.

pUChrA: Plásmido que surge de la clonación del fragmento pTRC*chrAHis* en el vector pUCP20, por lo que cuenta con las características de este vector (Aguilar, 2005) (Figura 6C).

III.2.D. Aislamiento y manipulación de DNA

III.2.D.i. Aislamiento de DNA plasmídico

El aislamiento de DNA plasmídico se realizó empleando el método de lisis alcalina, adaptado de Birnboim y Doly (1979) e Ish-Horowicz y Burke (1981), de acuerdo al siguiente protocolo:



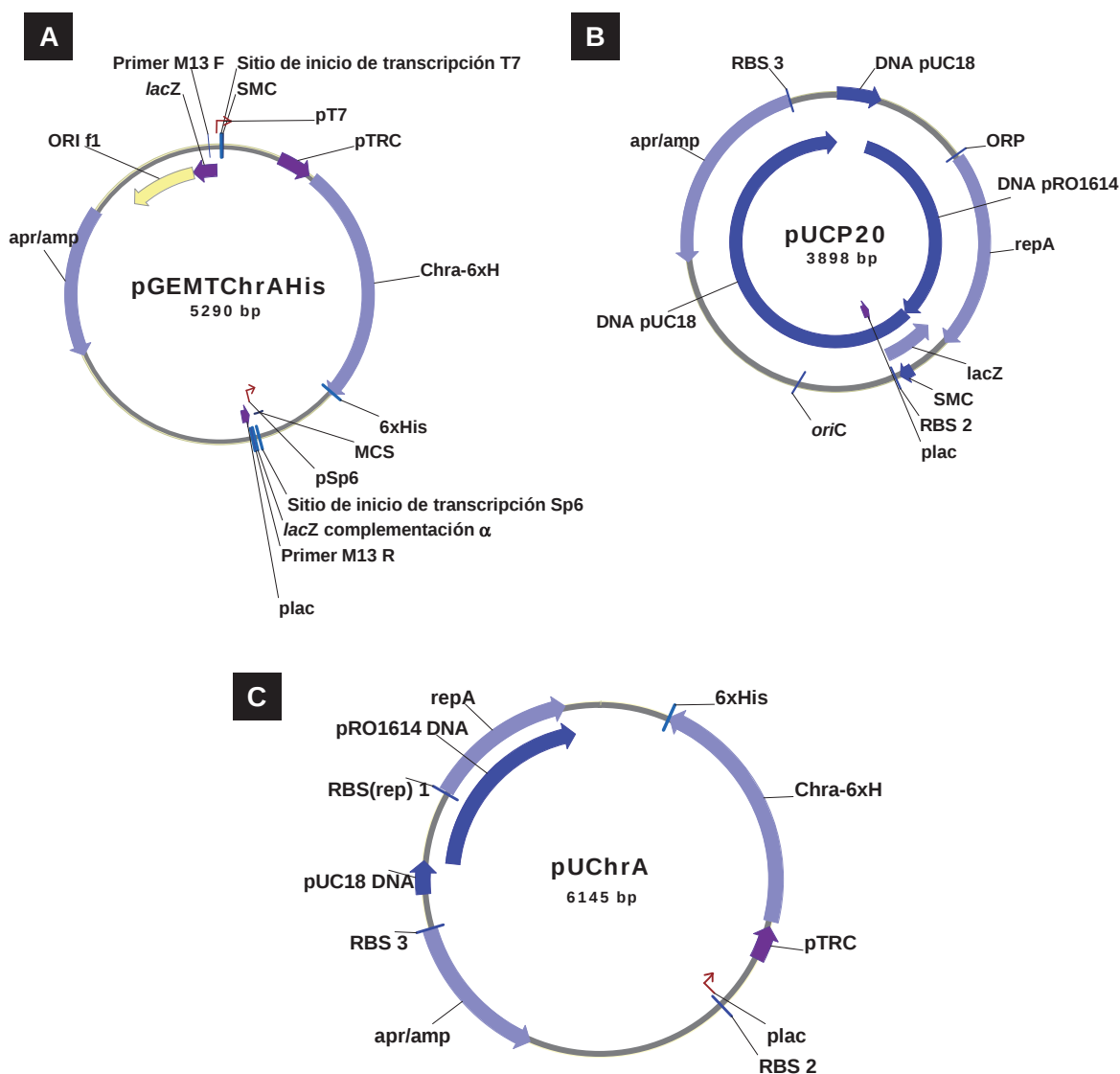


Figura 6 – Esquema de los plásmidos usados en este estudio. A) pGEMTChrAHIS, cuenta con el gen *chrA* bajo la regulación del promotor pTRC (Aguilar, 2005). B) pUCP20, vector binario que cuenta con orígenes de replicación para *Pseudomonas* (ORP) y *E. coli* (*oriC*) (West y cols., 1994). C) pUChrA, vector binario con el gen *chrA* bajo la regulación del promotor pTRC. Los tres cuentan con genes de resistencia a ampicilina (apr/amp) y sitios múltiples de clonación (SMC) (Aguilar, 2005).

En 4 ml de caldo Luria se creció un cultivo bacteriano durante 18-20 h a 37°C con agitación constante, posteriormente se distribuyó en tubos de microcentrifuga de 1.5 ml. Se centrifugaron durante 2 min a 10 000 g a temperatura ambiente. Posteriormente la pastilla se suspendió en 100 µl de la solución STE (sacarosa 50 mM, Tris-HCl 25 mM pH 8.0 y EDTA 10 mM) y se le adicionaron 20 µl de una solución de lisozima (20 mg/ml), se mezcló ligeramente y se incubó a 37°C durante 5 min. Después a la suspensión formada se le agregaron 200 µl de la solución II recién preparada [NaOH 0.2 N y SDS (dodecilsulfato de sodio)1%], se agitó suavemente y se incubó durante 5 min en hielo. Transcurrido este tiempo se le adicionaron 150 µl de una solución de acetato de potasio 5 M (acetato de potasio 5 M y ácido acético glacial pH 4.8), se mezcló ligeramente y se incubó 10 min en hielo. Se centrifugó durante 5 min a temperatura ambiente. Se le adicionaron 500 µl de la mezcla 1:1 fenol-cloroformo y se mezcló moderadamente. Se centrifugó durante 5 min y al sobrenadante se le agregaron 500 µl de cloroformo, se mezcló y se centrifugó durante 3 min. La fase acuosa superior se recuperó y se le adicionó 1 ml de etanol absoluto frío y se incubó durante 10 min a -20°C. Se centrifugó durante 15 min a temperatura ambiente. Por último el sedimento se lavó dos veces con 1 ml de etanol al 70% centrifugando durante 5 min entre cada lavado. Se recuperó el sedimento y se secó a temperatura ambiente para posteriormente resuspender la pastilla de DNA obtenida en 50 µl de agua desionizada estéril. Estas muestras se almacenaron a -20°C o fueron sometidas a los corrimientos electroforéticos en geles de agarosa.

III.2.D.ii. Electroforesis en geles de agarosa

El DNA aislado se sometió a un corrimiento electroforético, para ello se preparó un gel de agarosa (Sigma) al 1% (p/v) en amortiguador TAE (Tris-acetato 0.04 M y EDTA 0.001 M). El gel se depositó en una cámara de electroforesis horizontal, con el mismo amortiguador. Una muestra del DNA se mezcló con 2 µl de amortiguador de carga [azul de bromofenol al 0.05%, EDTA 0.1 M pH 8.0, SDS 0.5%, sacarosa 40%] y se descargó en los orificios del gel. La cámara se conectó a una fuente de poder (E-C Apparatus Corporation modelo EC452) y se sometió a un voltaje constante de 120 voltios durante 1 h. Posteriormente el gel se tiñó con una solución de bromuro de etidio al 0.01% durante 10 min. Las bandas de DNA teñidas se observaron en un transiluminador de luz UV de onda corta (UV-Products Inc.).

Como marcador de tamaño molecular de DNA lineal se utilizó el DNA del fago lambda digerido con las endonucleasas *EcoRI* y/o *HindIII*.



III.2.D.iii. Aislamiento del DNA de geles de agarosa

Para purificar el DNA empleado en las reacciones de ligación se empleó el kit comercial “Wizard SV Gel and PCR Clean-Up System” de Promega.

La banda de interés se cortó del gel de agarosa y se colocó en un tubo Eppendorf adicionando 10 µl de solución de unión a membrana por cada 100 mg de gel. Se mezcló bien en el agitador y se incubó a 65°C hasta disolver completamente el gel. Se insertó una minicolumna SV en un tubo de recolección, se transfirió la solución disuelta en la columna y se incubó 1 min a temperatura ambiente. Posteriormente se centrifugó 1 min a 7 000 g. Se lavó dos veces, primero con 700 µl de solución de lavado de membrana y centrifugando 1 min a 7 000 g y después con 500 µl con solución de lavado centrifugando ahora 5 min 7 000 g. La elución del DNA se hizo adicionando agua grado HPLC y centrifugando 1 min a 7 000 g. Una alícuota del DNA se corrió en un gel de agarosa al 1% para determinar la concentración del DNA recuperado.

III.2.D.iv. Tratamientos enzimáticos del DNA

III.2.D.iv.1. Restricción con endonucleasas

Los DNA plasmídicos se digirieron con endonucleasas empleando una unidad de enzima de restricción (Promega o Invitrogen) por microgramo de DNA e incubando a 37°C por un mínimo de 4 h o toda la noche. Después de este tiempo las muestras se sometieron a un corrimiento electroforético en un gel de agarosa al 1% para determinar el tamaño de los fragmentos obtenidos.

III.2.D.iv.2. Ligaciones

Después de la recuperación y purificación de los fragmentos de DNA del gel de agarosa, estos fueron ligados utilizando la enzima DNA ligasa del fago T4 (Promega). Las ligaciones se dejaron como mínimo toda la noche a temperatura ambiente y las proporciones inserto-vector variaron según el experimento realizado. Se emplearon de 3-5 µl de la mezcla de ligación en la transformación.

Para las ligaciones en el vector pGEMT, aproximadamente 100 ng de los fragmentos de PCR se mezclaron con 50 ng del vector. Estas ligaciones se incubaron como mínimo 1 h a temperatura ambiente o toda la noche a 4°C. Para la ligación se usó la ligasa provista con el vector. De estas mezclas se emplearon de 1-3 µl en la transformación.



III.2.E. Métodos de transformación bacteriana

III.2.E.i. Transformación de *Escherichia coli* por electroporación

III.2.E.i.1. Preparación de células competentes

De un cultivo de la cepa de *E. coli* crecido durante toda la noche se tomaron 4 ml para inocular 250 ml de CL contenidos en un matraz de 1L y se incubó durante 3 h a 37°C con agitación constante hasta obtener una absorbencia de 0.6 U.D.O. a 590 nm. Posteriormente se incubó el cultivo en hielo durante 20 min y se distribuyeron alícuotas en tubos para centrifugar durante 10 min a 5 000 g a 4°C. La pastilla se lavó dos veces por resuspensión con agua estéril fría y se centrifugó entre cada lavado con las condiciones antes descritas. Después del último lavado la pastilla se resuspendió en agua estéril más glicerol al 20%. Posteriormente se distribuyó en alícuotas de 0.2 ml en tubos de microcentrífuga de 1.5 ml. Estas células se utilizaron para la transformación o bien fueron almacenadas a -80°C.

III.2.E.i.2. Transformación

Las transformaciones de las diferentes cepas de *E. coli* empleadas se realizaron en un electroporador 2510 de Eppendorf. Las condiciones usadas variaron según el manual del proveedor en función de la cepa y la fuente de DNA empleada.

III.2.E.ii. Transformación de *Escherichia coli* por choque térmico

III.2.E.ii.1. Preparación de células competentes

Se preparó un preinóculo de la cepa de *E. coli* en 3 ml de medio LB y se incubó toda la noche a 37°C con agitación constante. Al siguiente día el preinóculo se vació a un matraz con 500 ml de medio LB y se incubó a 37°C hasta una absorbencia de 0.45-0.55 U.D.O. a 560 nm. El crecimiento se detuvo incubando en hielo por 5 min, se centrifugó a 2 500 g durante 10 min a 4°C. La pastilla se resuspendió en 25 ml de TFb1 (acetato de potasio 30 mM, KCl 100 mM, CaCl₂ 10 mM, MnCl₂ 50 mM y glicerol 15%, se ajustó el pH a 5.8 con ácido acético 0.2 M y se esterilizó por filtración en un filtro de 0.22 µm) frío, se incubó en hielo por 20 min y posteriormente se centrifugó a 6 000 rpm por 10 min. La pastilla se resuspendió en 2 ml de TFb2 frío [(MOPS (ácido 4-morfolino propano sulfónico) 75 mM, KCl 10 mM, CaCl₂ 100 mM y glicerol 15%], se ajustó el pH a 6.5 con KOH y se esterilizó por filtración en un filtro de 0.22 µm. Posteriormente se hicieron alícuotas de 100 µl de la suspensión bacteriana y se almacenaron a -80°C.

III.2.E.ii.2. Transformación

Se mezcló el DNA con 100 µl de la suspensión de células bacterianas competentes y se incubó en hielo 20 min. Después de este tiempo se incubaron a 42°C



en un baño de agua por 45 seg y los tubos se transfirieron al hielo incubándose por 3 min. Se les adicionó 1 ml de caldo LB precalentado, se incubaron a 37°C por un máximo de 1 h, se concentró el cultivo resuspendiendo en 100 µl y se plaqueó en cajas de LB con ampicilina.

III.2.E.iii. Electroporación de *P. aeruginosa* o *E. coli* sembradas recientemente en placa

Para la electroporación de *P. aeruginosa* se empleó el método de electroporación de células recién cosechadas de placa (Enderle y Farwell, 1998), para lo cual se sembró en una placa de LB la cepa de *P. aeruginosa* PAO1 y se incubó 16 h a 37°C, posteriormente la caja se mantuvo a temperatura ambiente. Se llenó un tubo de 1.5 ml con 500 µl de agua desionizada, estéril y fría y después se transfirieron aproximadamente 3 mg de células de la caja al tubo, usando un asa estéril. Los tubos se mezclaron moderadamente en el agitador hasta tener una mezcla homogénea y la suspensión celular se centrifugó 1 min a 14 000 rpm desechando el sobrenadante. Posteriormente las células se resuspendieron con la pipeta en 500 µl de agua desionizada estéril y fría y se centrifugaron nuevamente 2 min desechando el sobrenadante. Las células se resuspendieron en agua estéril y fría en un volumen de aproximadamente 500 µl, se mantuvieron en hielo y se usaron inmediatamente para la electroporación.

El DNA usado para la electroporación se diluyó en agua destilada estéril a una concentración final entre 5-50 ng de plásmido en 5-10 µl de volumen final. Después el DNA se adicionó a 100 µl de células y se mezcló moderadamente, la mezcla se transfirió a una celda de electroporación en hielo y se les dió un pulso de 1.8 ó 2.5 KV. Inmediatamente después se le adicionó 1 µl de medio SOC a la celda, se mezcló suavemente con la pipeta, se transfirió a un tubo con 1 ml de LB y se incubó de 1.5 a 4 h a 37°C con agitación constante. Transcurrido este tiempo se plaquearon 100 a 200 µl de cultivo en cajas de agar LB con carbenilicina (400 µg/ml).

III.3.MUTAGÉNESIS DIRIGIDA

III.3.A. Mutagénesis dirigida basada en PCR y digestión con *DpnI*

El método de mutagénesis dirigida empleado fue el reportado por Fisher & Pei (1997), que se basa en: (a) la inserción de la mutación mediante un juego de oligonucleótidos (a los que llamaremos oligos) complementarios, (b) la alta fidelidad de la enzima *Pfu* DNA-polimerasa, y (c) la digestión del DNA molde con la enzima de restricción *DpnI*. Para este método se empleó el plásmido pGEMT-ChrAHis como molde (Figura 7).



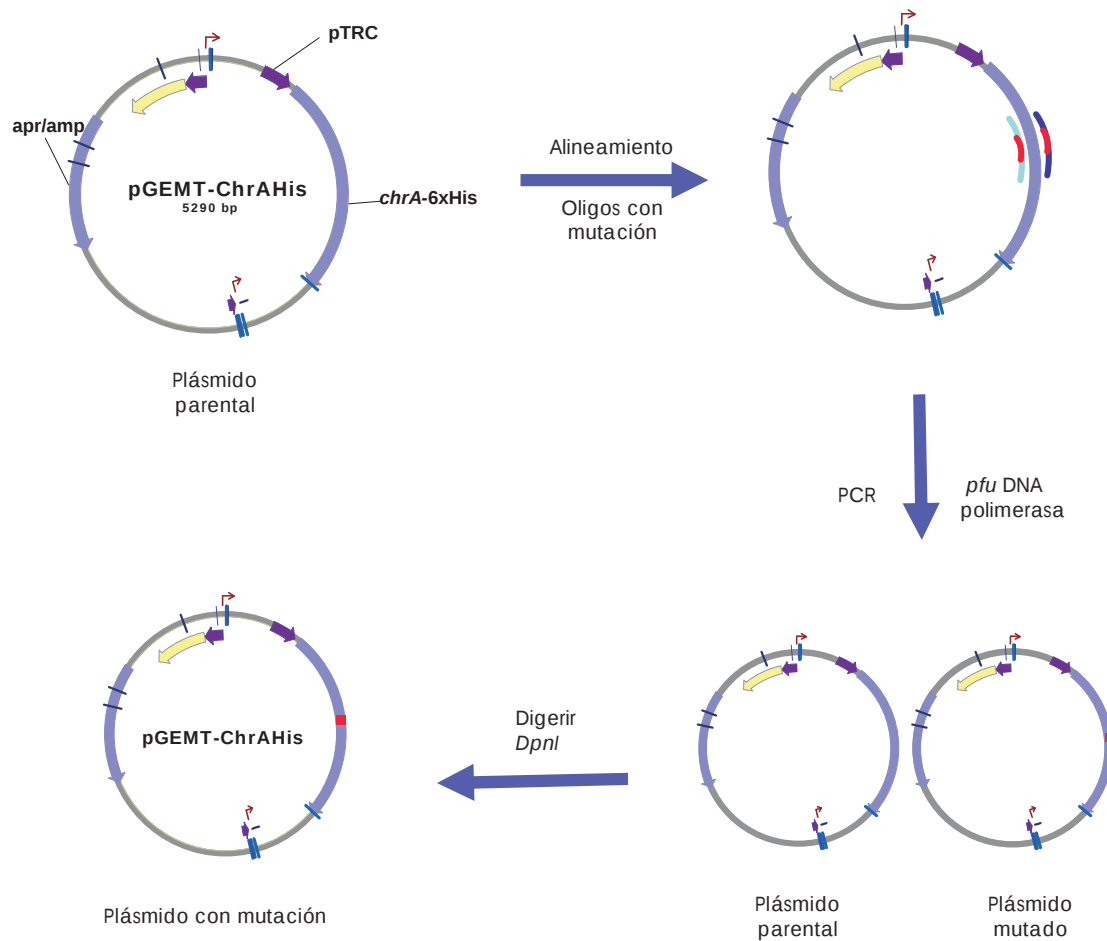


Figura 7 – Diagrama de la mutagénesis sitio dirigida basada en PCR y digestión con *DpnI*. Se muestra el vector original pGEMTChrAHis. En el texto se explica el método y es detallado por Aguilar (2005).

El fundamento del método es la utilización de DNA plasmídico como molde durante la PCR. Este DNA al proceder de una bacteria se encuentra metilado. Después de la reacción se obtendrán dos poblaciones de plásmidos, el plásmido parental no mutagenizado metilado, y el plásmido mutagenizado no metilado, ya que fue generado durante la PCR. El producto de la reacción se digiere con la enzima *DpnI*, la cual reconoce sólo al plásmido metilado, y por lo tanto lo digiere, dejando sólo el plásmido mutado intacto. Es importante señalar que la eficiencia de la mutagénesis aumenta gracias al uso de la enzima *Pfu* DNA polimerasa que tiene un porcentaje de error mucho menor que la enzima tradicionalmente usada en PCR, *Taq* DNA polimerasa.

El plásmido mutado se introduce mediante transformación por choque térmico a la cepa XL1-BLUE, la cual cuenta con sistemas de ligación de DNA que ligan al plásmido mutado (Stratagene). Los plásmidos mutantes confieren resistencia al antibiótico, por lo que se seleccionan las colonias resistentes que crezcan en medio con ampicilina.

III.3.A.i. Diseño de oligonucleótidos mutagénicos

Los oligos se diseñaron entre 25 a 45 nucleótidos, con una T_m igual o mayor a 78°C . La T_m se calculó aplicando la siguiente formula:

$$T_m = 81.5 + 0.41 (\%GC) - 675/N; \text{ donde } N \text{ no incluye las bases eliminadas.}$$

La mutación se localizó al centro de los oligos los cuales son complementarios entre sí. El diseño se realizó de manera que tuvieran un mínimo de GC de 40% y por lo menos una G o C en los extremos. Además, los oligos no fueron fosforilados. Los oligos fueron obtenidos de Invitrogen.

Se diseñaron 12 juegos de oligos, los cuales se listan en la Tabla 2.

III.3.A.ii. Síntesis de la cadena mutada

La síntesis de la cadena mutada se hizo por medio de PCR. Se puso la siguiente reacción: DNA (50-100 ng), oligonucleótido superior e inferior (125 ng), buffer de reacción 10X (5 μl), dNTPs mix 10 mM (1 μl), 1 μl de DNA polimerasa *Pfu* (Stratagene) (2.5 U/ μl), se ajustó la reacción a un volumen final de 50 μl con agua HPLC. La enzima siempre se adicionó al final.

Las condiciones del PCR usadas fueron las siguientes: $94^{\circ}\text{C}/1 \text{ min}$ (desnaturalización inicial), $94^{\circ}\text{C}/1 \text{ min}$ (desnaturalización), $62^{\circ}\text{C}/1 \text{ min}$ (alineamiento), $72^{\circ}\text{C}/7.5 \text{ min}$ (extensión) y $72^{\circ}\text{C}/7.5 \text{ min}$ (extensión final), las muestras se sometieron a 16 ciclos de amplificación.



Tabla 2.– Oligos diseñados para la mutagénesis dirigida de ChrA.

ID	Oligonucleótido 5'	Oligonucleótido 3'
G100	GCAAAGGGGCGGCTGGGGGCGTGTGGCGGG	CCCGCCAGCACGCCCCCAGCCGCCCTTTGC
G100→A	GCAAAGGGGCGGCTGGCCGGCGTGTGGCGGG	CCCGCCAGCACGCCGGCCAGCCGCCCTTTGC
G100→S	GCAAAGGGGCGGCTGAGCGGCGTGTGGCGGG	CCCGCCAGCACGCCGCTCAGCCGCCCTTTGC
G100→L	GCAAAGGGGCGGCTGCTGGGCGTGTGGCGGG	CCCGCCAGCACGCCCAGCAGCCGCCCTTTGC
G92	GAAATATGCGTTTCATTTGGGCATCCGAGCAAAGGG	CCCTTTGCTCGGATGCCCAAATGAACGCATATTTTC
G92→A	GAAATATGCGTTTCATTTGGCCATCCGAGCAAAGGG	CCCTTTGCTCGGATGGCCCAAATGAACGCATATTTTC
G92→S	GAAATATGCGTTTCATTTGAGCATCCGAGCAAAGGG	CCCTTTGCTCGGATGCTCAAATGAACGCATATTTTC
G92→L	GAAATATGCGTTTCATTTGCTGATCCGAGCAAAGGG	CCCTTTGCTCGGATCAGCAAATGAACGCATATTTTC
G45	GGCTTGCTCGCATGGGGCGGGCCTGTGGCTCAGATC	GATCTGAGCCACAGGCCCGCCCCATGCGAGCAAGCC
G45→A	GGCTTGCTCGCATGGGGCGCCCTGTGGCTCAGATC	GATCTGAGCCACAGGGGCGCCCCATGCGAGCAAGCC
G45→S	GGCTTGCTCGCATGGGGCAGCCCTGTGGCTCAGATC	GATCTGAGCCACAGGGCTGCCCCATGCGAGCAAGCC
G45→L	GGCTTGCTCGCATGGGGCCTGCCTGTGGCTCAGATC	GATCTGAGCCACAGGCAGGCCCCATGCGAGCAAGCC
G44	TTCGGCTTGCTCGCATGGGGCGGGCCTGTGGCTCAG	CTGAGCCACAGGCCCGCCCCATGCGAGCAAGCCGAA
G44→A	TTCGGCTTGCTCGCATGGGCCGGGCCTGTGGCTCAG	CTGAGCCACAGGCCCGGCCCATGCGAGCAAGCCGAA
G44→S	TTCGGCTTGCTCGCATGGAGCGGGCCTGTGGCTCAG	CTGAGCCACAGGCCCGCTCCATGCGAGCAAGCCGAA
G44→L	TTCGGCTTGCTCGCATGGCTGGGGCCTGTGGCTCAG	CTGAGCCACAGGCCCCAGCCATGCGAGCAAGCCGAA
W63	GTGGACGAGGAACGCTGGATCTCCAGCAAACGC	GCGTTTGCTGGAGATCCAGCGTTCTCGTCCAC
W63→A	GTGGACGAGGAACGCGCCATCTCCAGCAAACGC	GCGTTTGCTGGAGATGGCGCGTTCTCGTCCAC
W63→Y	GTGGACGAGGAACGCTACATCTCCAGCAAACGC	GCGTTTGCTGGAGATGTAGCGTTCTCGTCCAC
W63→F	GTGGACGAGGAACGCTTCATCTCCAGCAAACGC	GCGTTTGCTGGAGATGAAGCGTTCTCGTCCAC

Se subrayan los nucleótidos del codón que se va a mutar.

ID- Nombre del oligo, el cual representa la mutación.



III.3.A.iii. Digestión con *DpnI*

Después de la amplificación de la cadena mutada, la reacción se digirió con la enzima de restricción *DpnI* (New England Biolabs) para la eliminación del DNA molde. Se puso la siguiente reacción: DNA (17 µl), buffer 10X (2 µl) y 1 µl de *DpnI* (10-20 U/µl). La reacción se incubó toda la noche a 37°C.

III.3.A.iv. Transformación

La mezcla total de digestión se empleó para transformar células competentes de *E. coli* XL1-Blue por choque térmico (anteriormente descrito). La selección de las mutantes se hizo plaqueando en cajas de agar LB con 100 µg/ml de ampicilina. De las colonias obtenidas se seleccionaron algunas al azar y se les realizó aislamiento de DNA plasmídico por lisis alcalina para verificar la presencia del plásmido y su posterior secuenciación.

III.4. SECUENCIACIÓN DE LOS GENES MUTADOS

III.4.A. Limpieza del DNA con fenol:cloroformo y bromuro de etidio

Los plásmidos mutagenizados se sometieron a un proceso de limpieza para eliminar proteínas que puedan interferir con la secuenciación. Se aisló DNA de los plásmidos mutagenizados, se resuspendió en 200 µl de agua destilada, se adicionó 140 µl de acetato de amonio 7.5 M y se mezcló suavemente. Se adicionaron 5 ml de bromuro de etidio (10 mg/ml) y se agitó suavemente hasta que se incorporó. Se agregaron 400 µl de fenol:cloroformo:alcohol isoamílico (25:24:1) y se mezcló hasta que se formó una emulsión. Se procedió a centrifugar durante 5 min a 10 500 g. Se recuperó la fase acuosa, la cual se transfirió a un tubo nuevo. Se adicionaron 1000 µl de etanol absoluto, se mezcló suavemente e incubó en el congelador por un periodo que varió de 1 hora a toda la noche. Se centrifugó 10 min a 10 500 g y se eliminó el sobrenadante. Se lavó la pastilla con etanol al 70% y se centrifugó por 1 min a 10 500 g y retire el sobrenadante. Se repitió la operación de 2-4 veces. Se secó la pastilla a 37° C durante 7 min. El DNA se resuspendió en agua (20-40 µl) y se almacenó a -4°C.

III.4.B. Secuenciación

El DNA se mando secuenciar al laboratorio de secuenciación del CINVESTAV unidad Irapuato. Para la secuenciación se utilizó un oligo inverso (oligo 30) que comienza en la glutamina 125 de ChrA y cuya secuencia es: CGA ATC TAG ACT GGA AGT ACA ACC AGG AC.



III.5. TRANSFERENCIA DE LOS GENES MUTADOS A *P. aeruginosa*

III.5.A. Escisión y recuperación de los genes mutados

La subclonación (Figura 8) se efectuó cortando los plásmidos pGEMTChrAHis mutados con las endonucleasas *Xba*I y *Sph*I, para escindir el promotor pTRC y el gen *chrA* fusionado con las histidinas de dichos plásmidos. Se procedió a cortar el plásmido pUCP20 de la misma forma. Las restricciones se corrieron de manera independiente en geles de agarosa, y se procedió a recuperar las bandas de los fragmentos de DNA, los genes mutados (fragmento de 2.3 Kbases) y el vector linealizado (3 Kbases), Se procedió a ligar dichos fragmentos con la enzima T4 DNA ligasa. Se transformó por electroporación la cepa de *E. coli* JM101 con la mezcla de ligación antes mencionada.

III.5.B. Ligación y transferencia de los genes mutados a *P. aeruginosa*

Se ligaron los fragmentos del vector pUCP20 y los genes mutados durante toda la noche. Se utilizó 1 µl de la mezcla de ligación para transformar *E. coli* JM101 mediante electroporación. Se seleccionaron 10 colonias resistentes a ampicilina, a las cuales se les extrajo el DNA plasmídico el cual se verificó por restricción con las endonucleasas *Bgl*II, *Sph*I y *Xba*I. Con el DNA que dio el patrón correcto se transformó *P. aeruginosa* PAO1 recién crecida en placa.

III.6. PRUEBAS DE SUSCEPTIBILIDAD A CROMATO

Se prepararon tubos con 4 ml de CN con concentraciones variables de cromato. A cada tubo se le agregaron 40 µl de un cultivo de las cepas con los plásmidos mutantes, crecido en CN durante 18-20 h a 37°C con agitación constante. Como control de crecimiento bacteriano se incluyó un tubo que no contenía cromato. Los tubos inoculados se incubaron por un periodo de 18-20 h a 37°C con agitación constante. La susceptibilidad se midió en función de la absorbencia de los cultivos a 590 nm en un espectrofotómetro Spectronic 21, Milton Roy. Se usó como control sensible una cepa de *P. aeruginosa* PAO1 que contiene al plásmido pUCP20, el cual no lleva el gen *chrA*; como control resistente se usó la cepa de *P. aeruginosa* que se transformó con el plásmido pUChrA, el cual lleva el gen *chrA* sin mutar. Se efectuaron 6 repeticiones de cada prueba por cada mutante.



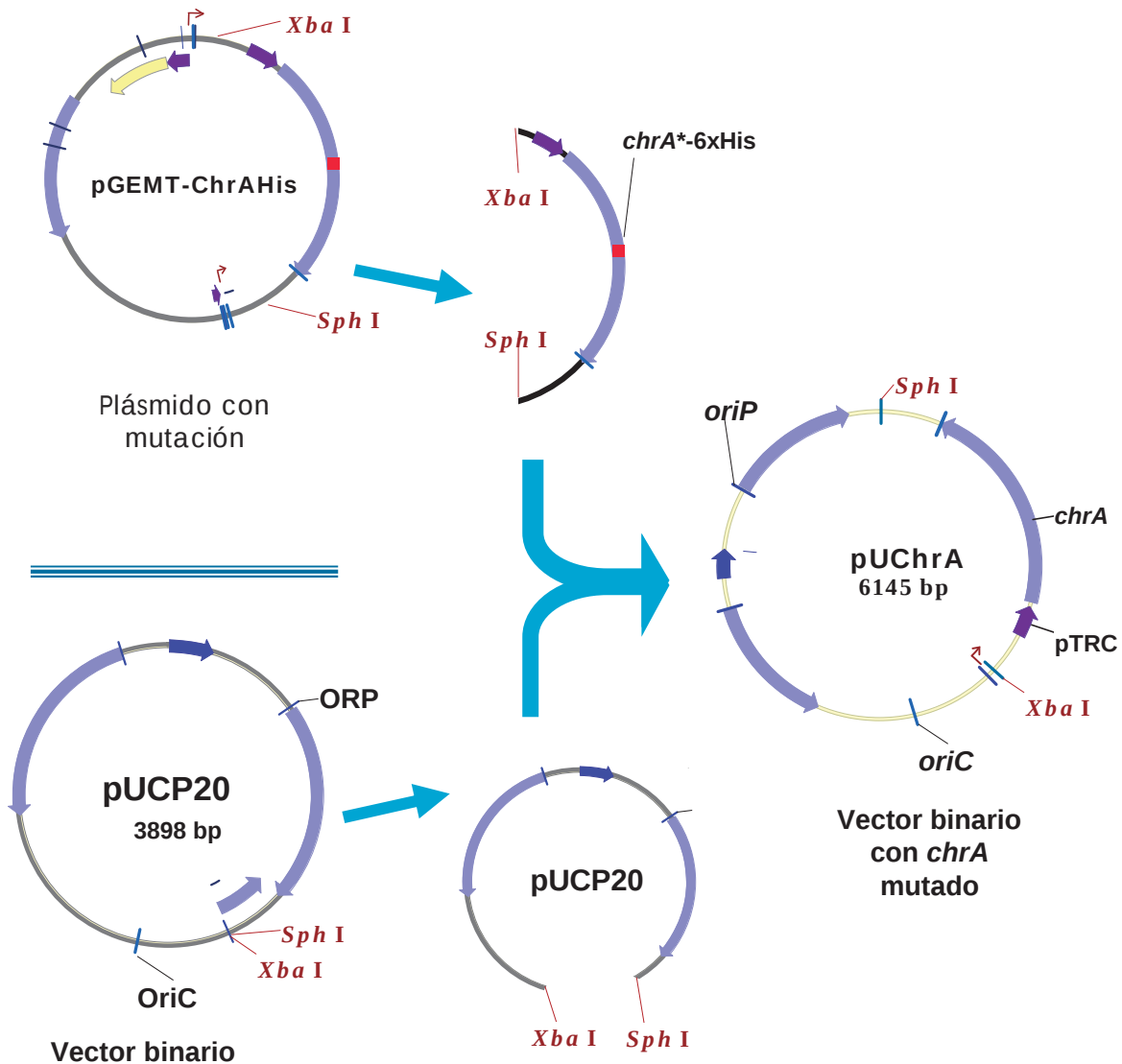


Figura 8 – Esquema de la subclonación de *chrAHis* en pUCP20. El gen *chrAHis* fue liberado con las enzimas de restricción *Xba*I y *Sph*I y se subclonó en el vector pUCP20, el cual se linearizó con las mismas enzimas. El plásmido resultante de la ligación se le denominó pUChrA.

IV. RESULTADOS

VI.1. FILOGENIA MOLECULAR

VI.1.A. Obtención e identificación de secuencias de la familia CHR

El primer paso a seguir para generar la filogenia de una proteína es obtener un juego de datos de trabajo, que en este caso consistió en secuencias de proteínas homólogas. Para esto se siguieron dos estrategias: la búsqueda por palabra y la búsqueda por similitud, de las cuales se obtuvieron 813 secuencias en total; en la Tabla 3, se resume el resultado de cada búsqueda de manera individual.

La información se extrajo en formato *geneprotein*, es decir, archivos con la información total sobre cada secuencia proteica (autores, gen, números de acceso, secuencia, bibliografía asociada, organismo de donde proviene, taxonomía del organismo).

Para el auxilio en los pasos de obtención de información y limpieza, se generó el script de PERL, *extract1.pl*, el cual a partir de los archivos *geneprotein* (.gp) extrae la siguiente información relevante sobre cada secuencia: número de acceso, secuencia fasta, número de aminoácidos de la proteína, nombre del gen, organismo, taxonomía del organismo, bibliografía asociada.

Cada una de las secuencias se identificó con el siguiente formato: cinco letras iniciales, número de acceso de la proteína y número de aminoácidos. De las cinco letras iniciales, las tres primeras corresponden al género del organismo de donde viene la secuencia y las dos siguientes corresponden a la especie, así, por ejemplo, la secuencia de ChrA de *P. aeruginosa* se identificó como PSEAE-P14285-416_aa.

Una vez obtenidas las secuencias, se procedió a hacer una limpieza de las mismas, es decir, si bien se buscaron en bases de datos de proteínas no redundantes, puede haber errores durante el curado de las bases de datos, e incluir proteínas redundantes y fragmentos de proteínas; por esta razón se siguió un protocolo de limpieza de secuencias repetidas y fragmentos de proteínas, el cual consistió en comparar todas las secuencias de cada fuente mediante un alineamiento múltiple, utilizando para ello el programa clustalW. Posteriormente se seleccionaron las secuencias que posiblemente sean repetidas y las que sean fragmentos, y se verificaron, secuencia por secuencia seleccionada, haciendo un alineamiento por pares. Si en este alineamiento hubo completa identidad entre las secuencias, ya sea en su totalidad o en un fragmento, y pertenecen al mismo organismo, una de las dos secuencias se eliminó.

Después de tres rondas de limpieza se obtuvieron 135 secuencias en total: 77



Tabla 3- Resultados de la búsqueda de secuencias de transportadores de cromato y proteínas similares a ChrA de *P. aeruginosa*

Tipo de búsqueda	Fuente	Número de secuencias	Grupo	Descripción
Palabra	NCBI	76		Base de datos de proteínas no redundantes
	SWISS-PROT	47		Base de datos de proteínas no redundantes
	COG	214	COG2059	Base de datos de familias de proteínas
	PFAM	213	Pfam02417	Base de datos de familias de proteínas
	SYSTEMS	47	152595	Base de datos de familias de proteínas
Similitud	BLAST	77		Búsqueda de homólogos
	PSI-BLAST	139		Búsqueda iterativa de homólogos



secuencias de cadena larga (CHR) o bidominio (mayores de 250 aminoácidos), y 58 secuencias de cadena corta (SCHR) o monodominio (menores de 250 aminoácidos). Se encontraron secuencias similares a ChrA en 81 organismos, comprendiendo especies de los tres dominios, *Archaea*, *Bacteria*, y *Eukaryota*. Las 135 secuencias del grupo de trabajo se enlistan en el Anexo I.

En el caso de las bacterias, se identificaron homólogos de ChrA en 10 clases diferentes de bacterias, que abarcan diversas morfologías, como bacilos, cocos y espiroquetas. No es de extrañar que proteínas de la familia CHR se encontraran en 73 proteobacterias distintas, ya que la proteína modelo ChrA proviene de una proteobacteria (Cervantes y cols., 1990). Sin embargo, aunque se encontraron en γ -*Proteobacterias*, no se identificó ninguna proteína similar en miembros de la familia *Enterobacteriaceae*, lo que quizá se relacione con los reportes que señalan que la proteína ChrA no es funcional en *E. coli* (Cervantes y cols., 1990; Nies y cols., 1990), tal y como ocurre con otras proteínas membranales expresadas en ese modelo de estudio (Laage y Langosch, 2001).

En el caso de las arqueas, sólo se encontró la ya reportada de *Methanococcus jannaschii* (AAB98712), la cual es una CHR de cadena larga o bidominio de 402 aa, que presenta un 24% de identidad con ChrA de *P. aeruginosa*. No se logró obtener ninguna otra proteína del dominio *Archaea* que tuviera similitud con ChrA, aunque buscamos mediante tBlastn (programa que busca genes relacionados a partir de una secuencia proteica) usando la secuencia de la proteína de *M. jannaschii*.

Igualmente, se encontraron proteínas similares a ChrA dentro de los genomas eucarióticos (ver Anexo I y II), en cinco especies de hongos, lo cual resulta interesante porque a la fecha no hay ningún reporte donde se mencione la presencia de homólogos de ChrA dentro de los eucariontes. Los hongos en donde se encontraron homólogos de ChrA son mohos filamentosos pertenecientes al phylum *Ascomycota*.

VI.1.B. Análisis de dominios

Las 135 secuencias se alinearon con clustalW para obtener un alineamiento preliminar, el cual sirvió de guía para analizar las familias preliminares, hacer un análisis de dominios y obtener, si los había, los aminoácidos más conservados en todas las proteínas del juego de datos.

Usando este alineamiento se procedió al análisis de las secuencias más parecidas entre sí, encontradas con base en la distribución dentro del árbol guía (producido con clustalW). Se encontraron un total de 10 grupos distintos (Figura 9), observándose que en



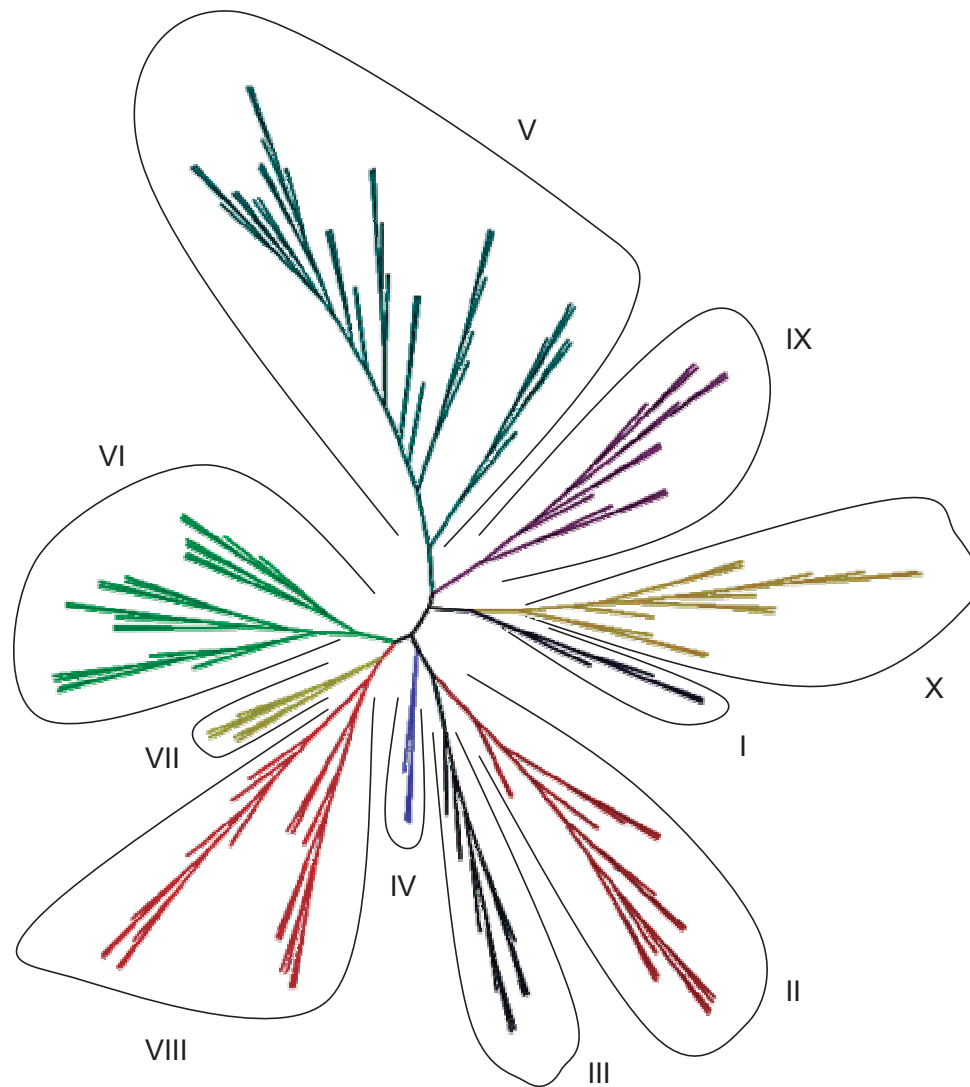


Figura 9.- Dendrograma preliminar de las 135 secuencias del grupo de trabajo. Este árbol se elaboró alineando el grupo de trabajo con clustalW, se aplicó una transformación logarítmica a la distancia entre las ramas. Se visualizó y editó con HyperTree.

general las proteínas de cadena corta (SCHR) se agruparon de manera independiente a las proteínas de cadena larga (LCHR). Cabe resaltar que este árbol es una guía preliminar, por lo que no es posible sacar una conclusión hasta que se hubieron generado los árboles filogenéticos definitivos.

El alineamiento preliminar también se utilizó para buscar los primeros aminoácidos conservados, sin embargo, dado que algunas proteínas menores se alinearon al extremo amino y otras al extremo carboxilo, no fue posible, con este alineamiento, encontrar aminoácidos con un alto porcentaje de conservación.

Estos 10 grupos de proteínas homólogas debieron de verificarse, para tener la certeza de que efectivamente formaran un grupo verdadero. Para dicho fin se efectuó un análisis de Blastp local para cada proteína, usando como base de datos las 135 proteínas encontradas en este trabajo como miembros de la familia CHR, considerando tres distintos grupos: el total de las 135 proteínas CHR más SCHR (chra_135), las proteínas de cadena larga o CHR (chra_77) y las proteínas de cadena corta o SCHR (chra_58). Se observó que en general los valores de E más significativos corresponden a los miembros del grupo a donde pertenecen las proteínas usada para la búsqueda, corroborando que el alineamiento preliminar agrupó a las proteínas más parecidas entre sí dentro de todo el juego de datos de trabajo.

Para el análisis de los resultados se utilizó el script *matriz_rep.pl*, el cual toma los resultados de Blastp de cada secuencia y construye una matriz de datos, con los valores de E de Blast de las proteínas del grupo de la búsqueda. Con este script se observó que en general los valores de E más altos para los miembros del grupo a donde pertenecen las proteínas usada para la búsqueda, corroborando que el alineamiento preliminar agrupó a las proteínas más parecidas entre sí dentro de todo el juego de datos de trabajo.

Una vez verificados los grupos preliminares se realizó un análisis de la distribución del tamaño de las proteínas pertenecientes a la familia CHR (Figura 10A), en el cual se observa que prácticamente todas las secuencias de proteínas se agrupan en sólo dos regiones (150-230 aa y 370-450 aa) y que entre ambas existe una ventana alrededor de 250 a 350 aminoácidos en donde no existen secuencias, lo cual nos sugiere la presencia de sólo dos tipos de proteínas: las de cadena corta o monodominio (SCHR) y las de cadena larga o bidominio (CHR), y que las proteínas pueden contener ya sea uno (proteínas SCHR) o dos dominio (proteínas LCHR) de ChrA de alrededor de 200 aminoácidos cada uno (Figura 10B). Notamos la presencia de una proteína que tiene 906



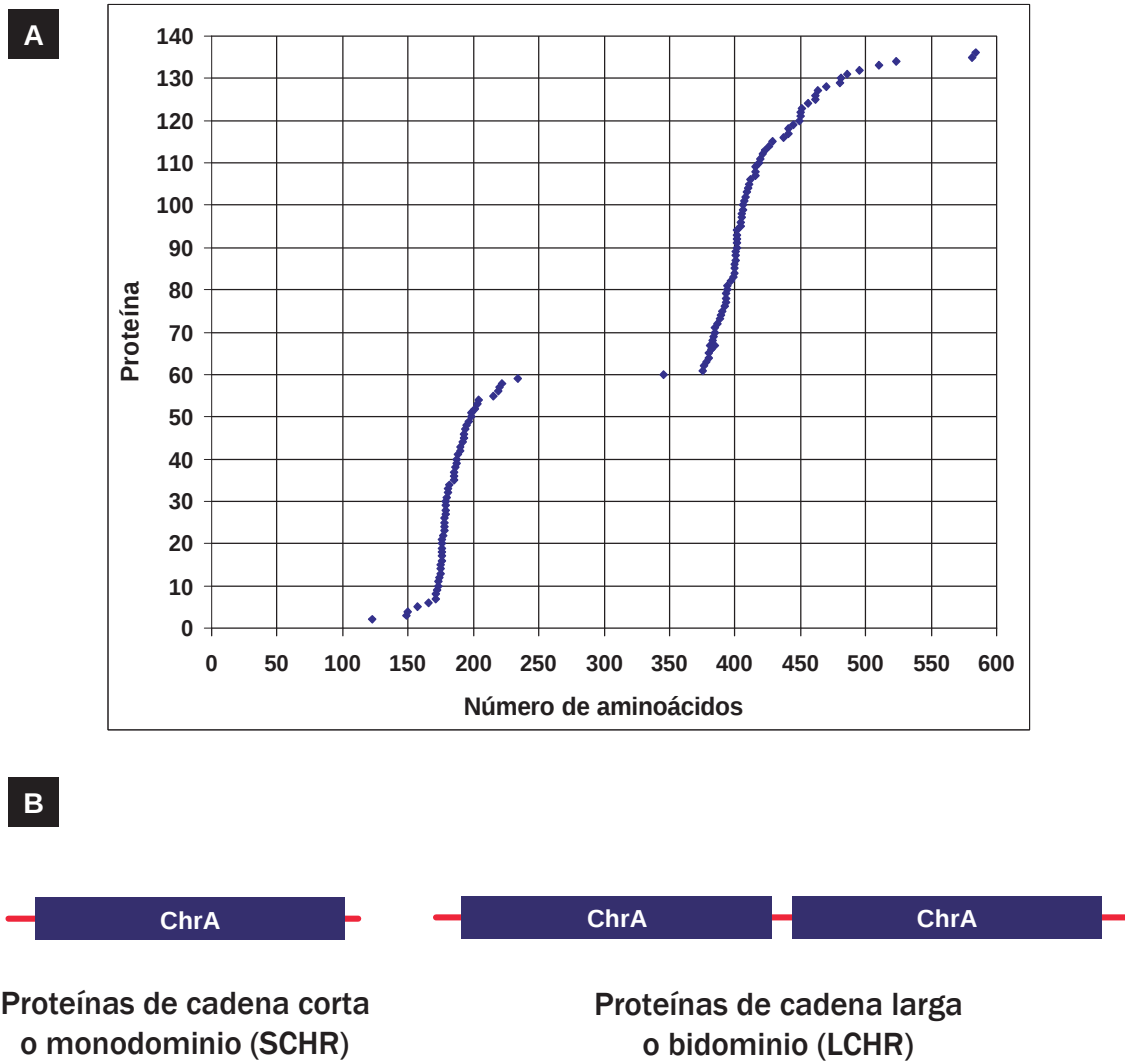


Figura 10 – Distribución del tamaño de las 135 proteínas del juego de datos. A) Se grafica la proteína (abscisa) contra el número de aminoácidos que forman a cada proteína (ordenadas). B) Diagrama de la estructura de las proteínas de cadena corta y las proteínas de cadena larga.

aminoácidos (XP_384019), lo cual es debido a la presencia de un dominio extra de 400 aminoácidos completamente distinto a ChrA.

Con la información obtenida, procedimos a hacer un análisis de localización de los dominios de ChrA dentro de las proteínas mayores, el cual se efectuó usando los resultados de Blastp. El análisis de Blastp se efectuó utilizando como secuencia a buscar cada una de las proteínas de cadena corta, y usando como base de datos las proteínas de cadena larga (chra_77), ya que así podemos saber qué parte de la proteína contiene el dominio de ChrA y si todas las proteínas de cadena larga efectivamente poseían dos dominios de ChrA. Para el análisis de dominios se diseñaron e implementaron los siguientes scripts (ver anexo III):

A. parse_Blast1.pl

Extrae datos de los archivos de salida para el análisis de dominios (E, identidad, homología, región).

B. reporte_Blast.pl

Mediante un archivo donde se proveen los grupos, reconoce cada secuencia, asigna el grupo a la secuencia y asigna el grupo a los resultados de Blastp.

C. ordena_Blast.pl

Ordena las fichas de salida en un solo archivo, colocando en orden cada salida de acuerdo al grupo al que pertenece cada secuencia.

D. linea_ord.pl

Ordena los datos de salida para poder ser analizados y leídos fácilmente en una hoja de cálculo.

Estos scripts permitieron un rápido y sencillo análisis de los datos crudos de salida de Blastp.

De este análisis se observó que todas las proteínas menores se alinearon a las proteínas mayores ya sea en el dominio amino, el carboxilo, y mayoritariamente en los dos dominios, sugiriendo que, efectivamente, las proteínas mayores poseen dos dominios de ChrA. Además, no se encontró una correlación en cuanto a cuál de los dos dominios de las CHR de cadena larga, tuviera mayor identidad o similitud con las proteínas de cadena corta SCHR.

Con la información recabada, se efectuó un análisis de localización de cada uno de los dos dominios que conforman las proteínas CHR de cadena larga, usando los resultados de Blastp. El análisis de Blastp se efectuó utilizando como secuencia a buscar



cada una de las proteínas SCHR o de cadena corta, y usando como base de datos las proteínas de cadena larga o CHR (chra_77). De los resultados del alineamiento de las proteínas SCHR contra las LCHR, se observaron las zonas de las proteínas mayores donde se alineaban las proteínas menores (lo cual se logró con modificaciones menores a los scripts de PERL antes mencionados).

Este hallazgo, aunado a los resultados anteriores, nos indicó que se pueden cortar las proteínas LCHR y dividir en los dominios de ChrA, amino y carboxilo, para analizarlos de manera independiente, y no como una sola proteína, para poder efectuar dos tipos de abordajes: el primero, analizar de manera independiente las proteínas LCHR completas, y el segundo los dominios de ChrA, tanto los dos dominios de las proteínas de cadena larga (LCHR) y el dominio de las proteínas de cadena corta (SCHR).

VI.1.C. Partición de dominios

Para entender mejor la relación entre las proteínas de cadena larga o CHR, y las de cadena corta o SCHR, las primeras se cortaron y se manejaron como dominios individuales. Para tomar la decisión de dónde se debería de cortar cada proteína CHR de cadena larga fue necesario establecer los límites de los dominios. Para este fin se usaron los resultados del alineamiento de las proteínas SCHR contra las CHR para localizar las zonas de las proteínas CHR de cadena larga donde se alineaban las proteínas SCHR. Como se esperaba, se identificaron regiones de baja similitud en el centro de cada proteína CHR de cadena larga.

Los datos obtenidos se correlacionaron con los alineamientos, de manera independiente, de cada grupo de proteínas de cadena larga. Como se esperaba, se pudieron identificar zonas de alta similitud y zonas de baja similitud, donde no alineó ninguna proteína SCHR y donde hay baja similitud inclusive entre las mismas CHR de cadena larga. Con esto se obtuvo una región interdominio, de tamaño ligeramente variable según el grupo de proteínas CHR, la cual se pudo utilizar para separar con seguridad los dos dominios que conforman las proteínas de cadena larga o CHR, ya que teníamos la seguridad de no cortar ningún dominio ChrA.

Se procedió a efectuar el corte de las proteínas, lo cual se llevó a cabo de manera automática usando el script de PERL *separa_sec.pl*, para evitar errores al momento de realizar el corte.

Una vez cortadas las 77 proteínas mayores, se contó con un juego de datos con un total de 212 secuencias proteicas (154 dominios y 58 monodominios), el dominio



amino se identificó con la terminación d1 y el dominio carboxilo se marcó con la terminación d2.

VI.1.D. Alineamiento múltiple

La parte medular de todo análisis filogenético es el alineamiento de las secuencias, ya que todos los métodos usados lo toman como base (Baldauf, 2003). El juego de datos final se alineó usando clustalW. Un paso que se tuvo que efectuar antes de proceder a aplicar los algoritmos filogenéticos es refinar el alineamiento, dado que los programas existentes para realizar alineamientos múltiples, con el fin de disminuir el tiempo de corrida y el uso de memoria de la computadora utilizada, cometen pequeños errores al momento de alinear las secuencias.

El refinamiento se realizó por inspección visual utilizando el editor BioEdit, con base en los resultados de los alineamientos de las búsquedas con Blastp para tomar las decisiones sobre dónde hay errores en el alineamiento múltiple. Cabe resaltar que el refinamiento del alineamiento se efectuó a mano, dado que no hay un programa suficientemente sensible para localizar y reparar dichos errores. Este último alineamiento fue el utilizado en la construcción de los árboles filogenéticos.

El alineamiento arroja un dendrograma preliminar (Figura 11), donde se pudo observar, para nuestra sorpresa, que los dominios se separaron claramente en tres grupos distintos: uno conteniendo a los dominios amino de las CHR de cadena larga, otro incluyó a los dominios carboxilo de las CHR de cadena larga, y un tercer grupo englobó a las proteínas SCHR o de cadena corta, corroborando los resultados de la figura 9.

VI.1.E. Construcción de los árboles filogenéticos

Como se mencionó, se pueden hacer dos tipos de análisis, uno de las proteínas de cadena larga completas (CHR) y otro analizando los dominios por separado. Se optó por hacer ambos para poder tener una visión completa de la historia evolutiva de las proteínas de la familia CHR.

Un paso muy importante en la construcción de un árbol filogenético es la colocación de la raíz. Para esto utilizamos dos miembros de la familia de proteínas intercambiadoras de disulfuro/tiol, ZP-00219994 y ZP-00278165 de las bacterias *Burkholderia cepacia* y *Burkholderia fungorum*, respectivamente, dado que secuencias de esta familia de proteínas se encontraron en la búsqueda con PSI-Blastp, aunque con una similitud baja (10 %), y por lo tanto un número E alto ($E > 0.01$). La colocación de la raíz



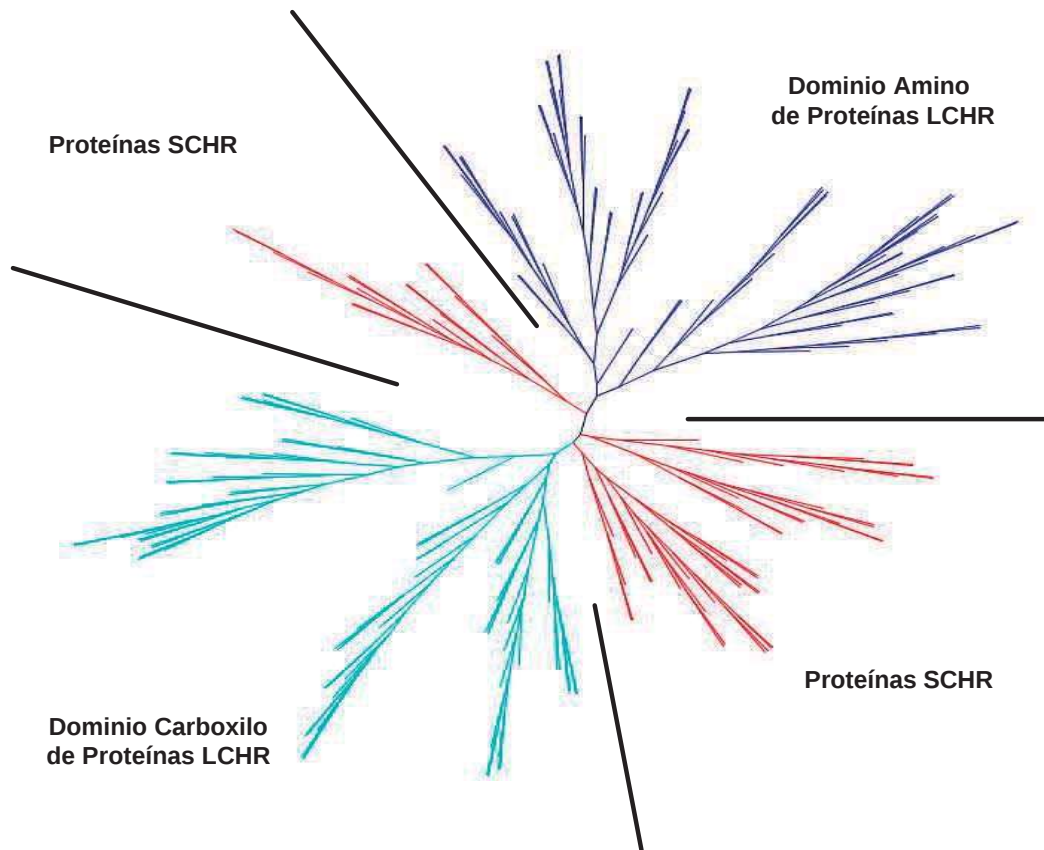


Figura 11 – Dendrograma preliminar de las secuencias de los 212 dominios del grupo de trabajo. Cada tipo de dominio se indica según la coloración de las ramas. Este árbol se elaboró alineando el nuevo grupo de trabajo con clustalW. Se aplicó una transformación logarítmica a la distancia entre las ramas. Se visualizó y editó con HyperTree.

se efectuó alineando el consenso del alineamiento refinado de las 212 secuencias con las dos secuencias seleccionadas.

Con el alineamiento refinado y ya con las secuencias raíz, se procedió a construir los árboles filogenéticos, para lo cual se usó el programa MEGA2 (Kumar y cols., 2001), el cual es muy versátil, fácil de usar y trabaja en un ambiente integrado para la visualización de los árboles. Para la construcción de los árboles, se utilizaron cuatro algoritmos filogenéticos distintos, tres del tipo MMP, UPGMA, NJ y ME, y uno del tipo MBC, MP. En dichos árboles se utilizó el modelo de evolución de Poisson y se realizó una eliminación apareada. Para la validación estadística de los árboles se efectuó un análisis de bootstrap de 1,000 repeticiones en cada caso.

VI.1.E.i. Reconstrucción de la filogenia con base en los dominios independientes de CHR

Como resultado obtuvimos árboles muy semejantes con los cuatro métodos, si bien la topología es ligeramente distinta entre cada método, los grupos formados en general son los mismos. Además cabe resaltar que se siguió observando la distribución por separado, de las proteínas menores, de los dominios amino y de los dominios carboxilo (Figura 12).

Se observaron siete grupos muy bien definidos (Figura 13), los cuales se nombraron de acuerdo a la distribución de los integrantes de los mismos, así el grupo CHRA es donde se encuentra la proteína ChrA de *P. aeruginosa*, el grupo CHRA emparentadas es el grupo más cercano al grupo de CHRA, CHRA de hongos y CHRA emparentadas de hongos y el grupo con las proteínas SCHR (Figura 14).

Los grupos donde se ubicaron los dominios amino y carboxilo de la proteína ChrA de *P. aeruginosa* constaron de diferentes números de secuencias, pero la rama (subgrupo) que agrupa a cada dominio de ChrA, contuvo el mismo número de secuencias, 15 dominios y una proteína menor (Figura 14). Cabe señalar que las proteínas que formaron tales

grupos son las mismas, sólo se diferenció el dominio. Una observación interesante fue que también en dichos grupos la proteína menor que se encuentra en el grupo amino es la ZP-00051015 y junto a los dominios carboxilos es la ZP-00052808, ambas del mismo organismo, *Magnetospirillum magnetotacticum* (Figura 12).

Las proteínas eucarióticas de hongos forman dos grupos independientes, uno con los dominios amino y otro con los dominios carboxilo (Figura 14). La única proteína



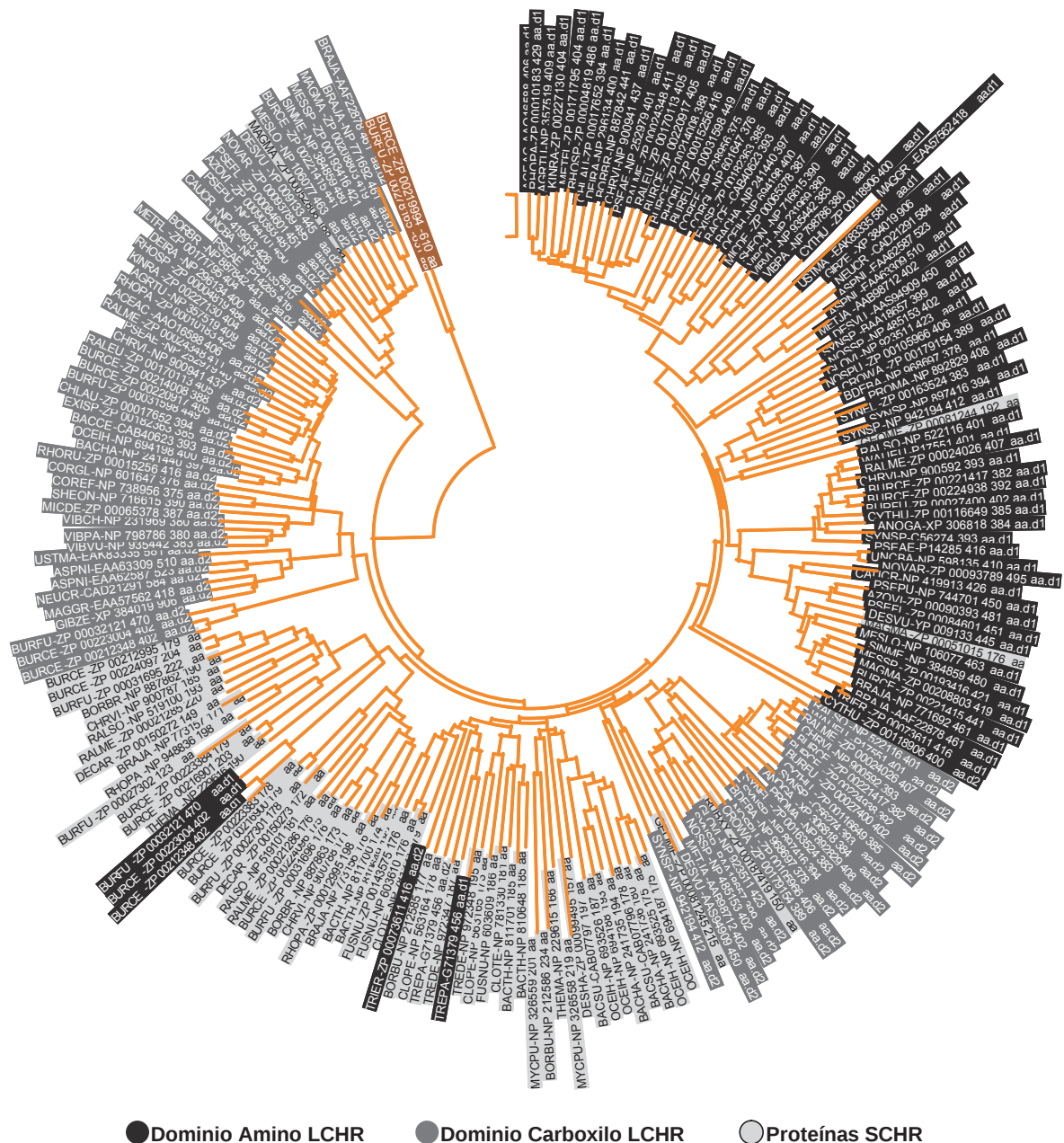


Figura 12 – Topología del árbol filogenético de la familia de CHR. El árbol se reconstruyó con las 212 secuencias de los dominios de ChrA, utilizando el método de NJ. d1-dominio amino; d2- dominio carboxilo.



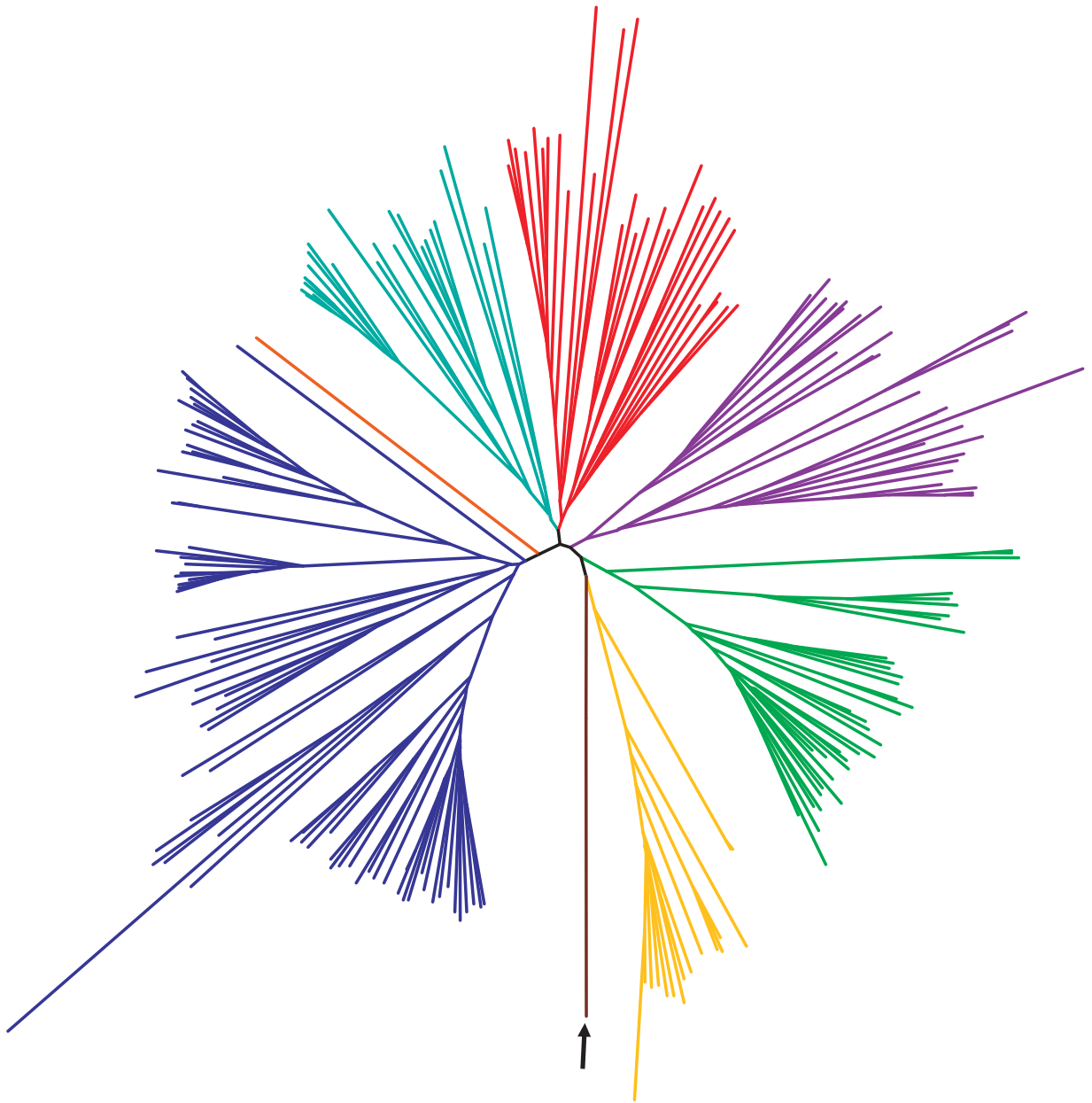
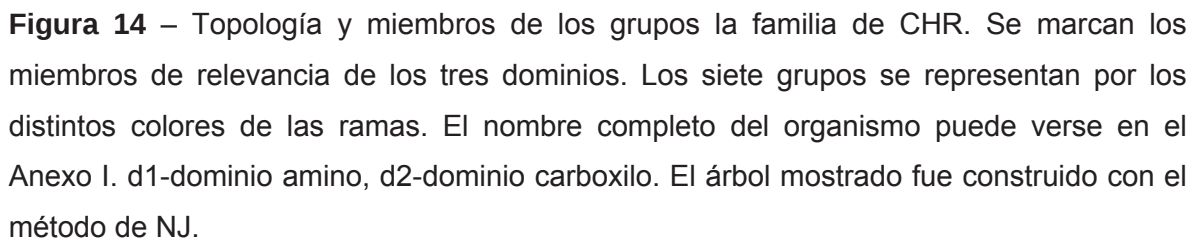


Figura 13 – Grupos de ChrA. La distancia de las ramas a la raíz, señalada con la flecha, se puede observar claramente la ramificación de los siete grupos que se forman. Los grupos se resaltan en los diferentes colores de las ramas. La flecha marca la raíz del árbol. Árbol mostrado construido con el método de NJ.



eucariótica no fúngica fue la XP-306818 de *Anopheles gambiae* (una especie de mosquito), la cual es la única proteína eucariótica que se encontró entre las bacterias (Ver anexo II).

Se efectuó un estudio taxonómico de los organismos (Figura 15), encontrándose representadas diez clases bacterianas. Una de las clasificaciones más usadas cuando se trabaja con bacterias, y que hoy en día es muy reconocida, es mediante la tinción Gram. Se aplicó esta forma de clasificación al árbol filogenético (Figura 16), donde pudimos ver que la amplia mayoría de los miembros del árbol son Gram negativos; además, salvo pocas excepciones, hubo un agrupamiento de acuerdo al tipo de tinción Gram que presentan las bacterias de donde provienen las proteínas del juego de datos. Es de interés resaltar que en el grupo donde se ubica ChrA de *P. aeruginosa* todos los organismos son proteobacterias, y por lo tanto Gram negativas.

Los valores de bootstrap para las ramas que forman a los grupos de proteínas donde se ubican los dos dominios de ChrA de *P. aeruginosa* en general fueron buenos (mayores del 70%; Figura 17), sin embargo, no en todos los métodos las ramas presentaron un valor de bootstrap elevado: siete ramas en el dominio amino y seis ramas en el dominio carboxilo tuvieron al menos un 70% de veracidad con los cuatro métodos. Esta tendencia del bootstrap se verificó para todas las ramas y grupos de los cuatro árboles filogenéticos reconstruidos, por lo que se reafirmó la veracidad de los grupos encontrados. Sin embargo, es pertinente señalar que mientras los dominios amino de las CHR de cadena larga se distribuyeron en un solo grupo, los dominios carboxilo se distribuyeron en dos grupos aparentemente separados, lo que muestra que la secuencia del dominio amino de las CHR de cadena larga está más conservado que el dominio carboxilo.

VI.1.E.ii. Reconstrucción de la filogenia de las proteínas LCHR

Dado que se observó que los dominios de las proteínas LCHR forman grupos claramente separados, se procedió a efectuar la filogenia de estas proteínas por separado. Para dicho análisis se procedió a alinear solamente las 77 proteínas LCHR completas, se efectuó el refinamiento del alineamiento, se corrieron los cuatro algoritmos filogenéticos y se realizó un análisis tipo bootstrap de 1000 repeticiones para cada árbol. En las secuencias analizadas se siguen conservando los miembros de los tres dominios, el grupo de eucariontes (formado por hongos y *A. gambiae*), la única *Archaea* (*M. jannaschii*) y un gran grupo de bacterias.



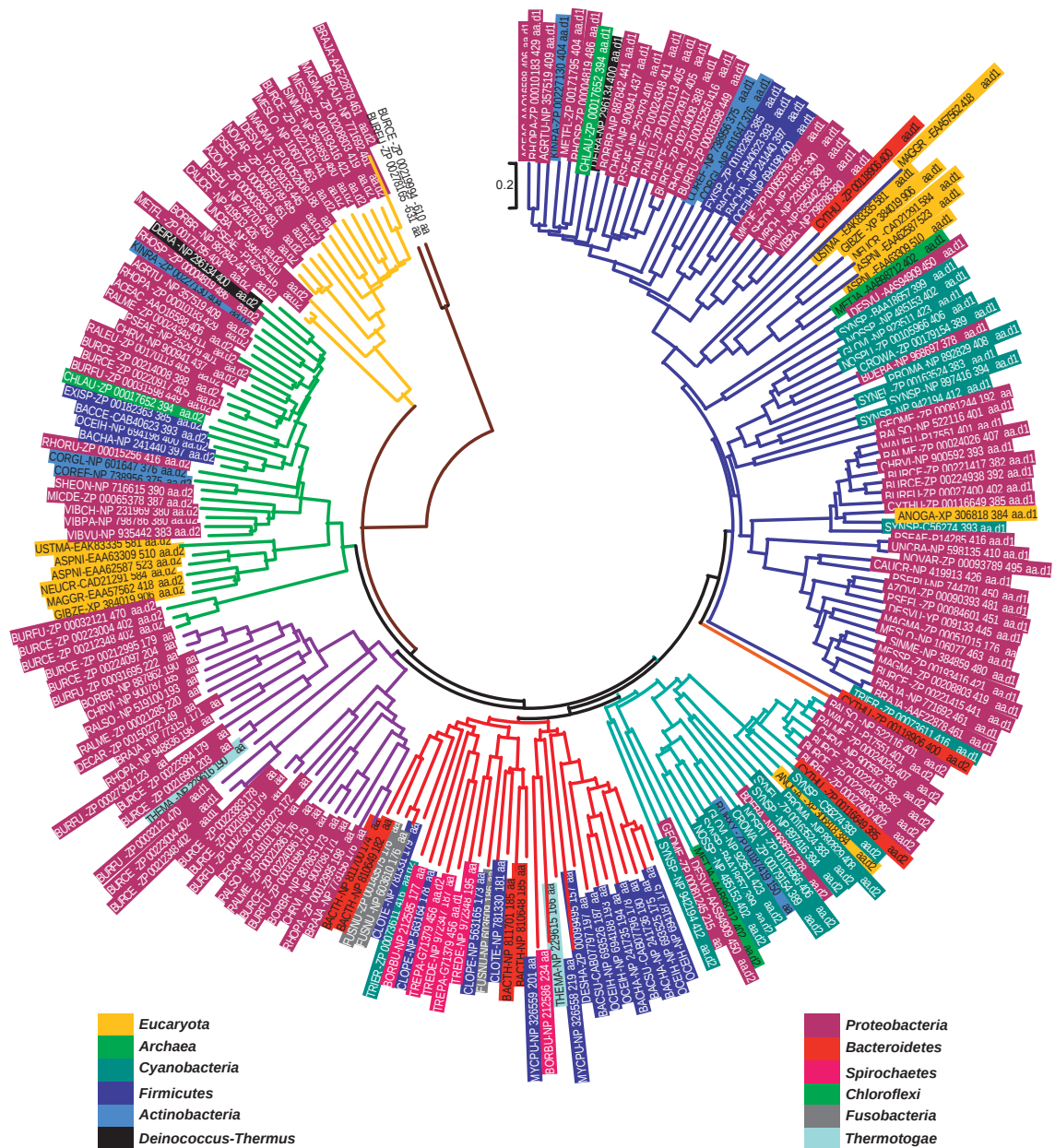
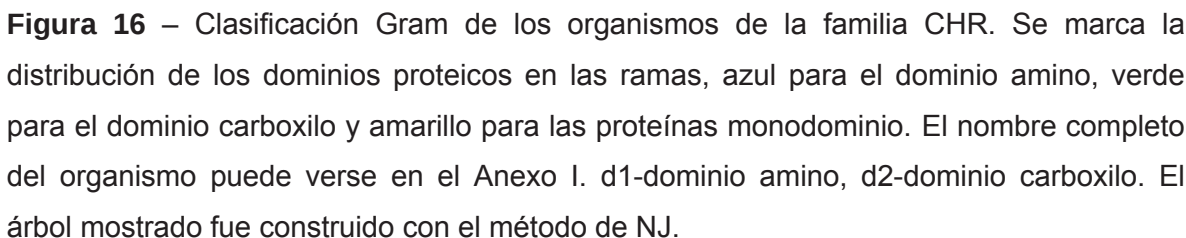


Figura 15 – Distribución taxonómica de las secuencias perteneciente a la familia CHR. Se representan los dominios *Eucarya*, *Archaea* y las clases de los organismos del dominio *Eubacteria*: Los grupos se representan por el color de las ramas. El nombre completo del organismo puede verse en el Anexo I. d1-amino, d2-dominio carboxilo. El árbol mostrado fue construido con el método de NJ.





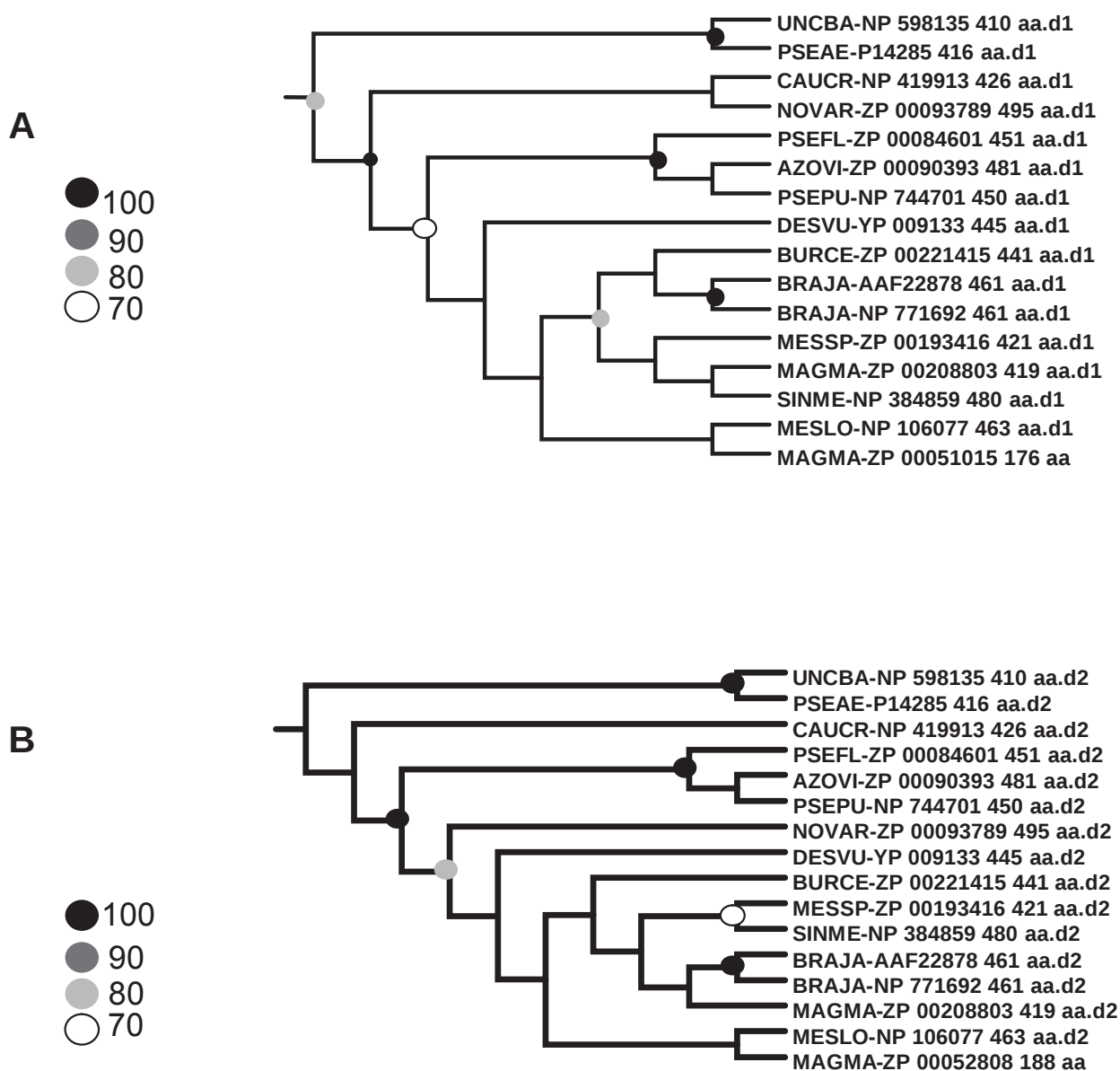


Figura 17 – Análisis Bootstrap de la rama donde se agrupa ChrA de *Pseudomonas aeruginosa*. A) Dominio amino. B) Dominio carboxilo. La leyenda en la izquierda se refiere al porcentaje que resultó de este análisis mediante los cuatro métodos, reportándose el más bajo de todos. PSEAE-P14285-416aa es la proteína ChrA de *P. aeruginosa*. El análisis se efectuó haciendo 1000 repeticiones. El nombre completo del organismo puede verse en el Anexo I. d1-dominio amino, d2-dominio carboxilo.

Al igual que se hizo anteriormente, se llevó a cabo un análisis de la taxonomía de los organismos dentro del árbol filogenético (Figura 18). Se observó que se redujo el número de clases bacterianas, ya ahora sólo hay miembros de los *Proteobacteria*, *Cyanobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Spirochaetes*, *Chloroflexi* y *Deinococcus*. Igualmente, se colocó la clasificación Gram de los organismos y, como se esperaba, hubo una agrupación dentro del árbol que también corresponde a la clasificación Gram, salvo pocas excepciones (Figura 18).

En la filogenia puede observarse claramente tres grupos principales de proteínas CHR (de cadena larga), cuya identificación está respaldada por el análisis estadístico de bootstrap (Figura 18).

El primer grupo o subfamilia contiene a las proteínas ChrA de *P. aeruginosa* y de *A. eutrophus*, que corresponden a las únicas dos proteínas funcionalmente caracterizadas; por ello proponemos denominar a este primer grupo como la subfamilia CHRA, cuya función es muy presumiblemente el transporte de cromatos.

Los otros dos grupos principales o subfamilias observados en la Figura 18 incluyen a las proteínas homólogas a ChrA presentes en hongos y otras bacterias. Ninguna de las proteínas contenidas dentro de estos dos grupos está caracterizada funcionalmente, por lo que proponemos denominarlas como CHR de hongos y proteínas emparentadas.

La única proteína homóloga de ChrA en el dominio *Archaea* identificada en este trabajo, corresponde a la previamente reportada en *M. jannaschii* (AAB98712), la cual mostró mayor similitud con la CHR de *Desulfovibrio vulgaris*, una proteobacteria anaerobia reductora de sulfatos. Ambas proteínas se agruparon juntas tanto en el árbol de las CHR (de cadena larga) (Figura 18), como en el árbol construido con los dominios individuales (Figura 16). La posición distante de esta proteína de *Archaea* (en el árbol mostrado en la figura 18) con respecto al resto de las CHR (de cadena larga), es más acorde con la posibilidad de que se trate de una proteína antigua y no el resultado de una transferencia horizontal reciente.

El análisis bootstrap (Figura 18) muestra un alto porcentaje, demostrando la veracidad de la topología de los árboles. De la misma manera, en general el porcentaje es elevado, aunque en uno de los cuatro métodos el porcentaje fue menor al 70% impuesto en este trabajo como límite mínimo.



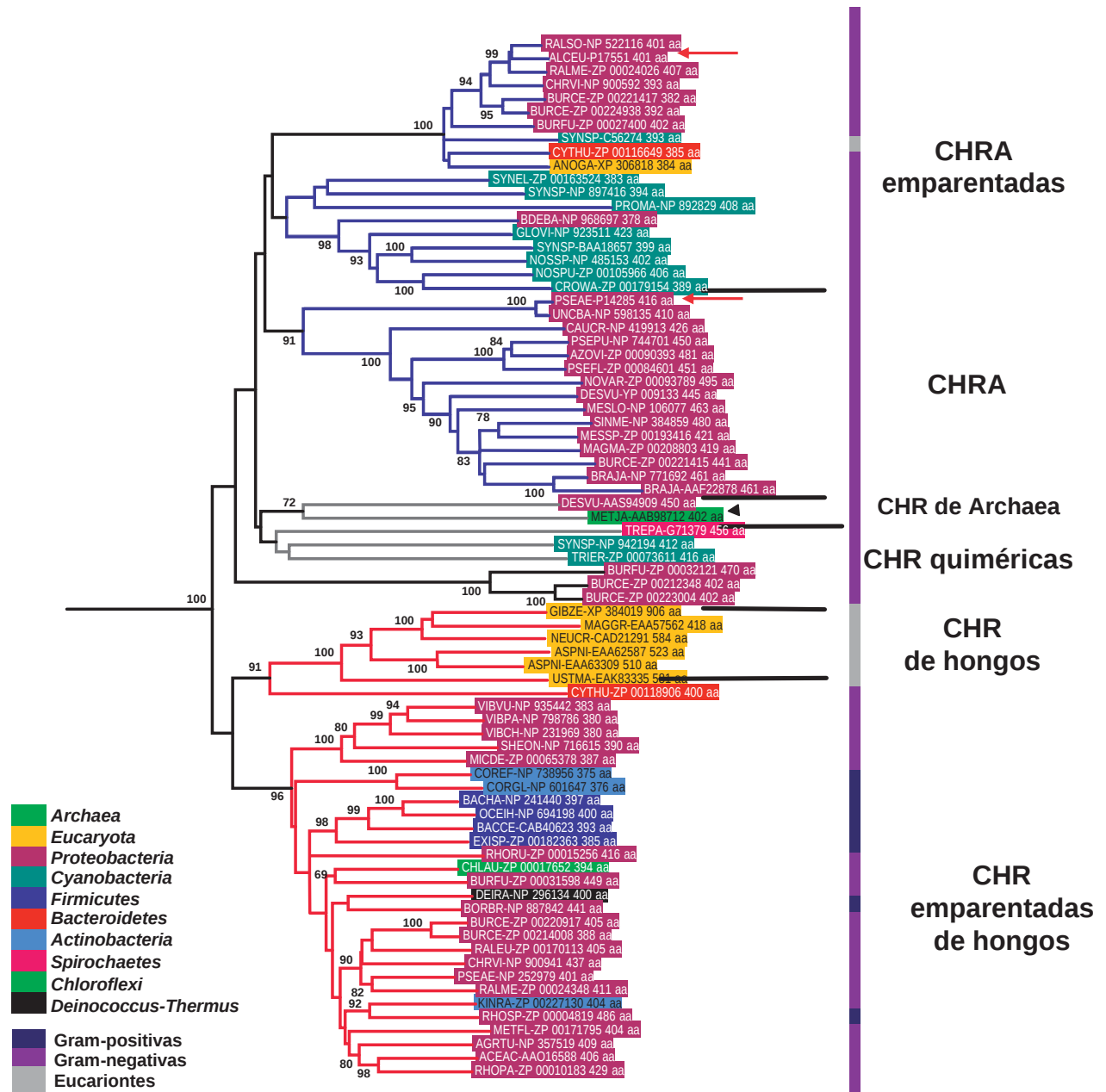


Figura 18 – Árbol filogenético de proteínas de cadena larga (LCHR) de la familia CHR. Las ramas de distinto color marcan a los tres distintos grupos. El color en la leyenda indica la distribución taxonómica de los grupos. La barra derecha muestra la clasificación Gram de las bacterias. Las flechas señalan a ChrA de *P. aeruginosa* y ChrA de *A. eutrophus*. La punta de flecha señala ChrA de *M. jannaschii*. Árbol reconstruido con ME y con un bootstrap de 1000.



VI.1.F. Identificación de aminoácidos conservados

Para la obtención de los aminoácidos más conservados se procedió a usar tres distintos abordajes. El primero, y más estricto, fue usar el alineamiento refinado de las 212 secuencias y visualizar los aminoácidos con una conservación superior al 80% de identidad usando el programa BioEdit. En la Figura 19 se muestran los 20 aminoácidos localizados con el nivel de conservación del 70 al 90%, de éstos 19 se conservan en un porcentaje mayor al 80%. El segundo abordaje fue utilizar únicamente los dominios donde se agrupa ChrA, es decir, los dominios más cercanos a los dominios de ChrA de *P. aeruginosa*. En este caso, se alinearon por separado los grupos amino y carboxilo, y se seleccionaron los aminoácidos con un porcentaje de identidad mayor. El tercer método consistió en alinear las proteínas mayores más cercanas a ChrA, para posteriormente seleccionar los aminoácidos más conservados; este método difiere del anterior en que se alinean las proteínas completas y no sólo los dominios, por lo que es más estricto que el método anterior, pero menos astringente que el primer método.

Debido a la cantidad de aminoácidos conservados que se localizaron, se decidió tomar a los aminoácidos del primer método, dado que son los mayormente conservados en todos los grupos proteicos. Además, para ser más estrictos aún, sólo se tomaron en cuenta aquellos aminoácidos con un porcentaje de conservación superior al 80%, por lo que es muy probable que dichos aminoácidos cumplan algún papel importante en la proteína, y al mutarlos podamos saber su importancia en la estructura o dentro del sistema de transporte (Figura 19).

Una vez identificados los aminoácidos conservados, se decidió seleccionar cinco de ellos para llevar a cabo la mutación sitio-dirigida. Los aminoácidos elegidos fueron (Figura 20): las glicinas 44 (G44) y 45 (G45) localizadas en el motivo KXGXXXXGG del primer segmento transmembranal (STM I), que están conservadas en un 96 y 89%, respectivamente; el triptófano 63 (W63), un aminoácido de gran tamaño, aparentemente localizado en el asa periplásmica 1 (P1), el cual está 100% conservado dentro del grupo que contiene a ChrA de *P. aeruginosa* y en un 82% dentro de los 212 dominios homólogos de ChrA; la glicina 92 (G92) que se localiza en el STM II, se encuentra conservada en un 96% en los dominios de ChrA; y la glicina 100 (G100), que presenta una conservación del 92% en las 212 secuencias de los dominios de ChrA, localizada en el STM III que además está en el motivo GGVLAG, muy conservado en el grupo donde se ubicó la proteína ChrA de *P. aeruginosa*.



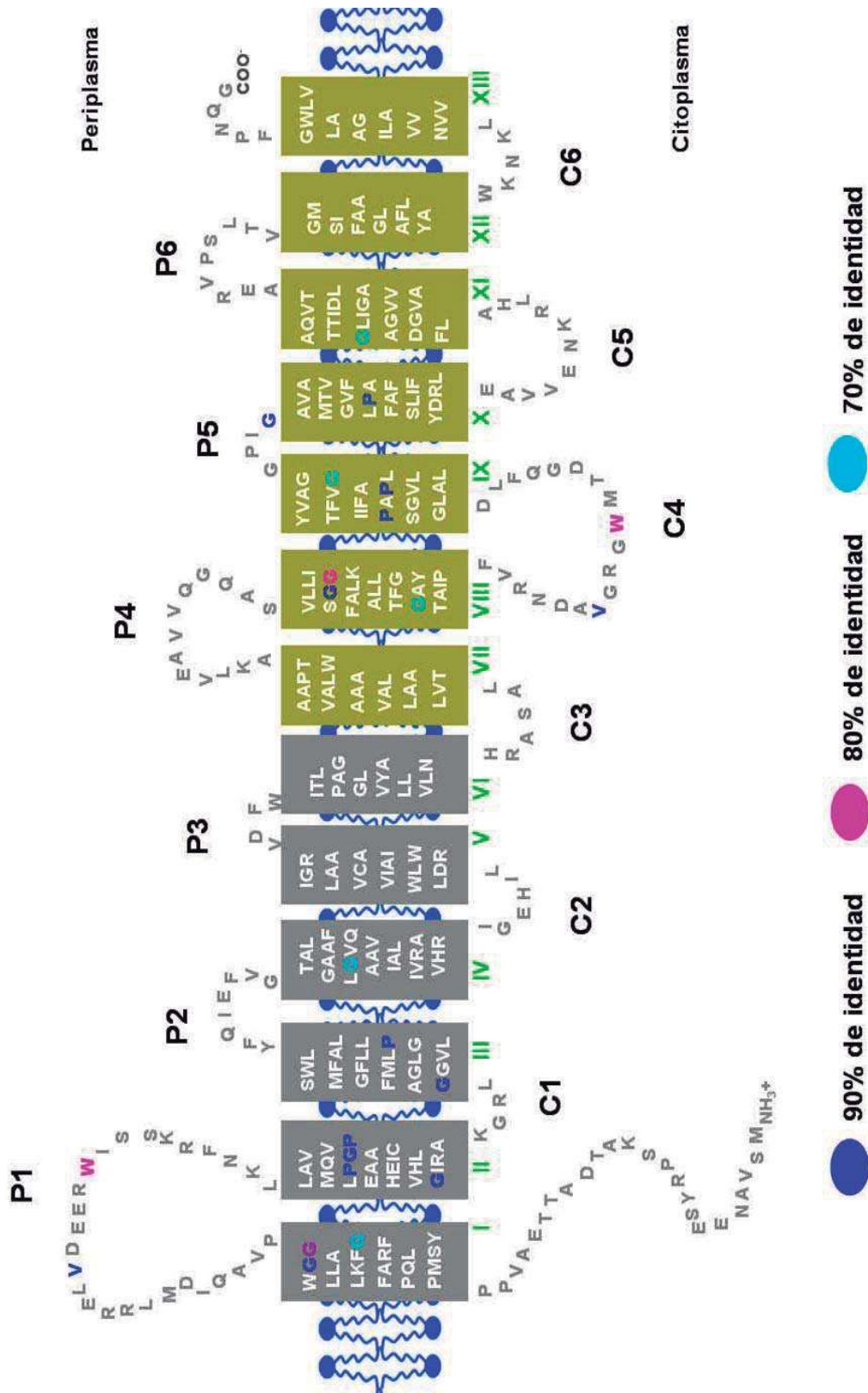


Figura 19 – Aminoácidos conservados en las 212 secuencias proteicas de la familia ChrA. El porcentaje corresponde a la identidad total dentro de las 212 secuencias. Se muestra la posible estructura secundaria de la proteína ChrA de *P. aeruginosa* como modelo para su visualización. El análisis se efectuó con el programa BioEdit, usando como fuente un alineamiento múltiple corregido.

IV.2. EVALUACIÓN DE LOS AMINOÁCIDOS CONSERVADOS EN LA FAMILIA CHR

Con el fin de evaluar la posible función de los aminoácidos más conservados dentro de la familia CHR, se decidió intentar hacer tres mutantes por cada uno de los aminoácidos seleccionados (Figura 21). Así, los cambios por la glicina, que es un aminoácido pequeño no polar, fueron por: (1) la alanina, el aminoácido no polar más parecido en cuanto a superficie y volumen; (2) la leucina, un aminoácido alifático más voluminoso; (3) la serina, el aminoácido polar más parecido en superficie y volumen (Figura 21). De la misma forma, el residuo aromático triptófano se cambio por: (1) la fenilalanina, que es el aminoácido hidrofóbico más parecido en cuanto a superficie y volumen; (2) la alanina, un aminoácido hidrofóbico más pequeño; (3) la tirosina, el aminoácido polar más parecido en cuanto a volumen y superficie (Figura 21).

IV.2.A. Obtención de las mutantes

Para llevar a cabo la modificación de los residuos, se seleccionó el método de mutagénesis sitio dirigida mediante PCR y digestión con la enzima de restricción *DpnI* (Fisher y Pei, 1997).

Se diseñaron 15 juegos de oligonucleótidos (Tabla 2), los cuales cumplieron con las características requeridas por el método (ver métodos, sección III.3.A.i). Se usó como molde el plásmido pGEMTChrAHis (Aguilar, 2005), el cual cuenta con el *chrA* bajo la regulación del promotor pTRC. Además, el producto del gen *chrA* está fusionado traduccionalmente a una cola de seis histidinas, el cual podrá ser utilizado para la futura detección de las proteínas mutantes mediante inmunolocalización; el vector cuenta con 29 sitios de restricción para la enzima *DpnI*.

Se procedió a hacer la PCR para las mutaciones establecidas, usando las condiciones descritas por Aguilar (Aguilar, 2005). Con dichas condiciones se obtuvieron las mutantes: G44A, G44L, G44S, G45L, G45S, W63F, W63Y, G92A, G92L, G92S, G100A y G100S. Sin embargo, las mutantes: G45A, W63A y G100L no se pudieron obtener con estos parámetros. La obtención de la mutante G45A, se pudo lograr al variar la concentración de los oligonucleótidos a 250 ng por reacción y la temperatura de alineamiento de 62 a 64°C. No se lograron obtener las mutantes W63A y G100L con ninguna variante del método.

Después de digerir el producto de la PCR y la transformación de la cepa XL1-BLUE se obtuvo un número variable de colonias para cada una de los de las mutaciones, desde 4 a más de 30. Se seleccionaron como máximo 20 colonias de cada placa. A cada



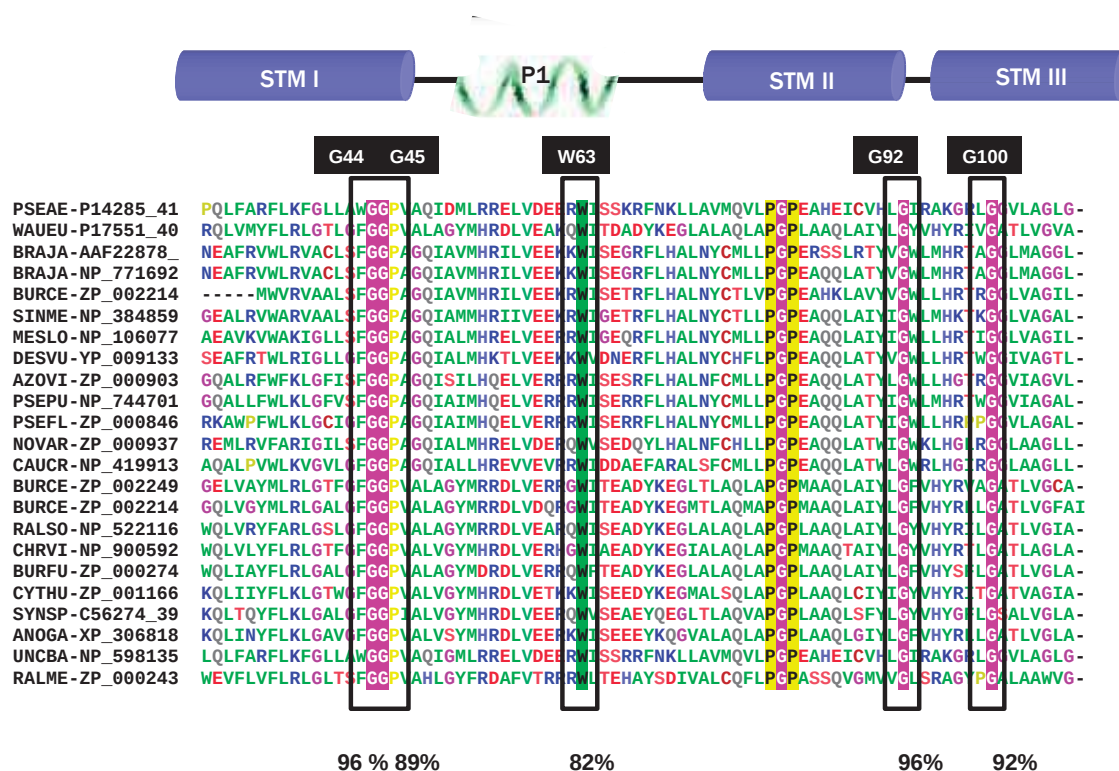


Figura 20 – Aminoácidos seleccionados para la mutagénesis sitio-dirigida. Se muestra un alineamiento parcial con las secuencias de los dominios más cercanos a ChrA de *P. aeruginosa*. En la parte superior se muestra la topología de la proteína ChrA donde se ubica la región mostrada. Los aminoácidos a mutar están dentro del recuadro. Los grados de conservación de los aminoácidos resaltados se muestran en la parte inferior. Para detalles de la leyenda ver el Anexo I. El análisis se efectuó con el programa BioEdit, usando como fuente un alineamiento múltiple corregido.

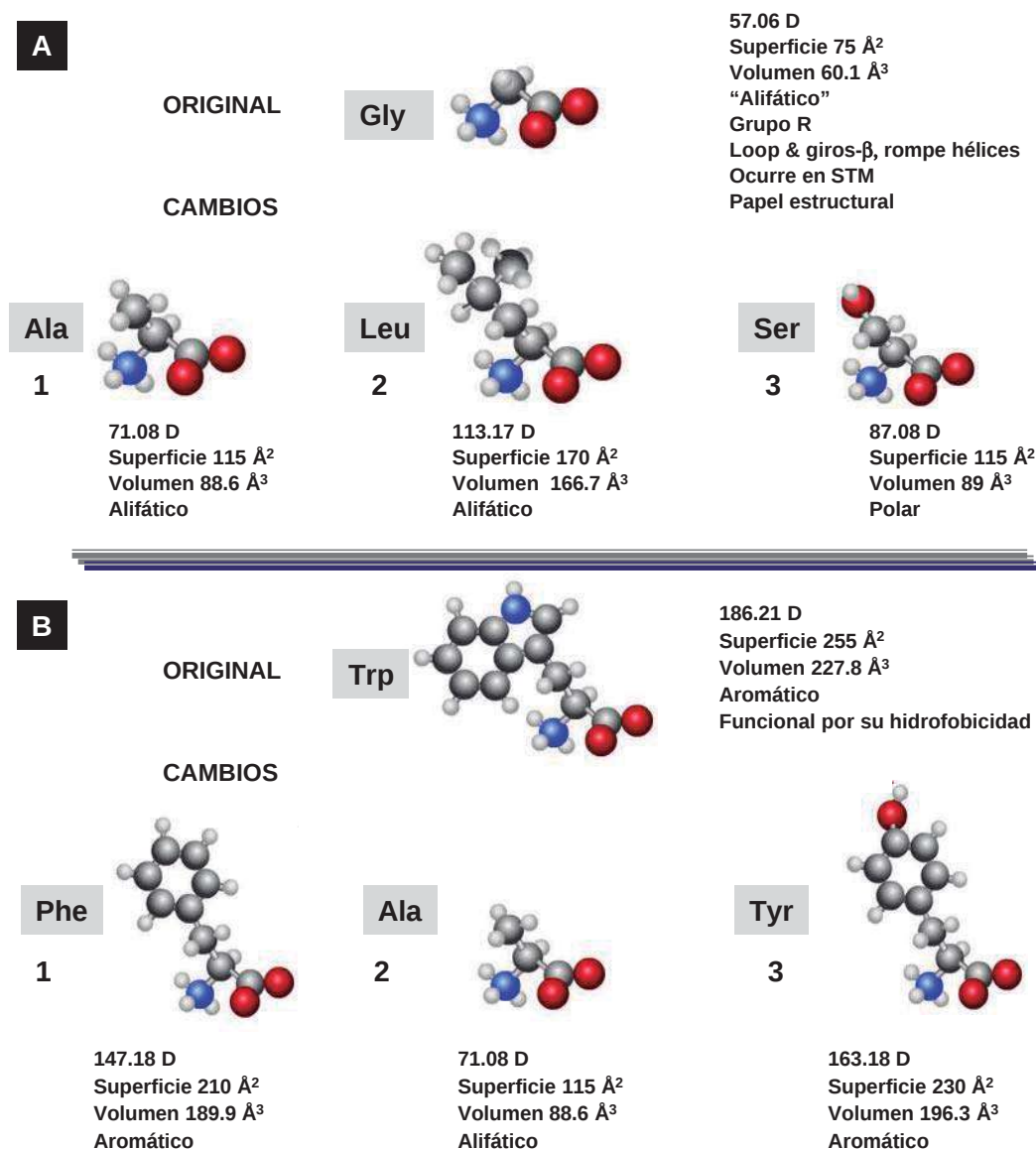


Figura 21 – Cambios propuestos para la evaluación de los aminoácidos conservados en la familia CHR. Se presentan las características principales de los aminoácidos presentados. Los cambios se hicieron de acuerdo a los siguientes razonamientos: (1) por el aminoácido más parecido al original; (2), por un aminoácido con un tamaño mayor o menor, según sea el caso del aminoácido nativo; y (3) por un aminoácido de igual tamaño pero con diferente polaridad al aminoácido presente en la proteína.

colonia se le extrajo el DNA plasmídico, el cual se procedió a analizar midiendo su tamaño molecular en geles de agarosa. A los plásmidos que tuvieron el mismo patrón de corrimiento que el control pGEMTChrAHis sin mutagenizar, se les efectuó un análisis de digestión con las enzimas *Bgl*II, *Sph*I y *Xba*I, las cuales liberaron los fragmentos esperados. Se muestra en la Figura 22 un ejemplo de la caracterización de los plásmido mutagenizados.

Para cada mutante se seleccionaron tres plásmidos que hubiesen cumplido satisfactoriamente las pruebas de restricción y se secuenciaron para comprobar que tuvieran la mutación en el codón establecido. En la Figura 23 se muestran dos ejemplos de la secuenciación de los plásmidos mutantes, donde se observa la mutación esperada.

Es oportuno mencionar que dicho método tiene la ventaja de tener una alta eficiencia, la cual en este estudio fue del 50%.

IV.2.B. Subclonación de los genes mutantes en pUCP20

Una vez que se comprobó que el gen *chrA* llevara la mutación deseada, se procedió a subclonarlo en el vector de expresión binario *E. coli*/ *Pseudomonas* pUCP20 (Holloway, 1969); esto debido a que el proceso de mutagénesis se lleva a cabo en *E. coli*, donde la proteína ChrA no es funcional, y el plásmido pGEMTChAHis no cuenta con un origen de replicación para *P. aeruginosa*.

Los plásmidos obtenidos en este paso se les denominó pUChrA-mutación, donde “mutación” se refiere al cambio de codón hecho; por ejemplo, al plásmido con la mutación en la glicina 44 por alanina se le denominó pUChrA-G44A.

Con los plásmidos ligados se transformó *E. coli* JM101. Se seleccionaron 10 colonias por plásmido mutante subclonado, a las cuales se les extrajo el DNA plasmídico para determinar su patrón de tamaño molecular. Igualmente, al DNA de las colonias que tuvieron el patrón de corrida semejante al control (pUChrA sin mutagenizar), se efectuó análisis de digestión con las enzimas *Bgl*II, *Sph*I y *Xba*I para corroborar que se tratara del plásmido pUChrA mutado (Figura 24).

Se seleccionó el DNA de una de las colonias que cumplió con las dos pruebas antes mencionadas para transformar por electroporación a *P. aeruginosa* PAO1. Se escogieron tres colonias, por cada uno de los plásmidos mutagenizados, para realizarles extracción de DNA plasmídico y a éste se le digirió con la enzima *Bgl*II para verificar que se tratara del vector con el gen *chrA* (Figura 25). Se escogió una colonia por cada plásmido con el gen *chrA* mutado para realizar pruebas de susceptibilidad a cromato.



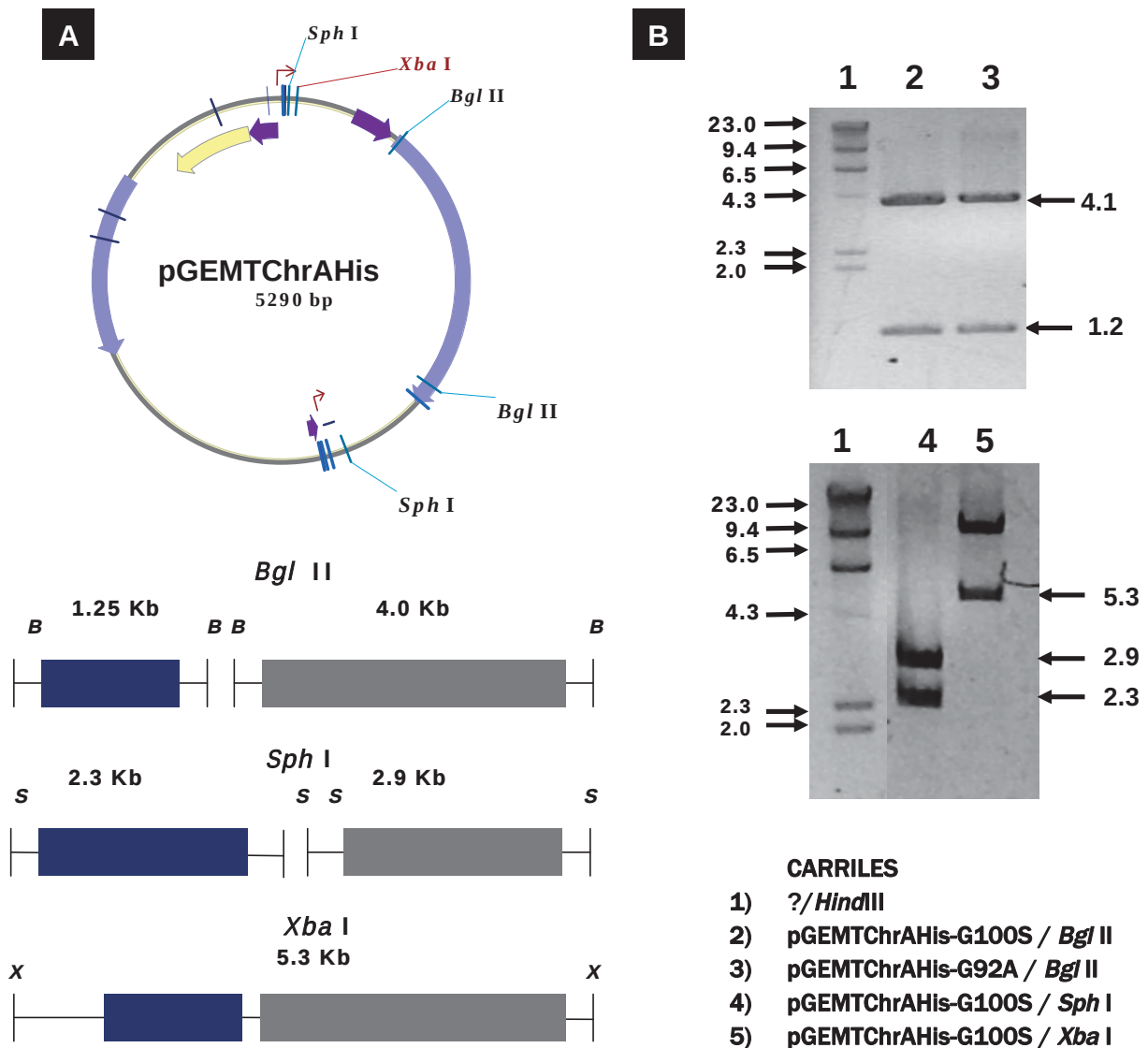


Figura 22 – Caracterización de los plásmidos pGEMTChrAHis mutagenizados. A) Esquema del plásmido pGEMTChrAHis donde se muestran los sitios de restricción para las enzimas *Bgl*II, *Sph*I y *Xba*I. B) Diagrama de los fragmentos de DNA esperados cuando se corta con esas enzimas. C) Análisis de restricción donde se muestra la fotografía de los geles de agarosa de los plásmidos pGEMTChrAHis-G100S y pGEMTChrAHis-G92A. En la parte inferior se indica el contenido de los carriles. Los números y flechas indican el tamaño molecular de marcador (izquierda) y de los fragmentos esperados (derecha) en kilobases. Este análisis se efectuó para todos los plásmidos mutagenizados, pero sólo se muestran dos resultados representativos.

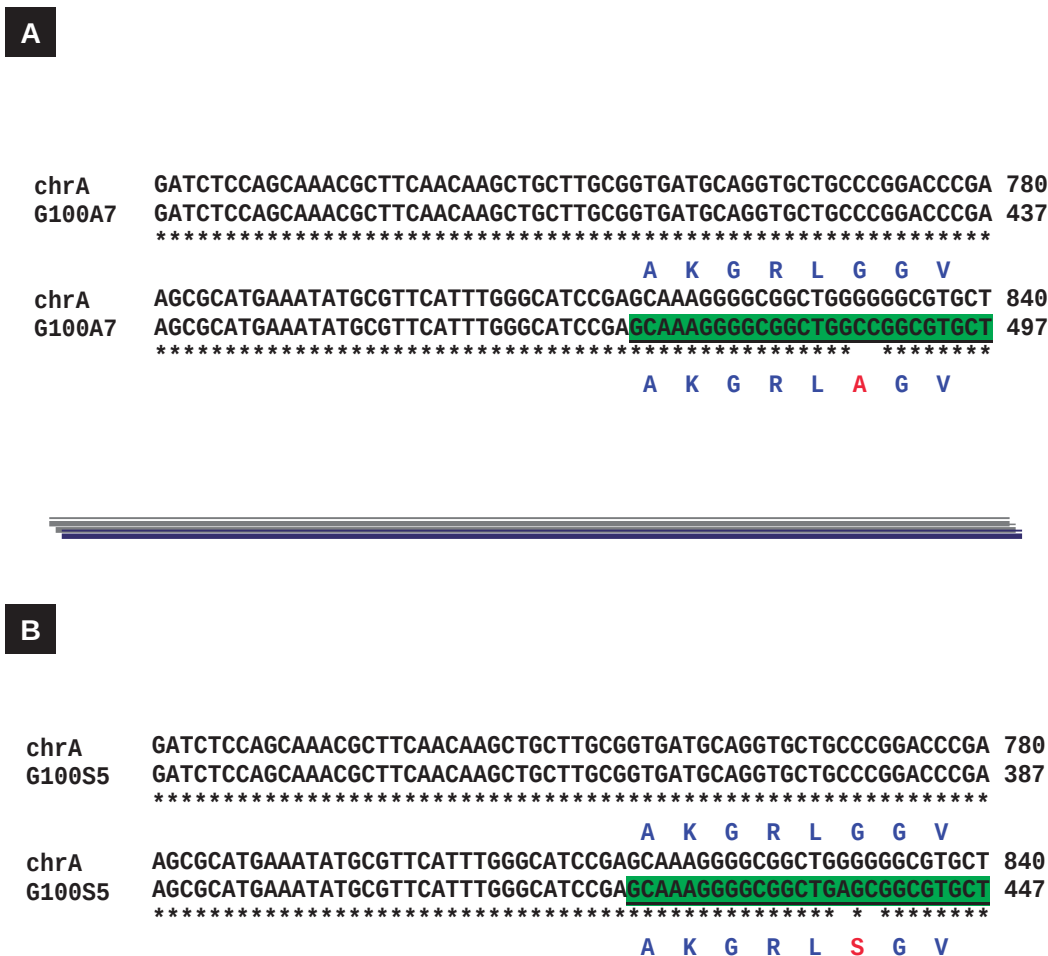


Figura 23 – Identificación de las mutaciones puntuales del gen *chrAHis* por secuenciación. Se muestra el alineamiento de los nucleótidos de la región donde se encuentra el cambio de gen *chrAHis* silvestre y el gen mutante; también se presenta una parte de la traducción del gen en la región donde sucedió el cambio. Este análisis se efectuó para todas los genes mutantes, y sólo se muestran las mutantes A) pGEMTChrAHis-G100A y B) pGEMTChrAHis-G100S.



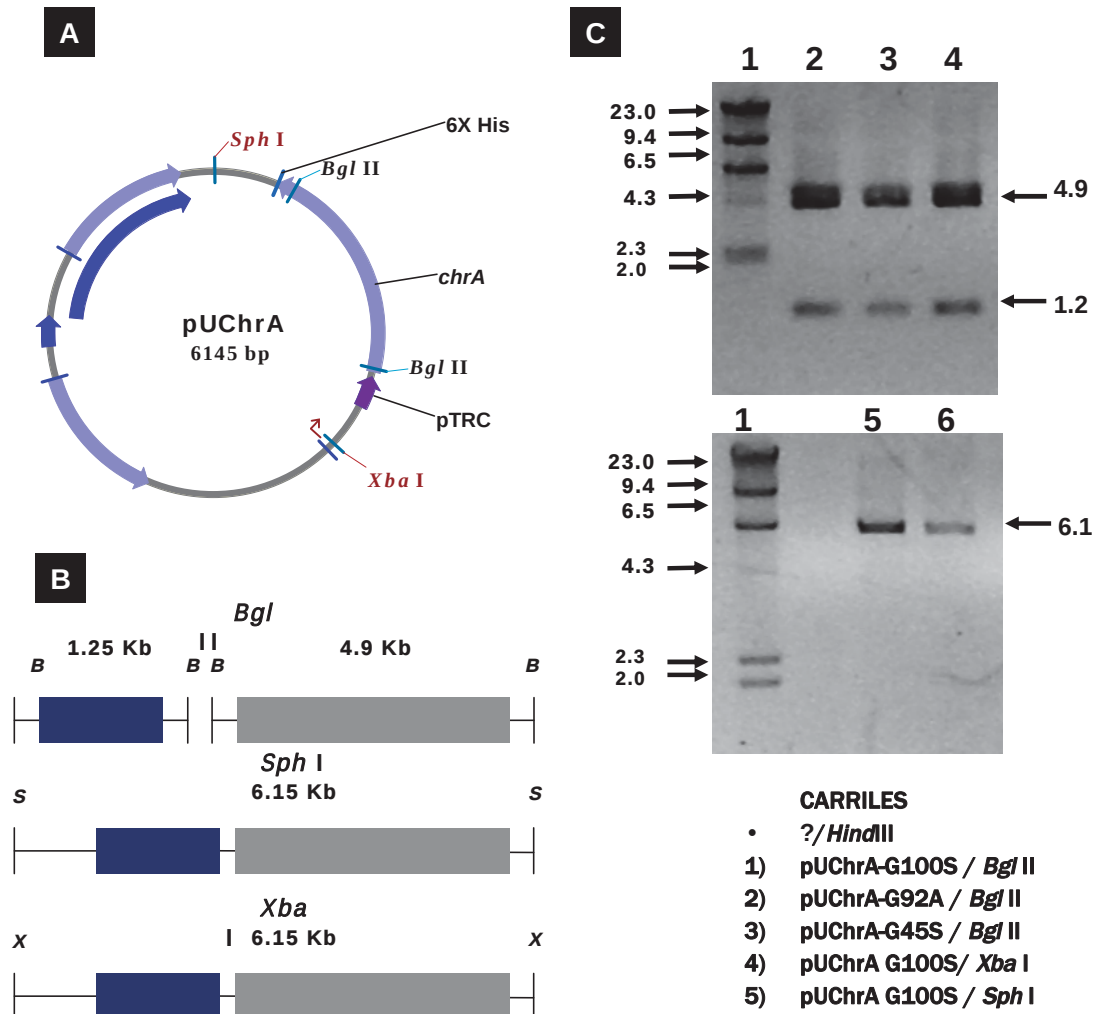


Figura 24 – Caracterización de los plásmidos pUChrA mutagenizados A) Esquema del plásmido pUChrA donde se muestran los sitios de restricción para las enzimas *Bgl*II, *Sph*I y *Xba*I. B) Diagrama de los fragmentos de DNA esperados cuando se cortan con esas enzimas. C) Análisis de restricción donde se muestra la fotografía de los geles de agarosa de los plásmidos pUChrA-G100S, pUChrA-G92A y pUChrA-G45S. En la parte inferior se indica el contenido de los carriles. Los números y flechas indican el tamaño molecular de marcador (izquierda) y el tamaño de los fragmentos esperados (derecha) en kilobases. . Este análisis se efectuó para todos los plásmidos mutagenizados, pero sólo se muestran algunos resultados representativos.

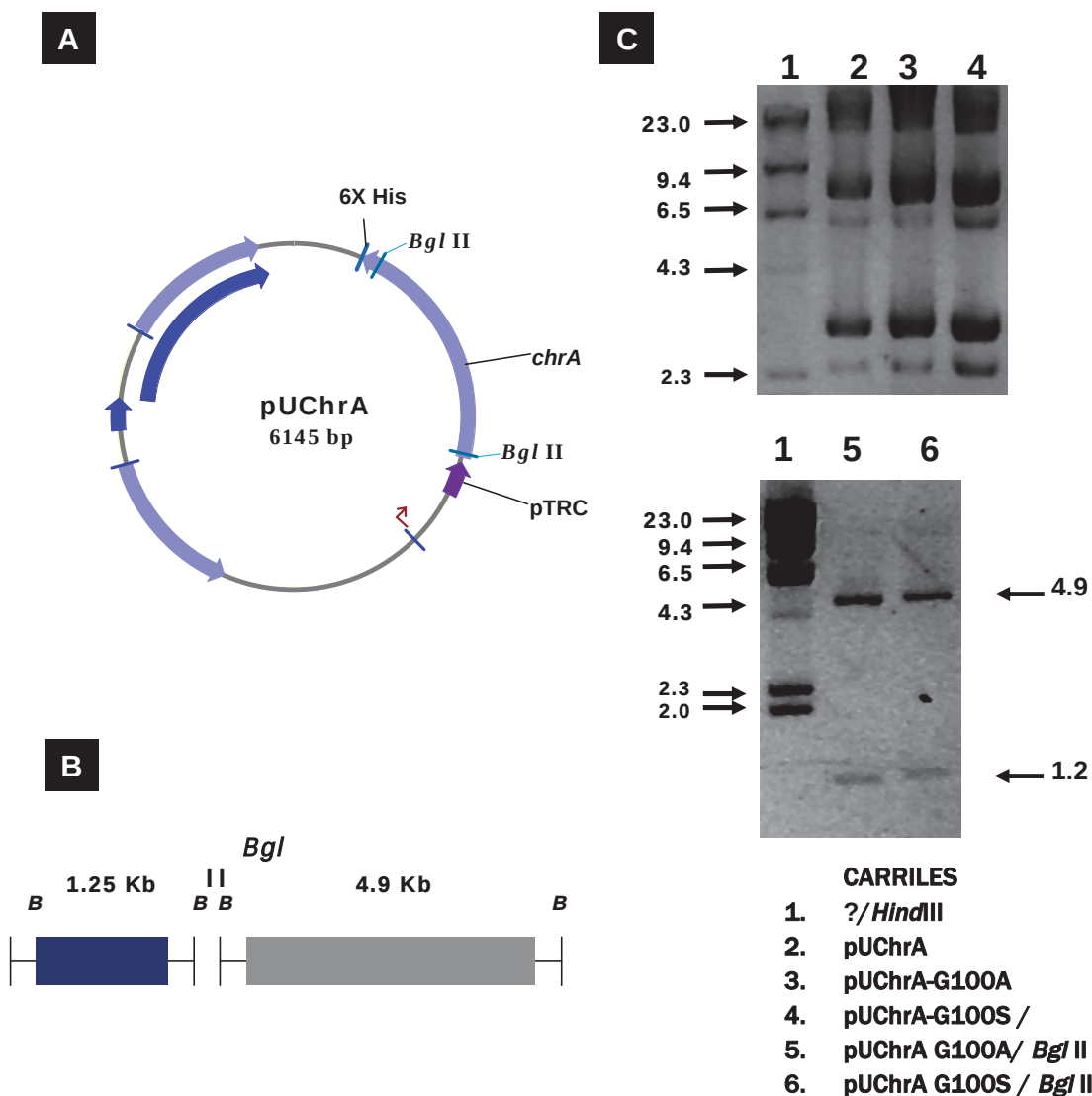


Figura 25 –Caracterización de los plásmidos pUChrA mutagenizados y transferidos a *P. aeruginosa* PAO1. A) Esquema del plásmido pUChrA donde se muestran los sitios de restricción para la enzima *Bgl*II. B) Diagrama de los fragmentos de DNA esperados cuando se corta con esa enzima. C) Análisis de restricción donde se muestra la fotografía de los geles de agarosa de los plásmidos pUChrA-G100S y pUChrA-G100A. En la parte inferior se indica el contenido de los carriles. Los números y flechas indican el tamaño molecular del marcador (izquierda) y el tamaño de los fragmentos esperados (derecha) en kilobases. Este análisis se efectuó para todos los plásmidos mutagenizados, pero sólo se muestran algunos resultados representativos.

IV.2.C. Pruebas de susceptibilidad a cromato

Con el fin de probar la importancia de los aminoácidos conservados en la funcionalidad de la proteína ChrA, se efectuaron pruebas de susceptibilidad a cromato a las mutantes que se obtuvieron anteriormente. En las pruebas de susceptibilidad en medio líquido se usó como control sensible la cepa de *P. aeruginosa* PAO1(pUCP20), que no lleva el gen *chrA*, y como control resistente la cepa de *P. aeruginosa* PAO1(pUChrA) la cual lleva el gen *chrA* silvestre.

IV.2.C.i. Glicina 44

En la prueba de susceptibilidad de las tres mutantes de la glicina 44, se observaron dos comportamientos distintos (Figura 26): en las mutantes G44L y G44S, donde se cambió la glicina por leucina y serina, respectivamente, se abatieron por completo los niveles de resistencia a cromato, comportándose como la cepa sensible. Mientras que cuando el cambio fue menos drástico, en la mutante G44A, el nivel de resistencia a cromato sólo se redujo parcialmente. Este resultado resalta la importancia de este residuo dentro de la proteína. El que la glicina 44 pueda soportar parcialmente un cambio muy sutil, sin embargo, nos indica que este residuo no es completamente esencial para que ChrA pueda llevar a cabo su función.

IV.2.C.ii. Glicina 45

Las tres mutaciones obtenidas de la glicina 45: G45A, G45L y G45S, provocaron susceptibilidad a cromato de manera similar a la del control sensible, como se observa en la Figura 27. El abatimiento de la resistencia a cromato, aun en el cambio más sutil (G45A), nos indica que este aminoácido es totalmente esencial en la función de la proteína ChrA.

IV.2.C.iii. Triptófano 63

En la Figura 28 se observa el comportamiento de las mutantes W63F y W63Y, las cuales son claramente sensibles a cromato. Este comportamiento nos indica que el triptófano en el asa P1 es sumamente importante para la correcta función de la proteína ChrA, ya que ni el cambio por el aminoácido más parecido a éste (W63F) logra recuperar el fenotipo resistente.



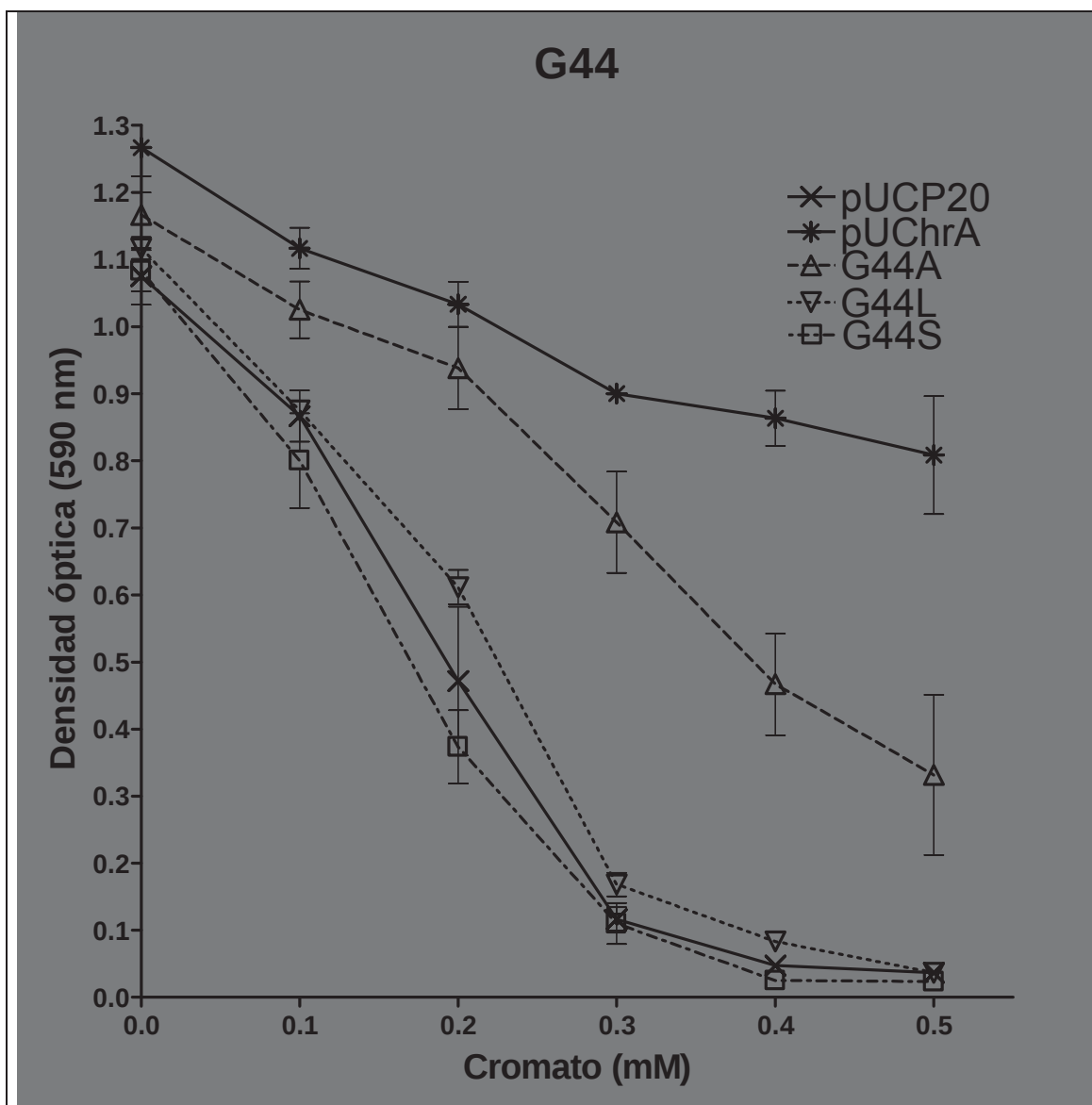


Figura 26 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 44 de la proteína ChrA de *P. aeruginosa*. Los cultivos se crecieron de 18-20 h a 37°C con agitación constante, utilizando las concentraciones de cromato indicadas y se determinó el crecimiento bacteriano por la densidad óptica de las muestras a 590 nm. Se emplearon las cepas: *P. aeruginosa* PAO1(pUCP20) -control sensible-, *P. aeruginosa* PAO1(pUChrA) -control resistente con el gen *chrA* silvestre- y las mutantes señaladas. Se muestran la barras del error estándar de la media, n=6.

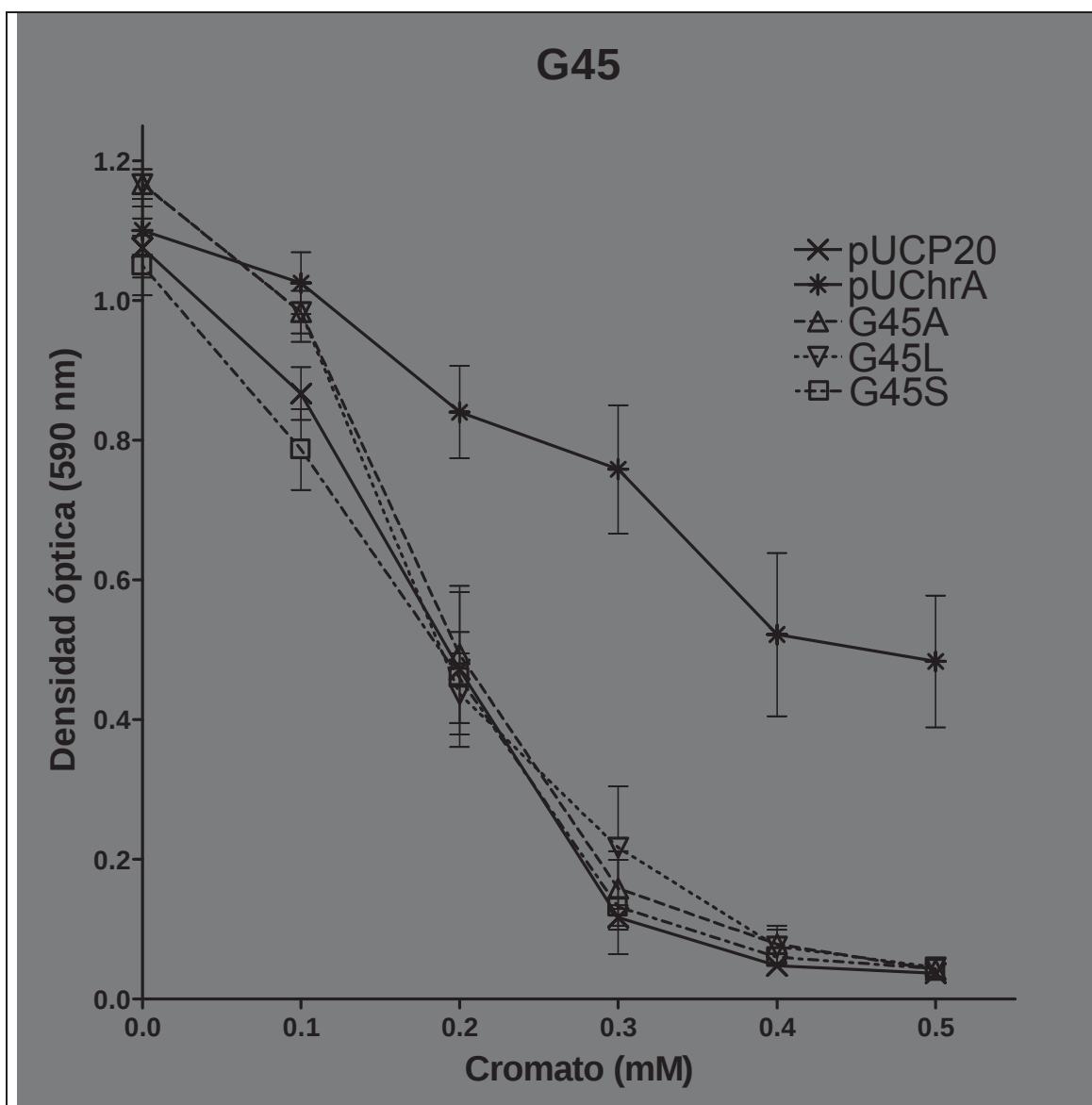


Figura 27 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 45 de la proteína ChrA de *P. aeruginosa*. Los cultivos se crecieron de 18-20 h a 37°C con agitación constante, utilizando las concentraciones de cromato indicadas y se determinó el crecimiento bacteriano por la densidad óptica de las muestras a 590 nm. Se emplearon las cepas: *P. aeruginosa* PAO1(pUCP20) -control sensible-, *P. aeruginosa* PAO1(pUChrA) -control resistente con el gen *chrA* silvestre- y las mutantes señaladas. Se muestran la barras del error estándar de la media, n=6.

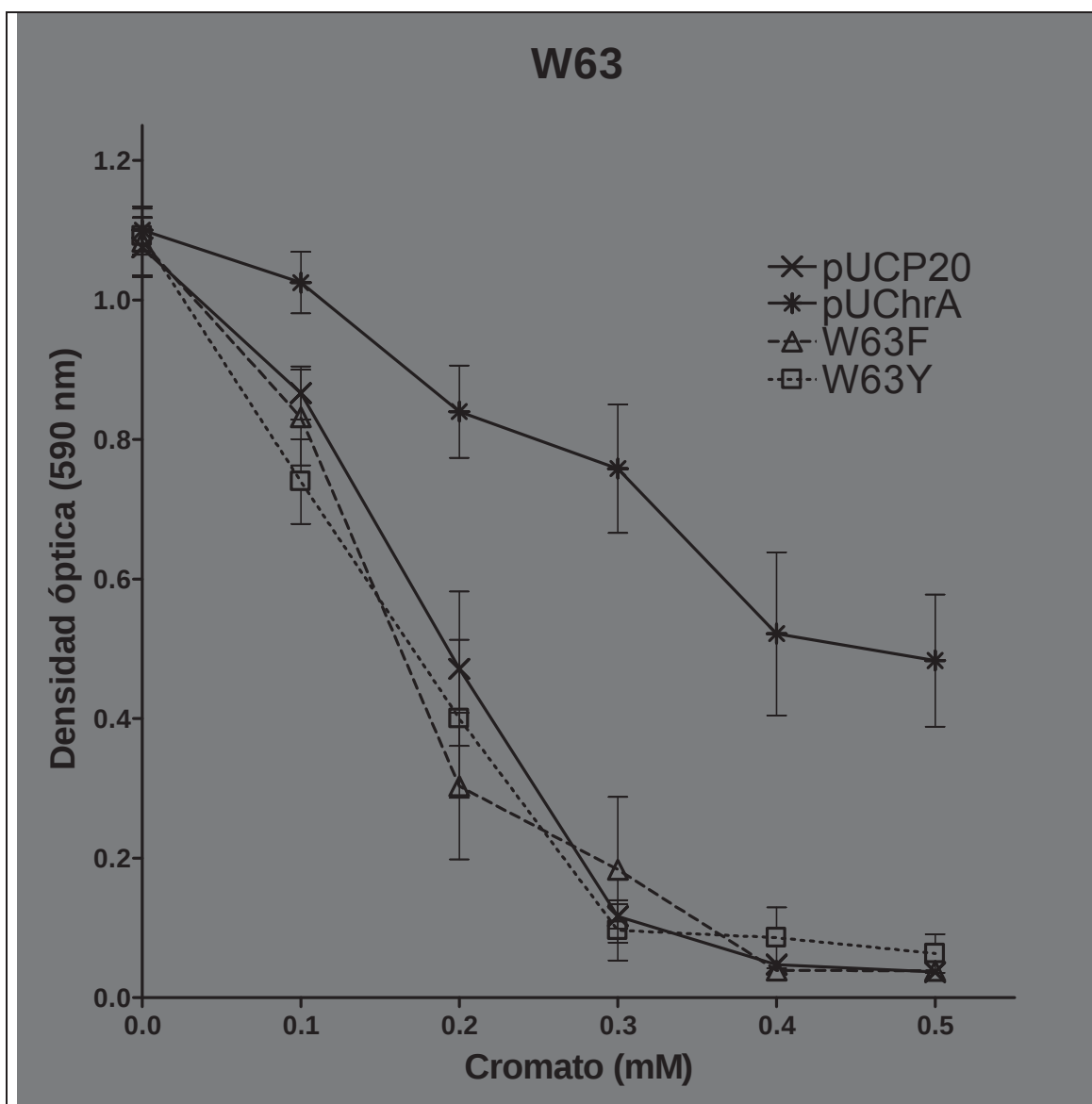


Figura 28 – Pruebas de susceptibilidad a cromato de las mutantes del triptófano 63 de la proteína ChrA de *P. aeruginosa*. Los cultivos se crecieron de 18-20 h a 37°C con agitación constante, utilizando las concentraciones de cromato indicadas y se determinó el crecimiento bacteriano por la densidad óptica de las muestras a 590 nm. Se emplearon las cepas: *P. aeruginosa* PAO1(pUCP20) -control sensible-, *P. aeruginosa* PAO1(pUChrA) -control resistente con el gen *chrA* silvestre- y las mutantes señaladas. Se muestran la barras del error estándar de la media, n=6.

IV.2.C.iv. Glicina 92

Las mutantes en los tres cambios obtenidos en este aminoácido, G92A, G92L y G92S, se comportaron de igual manera que el control sensible (Figura 29), es decir, su crecimiento se vio inhibido a bajas concentraciones de cromato por lo que, de la misma forma que la glicina 45, podemos deducir que este aminoácido es absolutamente esencial para la funcionalidad del transportador ChrA.

IV.2.C.v. Glicina 100

Se lograron dos cambios en este residuo, G100A y G100S. Podemos observar claramente (Figura 30) que la mutación G100S abatió por completo la resistencia a cromato; sin embargo, el cambio por alanina en esta posición generó un fenotipo silvestre de resistencia a cromato. Estos resultados revelan que la presencia de un residuo pequeño no polar, aunque ligeramente distinto al residuo original, es lo importante para la que se lleve a cabo la correcta función de la proteína ChrA, por lo que la G100 es importante, más no completamente esencial, para la correcta función de la proteína ChrA.



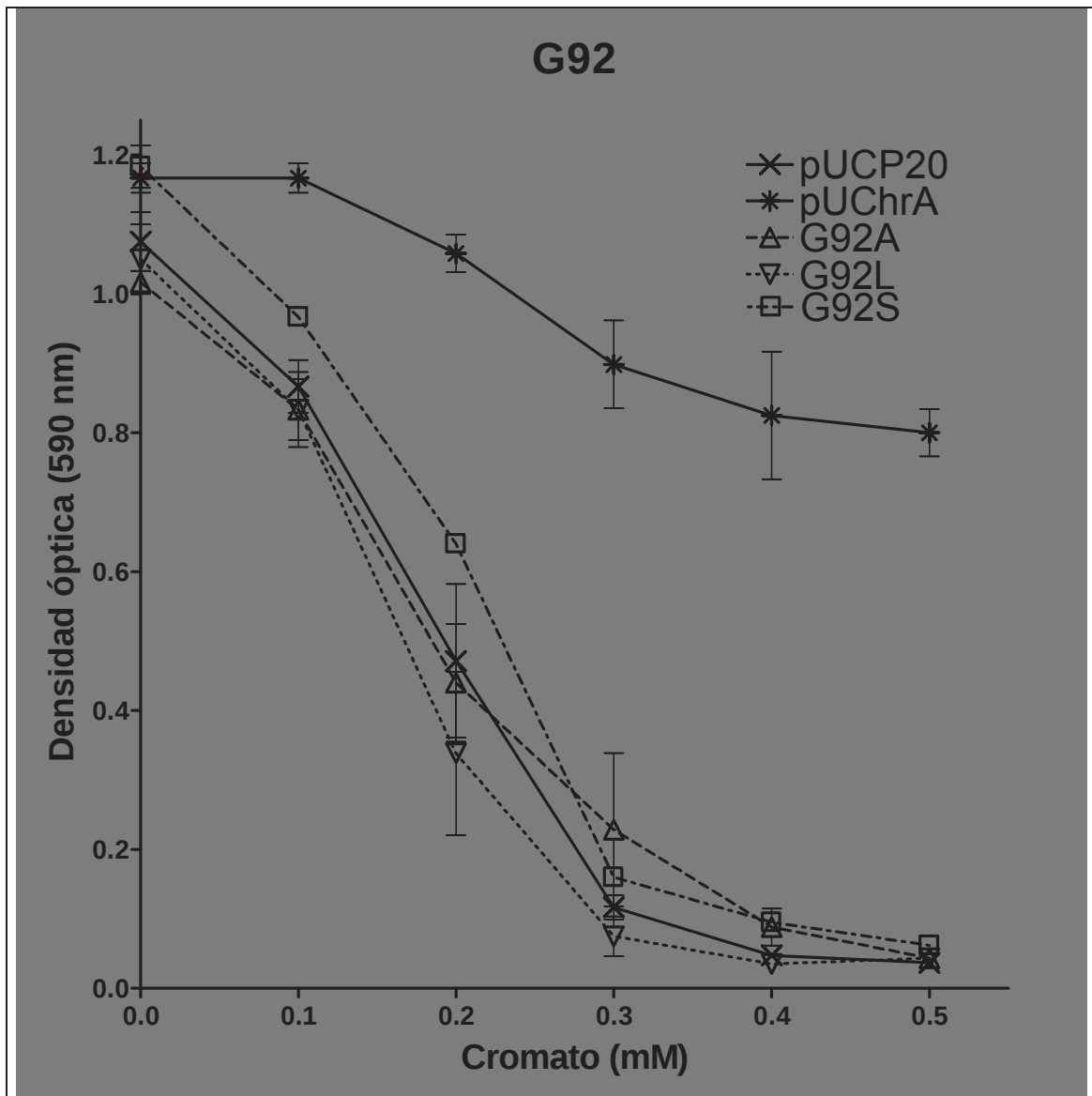


Figura 29 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 92 de la proteína ChrA de *P. aeruginosa*. Los cultivos se crecieron de 18-20 h a 37°C con agitación constante, utilizando las concentraciones de cromato indicadas y se determinó el crecimiento bacteriano por la densidad óptica de las muestras a 590 nm. Se emplearon las cepas: *P. aeruginosa* PAO1(pUCP20) -control sensible-, *P. aeruginosa* PAO1(pUChrA) -control resistente con el gen *chrA* silvestre- y las mutantes señaladas. Se muestran la barras del error estándar de la media, n=6.

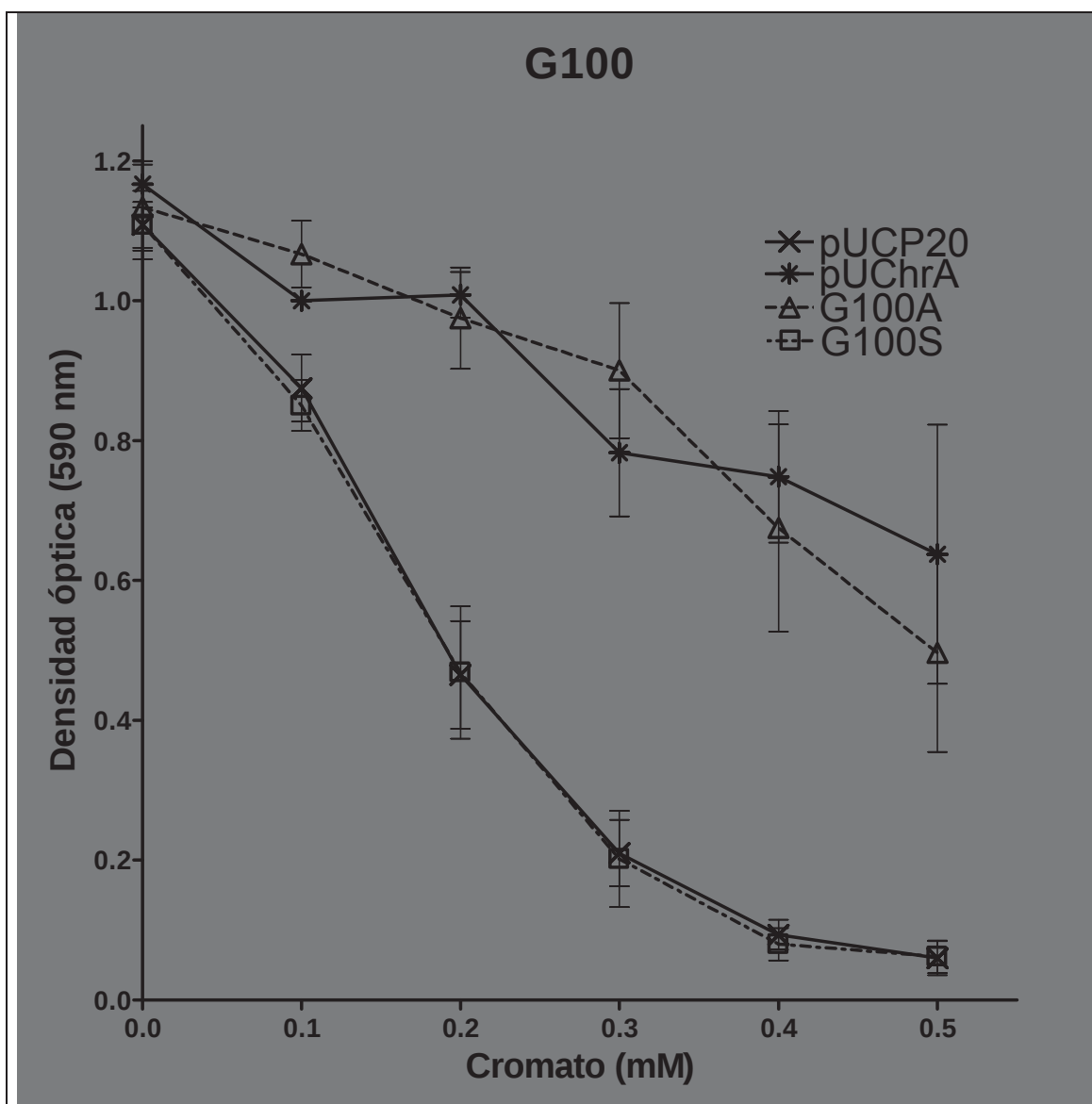


Figura 30 – Pruebas de susceptibilidad a cromato de las mutantes de la glicina 100 de la proteína ChrA de *P. aeruginosa*. Los cultivos se crecieron de 18-20 h a 37°C con agitación constante, utilizando las concentraciones de cromato indicadas y se determinó el crecimiento bacteriano por la densidad óptica de las muestras a 590 nm. Se emplearon las cepas: *P. aeruginosa* PAO1(pUCP20) -control sensible-, *P. aeruginosa* PAO1(pUChrA) -control resistente con el gen *chrA* silvestre- y las mutantes señaladas. Se muestran la barras del error estándar de la media, n=6.

V. DISCUSIÓN

La proteína ChrA es una proteína membranal perteneciente a la familia CHR, en la cual se han localizado proteínas de cadena corta (SCHR) y proteínas de cadena larga (LCHR) (Nies y cols., 1998). Sólo se conoce la función de dos de miembros de esta familia, ChrA de *P. aeruginosa* (No. acceso P14285) y ChrA de *A. eutrophus* (No. acceso P17551), las cuales confieren resistencia a cromato. Se sabe que ChrA de *P. aeruginosa* requiere del potencial generado por los componentes de la cadena respiratoria presente en la membrana (Pimentel y cols., 2002).

Para tratar de conocer la relación estructura-función de las proteínas de membrana, tomando de modelo a ChrA de *P. aeruginosa*, estudios de mutagénesis al azar llevados a cabo en nuestro grupo de trabajo llevaron a la identificación de 12 residuos cuyo cambio provocaba sensibilidad a cromato (Aguilera y cols., 2004). Sin embargo, los cambios que se lograron en este estudio fueron, en general, demasiado drásticos, por lo tanto se efectuó un segundo abordaje, también se tomó en cuenta el grado de conservación de los residuos mutados. Como la glicina 81, la cual si se cambiaba aún por el aminoácido más parecido a ella se perdía la resistencia a cromato, llevó a la identificación de este residuo como esencial en ChrA (Aguilar, 2005).

Dado que la evolución adaptativa nos indica que podemos identificar aminoácidos potencialmente importantes para que las proteínas puedan ser útiles al organismo en su medio (Yokoyama, 2002), este trabajo se enfocó en la identificación de residuos conservados dentro de la familia CHR, para tratar de encontrar más residuos esenciales. Para ello se buscó como primer objetivo obtener la filogenia de la familia CHR.

V.1. FILOGENIA MOLECULAR

La filogenia molecular se define como el establecimiento de relaciones evolutivas, mediante el uso de macromoléculas, ya sean éstas ácidos nucleicos o proteínas (Baldauf, 2003). Esta es una herramienta que puede ayudar a obtener una gran cantidad de información, ya sea estructural o funcional, al analizar grandes grupos de genes o proteínas relacionados; por ejemplo, el uso de estas técnicas enfocadas al análisis proteico puede proveer de información suficiente para tratar de dilucidar los mecanismos, estructuras y aminoácidos esenciales que intervienen en la función en las proteínas (Dopazo y Carazo, 1997).



V.1.A. La torre de babel: un viejo problema en la actualidad

La filogenia molecular, al ser una herramienta para el análisis a gran escala de los genes y proteínas de organismos relacionados, es una de las especialidades que se ha beneficiado directamente de los proyectos de secuenciación de genomas, ya que se nutre directamente de los datos aportados por estos proyectos (Rehm, 2001). Esto se hace evidente con la publicación de distintas bases de datos de familias relacionadas (Tatusov y cols., 2003; Bateman y cols., 2004; Meinel y cols., 2005), además, también se ha visto favorecida con la aparición y el mejoramiento de los programas bioinformáticos, muchos de ellos disponibles gratuitamente, para la obtención de secuencias, y la minería de datos, lo que permite llevar a cabo estos análisis en menor tiempo (Rehm, 2001; Jamison, 2003).

La bioinformática surgió hace pocos años, sin embargo, ha avanzado a pasos agigantados ayudando a realizar grandes proyectos a velocidades que hace una década jamás se hubiera pensado (Jamison, 2003). También ha promovido que se haya podido recabar una cantidad enorme de información; gracias a esto, en este trabajo pudimos reunir inicialmente 813 secuencias mediante dos tipos de búsqueda. Sin embargo, mediante un análisis preliminar de las secuencias pudimos constatar que había un gran número de secuencias redundantes y posibles fragmentos. Este resultado no es de extrañar, dada la gran redundancia de información entre las diferentes bases de datos y a que se siguen generando más y más secuencias. De hecho, este problema no sólo ocurre entre diferentes bases de datos, sino también dentro de una misma base; esto se hizo evidente en los pasos de depuración de secuencias, ya que de una sola base de datos se llegaron a eliminar hasta más de 100 secuencias repetidas.

Además, se presentó otro problema: la estandarización de formatos, uno de los más graves que desde hace años se ha tratado de resolver, y al que muchos nombran “la nueva torre de Babel” (Attwood, 2000). Históricamente, cada base de datos tiene su propio formato estándar, el cual se actualiza cada cierto tiempo; sin embargo, esta actualización muchas veces trae incompatibilidades dentro de los programas para el manejo de las secuencias, lo cual hace más difícil la automatización de procesos sencillos, pero tediosos, como son la obtención y el manejo de secuencias para la rápida colecta de información (Wickware, 2004).

Perl es un lenguaje de programación cuyo principal propósito es el fácil manejo de cadenas de texto para el manejo de bases de datos, por lo que ha cobrado una gran importancia en el manejo de información genética desde hace varios años (Stein, 2004), incluso ha resultado vital para proyecto genómicos de gran envergadura (Stein, 1996). En



este trabajo se desarrollaron varios scripts de PERL (ver anexo II) para enfrentar los problemas antes mencionados fácil y eficientemente, ya que de manera automática o semiautomática se pudieron llevar a cabo distintos pasos de la metodología, como en el manejo, uso y clasificación de los datos y las secuencias, logrando que análisis que tomarían días en llevarse a cabo manualmente se hicieran en horas. Mas aún, estos pequeños scripts jugaron un papel doble, ya que también nos ayudaron a manejar la información para conectar los distintos pasos de la filogenia, y llevarla a cabo en un menor tiempo, al evitar tediosos y tardados pasos posteriores al almacenar la información en bases de datos de acuerdo a nuestras necesidades.

Al trabajar con diferentes bases de datos, tuvimos que correr pasos de limpieza independiente, y después volver a correr pasos de limpieza generales, para llegar a obtener un juego limpio de datos. Se partió de 813 secuencias iniciales hasta llegar a tener 135 secuencias homólogas, no repetidas ni fragmentadas (Tablas 2 y 3). Se buscó el formato de la base de datos que nos diera la mayor información posible, ya que se requiere tener la mayor cantidad de información para cada secuencia, así el formato que cumplió con estas características fue el formato *genepept*.

Se diseñó el script en PERL *extract1.pl* para extraer la información más relevante sobre la secuencia proteica, como taxonomía del organismo, bibliografía reportada, nombre del gen y número de aminoácidos, por mencionar algunos datos. Además, *extract1.pl* crea y formatea archivos para su uso directo en los programas clustalW y formatDB (programa de formateo de bases de datos para Blastp); en la generación de bases de datos en el gestor MySQL; así como para el análisis de secuencias por parte del investigador. Estas ventajas nos ayudaron a tener un rápido y fácil acceso a los datos.

V.1.B. Filogenia de la familia CHR

Nuestro grupo de datos consistió de 135 secuencias homólogas, en este conjunto se encontraron las 10 proteínas identificadas previamente por Nies y cols. (1998) como integrantes de la familia CHR, las cuales son consideradas como las proteínas modelo de dicha familia, por lo que podemos concluir que nuestra búsqueda se efectuó correctamente. Se observó que todas las fichas de dichas proteínas estaban repetidas, bajo distintos números de acceso, por lo que estas fichas fueron eliminadas al inicio de los pasos de limpieza, lo cual reafirma el problema de la redundancia de datos, además de que los scripts usados cumplieron su cometido exitosamente.



Una vez recabado nuestro juego de datos, hicimos una serie de alineamiento de todas las secuencias, en los que se observó que en la región media de las proteínas de cadena larga había una región de baja similitud, resultado que concuerda con la distribución de tamaño de proteína observada (Figura 8). Estos datos confirman el hecho de que las proteínas de la familia CHR, así como otras familias de transportadores (Saier, 2000; 2003), pueden contener uno o dos dominios homólogos a ChrA, de alrededor de 200 aminoácidos cada uno (Nies y cols., 1998). Este resultado también sugiere que el origen de las proteínas de cadena larga LCHR, por ejemplo ChrA, fue por un evento de duplicación génica y una posterior fusión de los dos genes, para dar como resultado una proteína LCHR con dos dominios (Nies y cols., 1998).

Es pertinente señalar que se identificó la presencia de una proteína de cadena muy larga (906 aa) en el hongo *Gibberella zeae* (XP_384019); esto es debido a que esta proteína contiene un dominio extra de 451 aminoácidos en el extremo carboxilo, el cual tiene similitud con el dominio DadA, encargado de la oxidación de D-aminoácidos y glicina, y que es completamente distinto a ChrA. Este caso es único ya que ninguna otra proteína CHR tiene unido un dominio diferente; todas las proteínas de cadena larga contienen sólo los dos dominios duplicados ya mencionados. Para los análisis posteriores se procedió a cortar el dominio de DadA de la proteína XP_384019, dado que para fines evolutivos y de conservación de aminoácidos de la familia de ChrA, no aportaría información relevante.

Ya que no existe un método ideal para reconstruir todos los árboles filogenéticos, y además, cada método tiene sus ventajas y sus deficiencias, se recomienda el uso de distintos algoritmos filogenéticos, lo que permite obtener con certeza un consenso más real, que el que se obtendría si se usara sólo un algoritmo (Baldauf, 2003). En este trabajo se utilizaron cuatro algoritmos distintos, englobando las dos metodologías más usadas: los métodos basados en distancia (MBD) y los métodos basados en caracteres (MBC). Los métodos UPGMA, NJ y ME son MBD y MP es un MBC.

V.1.B.i. Reconstrucción de la filogenia

Los cuatro algoritmos nos dieron árboles muy semejantes entre sí, aunque presentaron algunas variaciones menores en la topología (Figuras 12-14); sin embargo, esto es de esperarse ya que los métodos pesan de distinta manera las características de las secuencias proteicas. Un hecho que nos da más seguridad de la veracidad del árbol, es que los grupos generados en los cuatro métodos fueron prácticamente los mismos. En



la Figura 10 podemos ver la distribución de los dominios amino y carboxilo de las proteínas LCHR y de las proteínas SCHR. La separación obtenida en los cuatro métodos de las proteínas SCHR de los dominios de amino y carboxilo de las proteínas LCHR nos indica fuertemente la posible paralogía de las proteínas de cadena pequeña con respecto de las proteínas de cadena larga, de las cuales, éstas últimas formarían los verdaderos transportadores de cromato de la familia CHR.

El hecho de que los dominios amino y carboxilo de las proteínas LCHR se agrupen de manera independiente, nos indica la probable separación evolutiva de los dos dominios. Se postula que las proteínas mayores, como sucede dentro de miembros de la MFS, surgieron de un evento de duplicación génica (Rubin y cols., 1990; Pao y cols., 1998), por lo que se esperaría que los dos dominios de una misma proteína se encontraran relativamente cerca, cosa que no ocurre. Este hecho indica que muy probablemente el evento de duplicación se haya dado hace mucho tiempo, y que el gen duplicado, ahora formando los dos dominios, haya evolucionado a velocidades distintas, provocando que el grado de conservación de los dominios sea diferente, y por tanto, que la distribución de los mismos dentro del árbol filogenético sea por separado.

Un hecho muy importante es que el dominio amino de las proteínas de cadena larga se agrupa en un solo grupo, mientras que el dominio carboxilo lo hace en dos grupos, aparentemente separados, lo que muestra que la secuencia del dominio amino de las proteínas de cadena larga LCHR está más conservado que el dominio carboxilo. Una observación interesante, que se advierte en la Figura 31, es la distribución de los dominios amino en el árbol de análisis de dominio, en un solo grupo. El grupo muestra casi la misma topología que el árbol de proteínas LCHR cuando se analizaron de manera individual (Figura 16); la única gran diferencia es la localización del grupo de bacterias del género *Burkholderia*, cuyo dominio amino parece derivar de las SCHR (ver Figura 31), lo que sugiere un evento de sustitución de dominios.

Esto último parece correlacionar con el hecho de que en experimentos de mutagénesis al azar de ChrA de *P. aeruginosa*, donde se seleccionaron mutantes susceptibles a cromato, el dominio amino parece ser más sensible a las mutaciones que el dominio carboxilo (Aguilera y cols., 2004). Estos datos en conjunto nos permiten proponer, a manera de hipótesis, que el dominio amino (menos variable) es el responsable del transporte y que el dominio carboxilo tiene más bien un papel auxiliar, como por ejemplo, contribuir a la estabilidad de la proteína en la membrana.



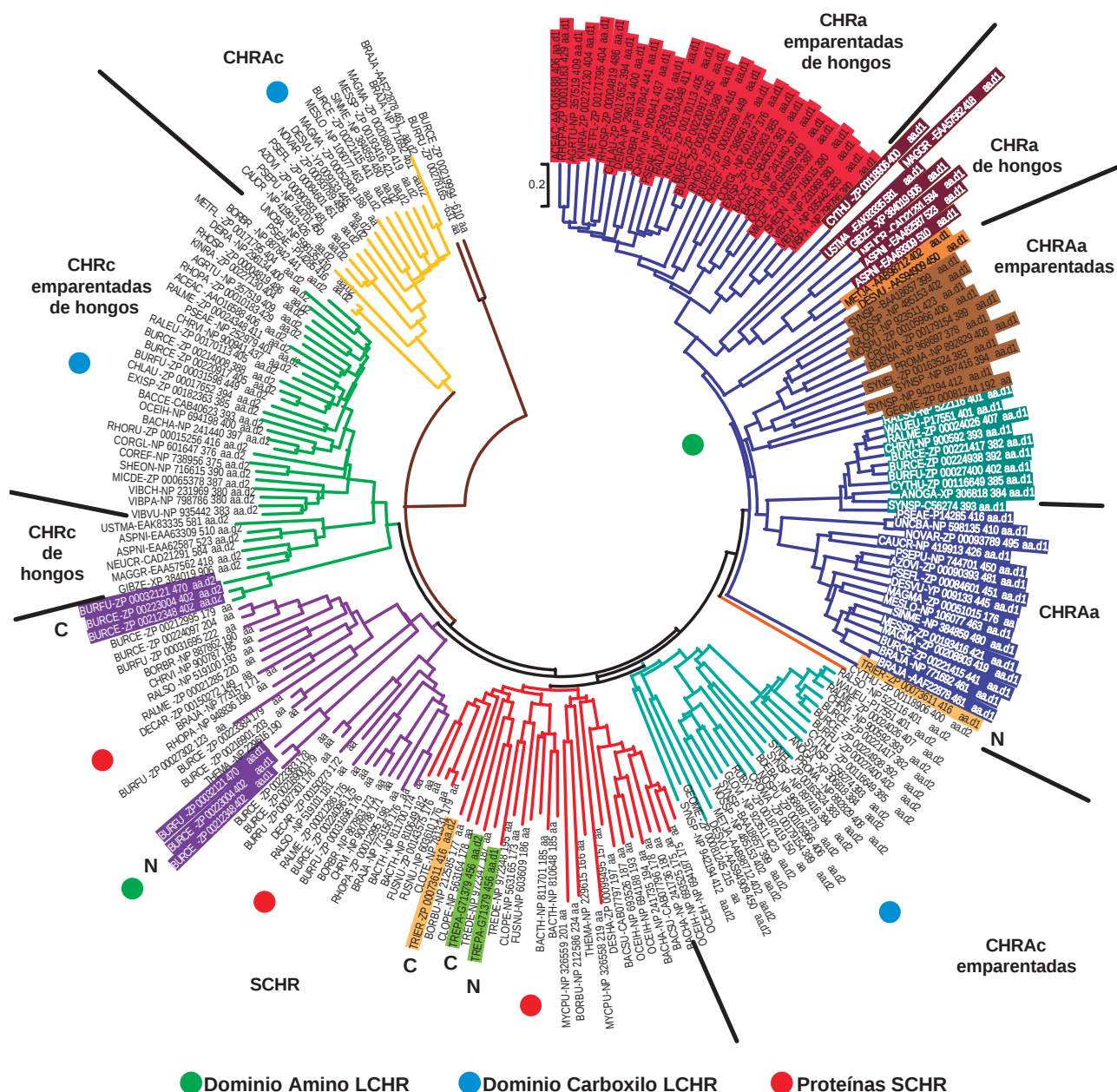


Figura 31 – Árbol filogenético de la familia CHR. En las ramas se marcan los grupos obtenidos en el árbol cuando se analizaron los dominios de forma independiente, mientras que se marca el fondo de las secuencias que se agruparon de esa manera en el análisis de las proteínas mayores. d1-domínio amino, d2-domínio carboxilo. Las leyendas indican el nombre de los grupos. Árbol generado con el algoritmo de NJ.



Es interesante hacer notar también que la topología con que se agruparon los dominios amino y carboxilo de las proteínas LCHR resultó muy similar (con la salvedad de que las ramas que representan los dominios carboxilo en el árbol se observan distribuidas en dos grupos separados) aunque, como se mencionó antes, separados de las SCHR (Figura 31); esto sugiere un origen monofilético para la mayoría de las proteínas LCHR. Así, el evento primario de duplicación y fusión de genes, que posiblemente dio origen a la mayoría de las LCHR, ocurrió hace mucho tiempo, por lo que es probable que las SCHR actuales correspondan a un grupo parálogo a las proteínas bidominio LCHR, esto es, que posean una función diferente al transporte de cromatos.

La presencia de una proteína homóloga a ChrA en una arquea (*Methanococcus jannaschii*), y la ausencia de datos que sugieran su presencia como resultado de una transferencia horizontal, sugiere que la familia CHR corresponde a un grupo de proteínas muy antiguo. Puesto que a la fecha se ha terminado la secuencia de 19 genomas completos de arqueas, se podría esperar la presencia de más proteínas de la familia CHR entre los miembros de ese dominio, lo cual no ocurrió. Esto puede deberse a que esta proteína haya sido adquirida por un evento de transferencia horizontal no muy reciente (Philippe y Douady, 2003), aunque no hay reportes que justifiquen tal evento; este resultado también puede deberse a la pérdida de genes, fenómeno ya reportado en eucariotes (Salzberg y cols., 2001).

V.1.C. HISTORIA EVOLUTIVA DE LA FAMILIA CHR

Algunos dominios amino y carboxilo de ciertas proteínas LCHR presentan una posición anómala dentro del árbol filogenético mostrado en la Figura 31, la cual puede ser explicada por fenómenos de duplicación, fusión y/o reemplazo de dominios en eventos ulteriores. Así, por ejemplo, los dominios amino y carboxilo de la proteína LCHR de *Treponema pallidum* (G71379) se encuentran agrupados junto con las proteínas SCHR, lo que sugiere fuertemente que su origen es un evento de fusión de dos genes de SCHR adyacentes.

En *Burkholderia fungorum* y *B. cepacia* R1808 y R18194, existen dos proteínas LCHR, una de las cuales presenta una posición anómala para su dominio amino (ZP_00032121, ZP_00223004 y ZP_00212348), el cual se agrupa junto con las SCHR; su dominio carboxilo, en cambio, se agrupa junto con el resto de los dominios carboxilo de las LCHR (de cadena larga), lo que sugiere que el dominio amino en estas tres proteínas fue sustituido por una SCHR. De la misma forma, el dominio carboxilo de la proteína



LCHR de *Trichodesmium erythraeum* (ZP_00073611) se agrupa dentro de las SCHR, no así su dominio amino, que se agrupa junto con el resto de los dominios amino de las LCHR; lo que sugiere en este caso que el dominio carboxilo original de la CHR de *T. erythraeum* pudo ser reemplazado por un dominio SCHR.

El hecho de que estos últimos casos, en donde una secuencia del tipo SCHR se encuentra presente en proteínas LCHR, y considerando que presentan una distribución filogenética tan limitada, permite suponer que corresponden al resultado de eventos secundarios relativamente recientes, en comparación con la duplicación y fusión primaria que dio origen al resto de las LCHR, por lo que pueden considerarse como variantes quiméricas de las CHR de cadena larga *bona fide*. Este hecho se resalta dado que las CHR de cadena larga, que resultan de la sustitución de uno de sus dominios por una proteína del tipo SCHR, se colocaron dentro del árbol de la Figura 16 en ramas separadas del resto de las CHR de cadena larga, y están identificadas en dicho árbol de manera genérica como CHR quiméricas.

Por otra parte, es interesante hacer notar que algunas proteínas SCHR o de cadena corta presentan también una posición anómala dentro del árbol filogenético mostrado en la Figura 31, ya que se encontraron agrupadas dentro de los dominios amino o carboxilo de las LCHR. Así, por ejemplo, en *Geobacter metallireducens* y en *Magnetospirillum magnetotacticum* existen dos proteínas de cadena corta o SCHR, las cuales se agrupan dentro de los dominios amino (ZP_00081244 y ZP_00051015) o carboxilo (ZP_00081245 y ZP_00052808) de las proteínas de cadena larga o LCHR. La existencia de proteínas SCHR con una alta identidad y similitud a los dominios amino y carboxilo de las LCHR sugiere que su origen es a través de la escisión de una proteína de cadena larga LCHR, y apoya indirectamente la idea de que estas últimas pudieron originarse a partir de la fusión de dos proteínas de cadena corta para formar proteínas bidominio, ya que si los dominios de las LCHR pueden escindirse, implica también que pueden fusionarse. Sin embargo, cabe la posibilidad de que estas proteínas SCHR deriven directamente de una CHR ancestral (monodominio), aunque como no se agrupan entre sí, es más probable que representen una fragmentación ulterior de una LCHR.

Es pertinente señalar que en muchas otras familias de proteínas, incluyendo la superfamilia de transportadores MFS, también se ha documentado la fusión de dominios para generar proteínas de mayor tamaño (Pao y cols., 1998). En la Figura 32 se ilustran los posibles eventos de duplicación, fusión y reemplazo de dominios que dieron origen a las diversas proteínas que forman la macrofamilia CHR.



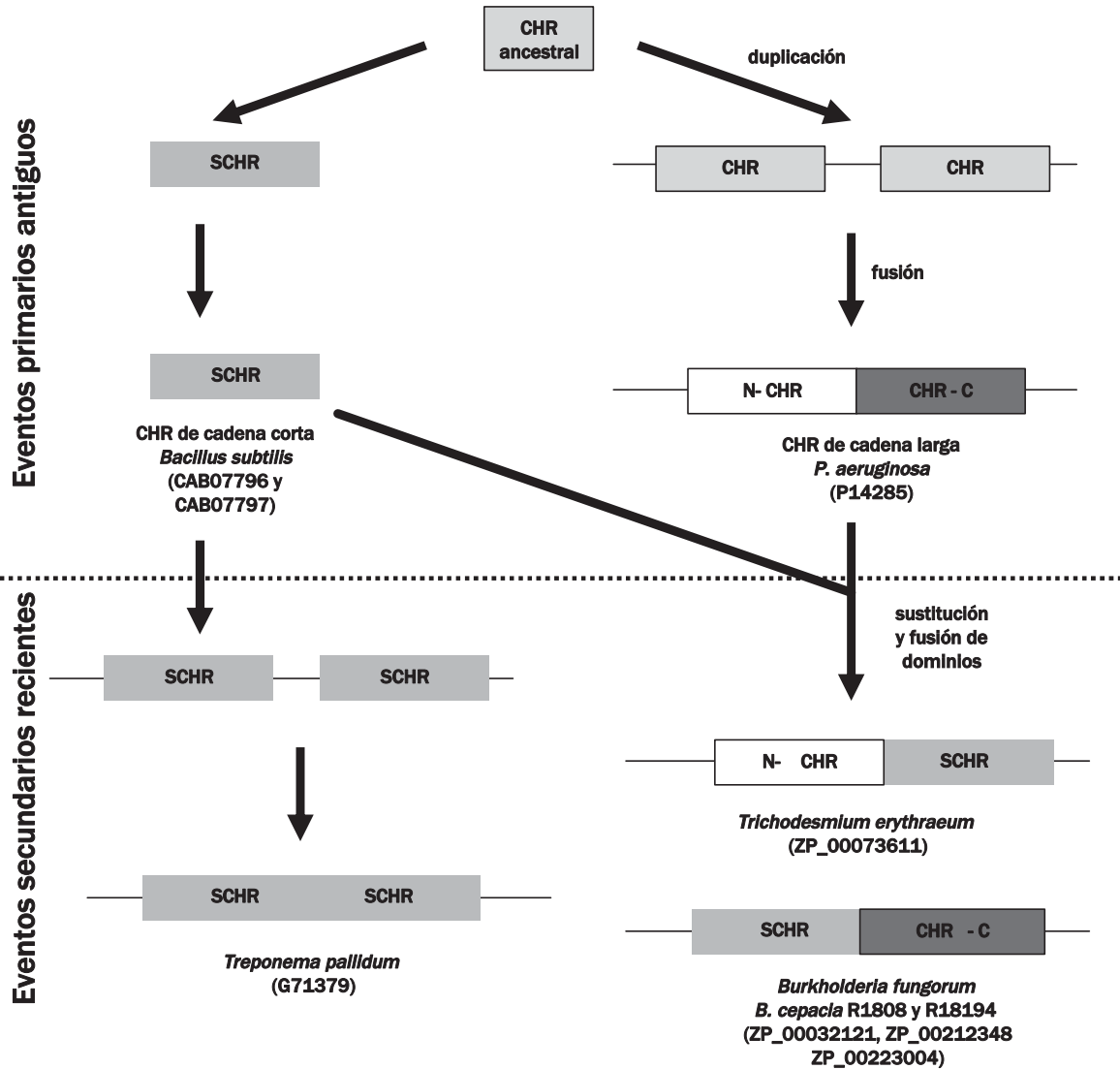


Figura 32 – Posibles eventos de duplicación, fusión y reemplazo de dominios que dan origen a las diversas proteínas que forman la familia CHR. En el texto se explican con detalle dichos eventos.

V.1.D. LA MACROFAMILIA CHR

Una interesante cuestión después de llevar a cabo la filogenia de la familia CHR es decidir qué nivel taxonómico, dentro de la clasificación de proteínas, debe tener este grupo de proteínas. Hasta el momento hemos visto que hay dos grupos de proteínas claramente definidos, las SCHR y las LCHR, que comparten un ancestro común y que además tienen gran similitud en su secuencia de aminoácidos. Sin embargo, es probable que estos dos grupos de proteínas lleven a cabo funciones distintas (Nies y cols., 1998), y sólo se ha probado funcionalidad en dos de los subgrupos de las proteínas LCHR (CHRA y emparentados a CHRA). En la propuesta de Riveros-Rosas (2004), se definen los siguientes tres grupos taxonómicos para la clasificación de proteínas:

- 1) Subfamilias, que consisten en grupos cerrados, conformados idealmente por conjuntos de proteínas ortólogas que desarrollan la misma función.
- 2) Familias, cada una compuesta por un conjunto monofilético de subfamilias que poseen una identidad significativa al comparar sus secuencias entre ellas y que pueden o no, compartir sustratos y/o mecanismos de reacción
- 3) Macrofamilias, cada una compuesta por un conjunto monofilético de familias de proteínas presentes en los tres dominios en que se distribuyen los seres vivos, e incluye al menos a una subfamilia de proteínas que posee una actividad perteneciente a una vía metabólica muy antigua.

Esta clasificación es compatible con las clasificaciones ya establecidas basadas en el fenotipo o la estructura de las proteínas, como la clasificación SCOP (Lo Conte y cols., 2000), o CATH (Orengo y cols., 1997). Por tanto, las proteínas CHR, si tomamos en cuenta la clasificación CATH, serán parte de una clase, arquitectura, topología (13 STM's), superfamilia homóloga y superfamilia. Además, de acuerdo a esta propuesta CHR formará en sí una macrofamilia, integrada a su vez por dos familias, la familia LCHR y la familia SCHR, cada una constituida por sus propias subfamilias (Figura 33).

En el caso de la familia LCHR, podemos definir a los grupos CHRA y emparentados a CHRA como una subfamilia, ya que es probable que compartan la función, además de que hemos podido establecer homología entre las secuencias de ChrA de *P. aeruginosa* y de *A. eutrophus* (prr score = 10^{-15}), según el método propuesto por Saier (Saier, 1994).



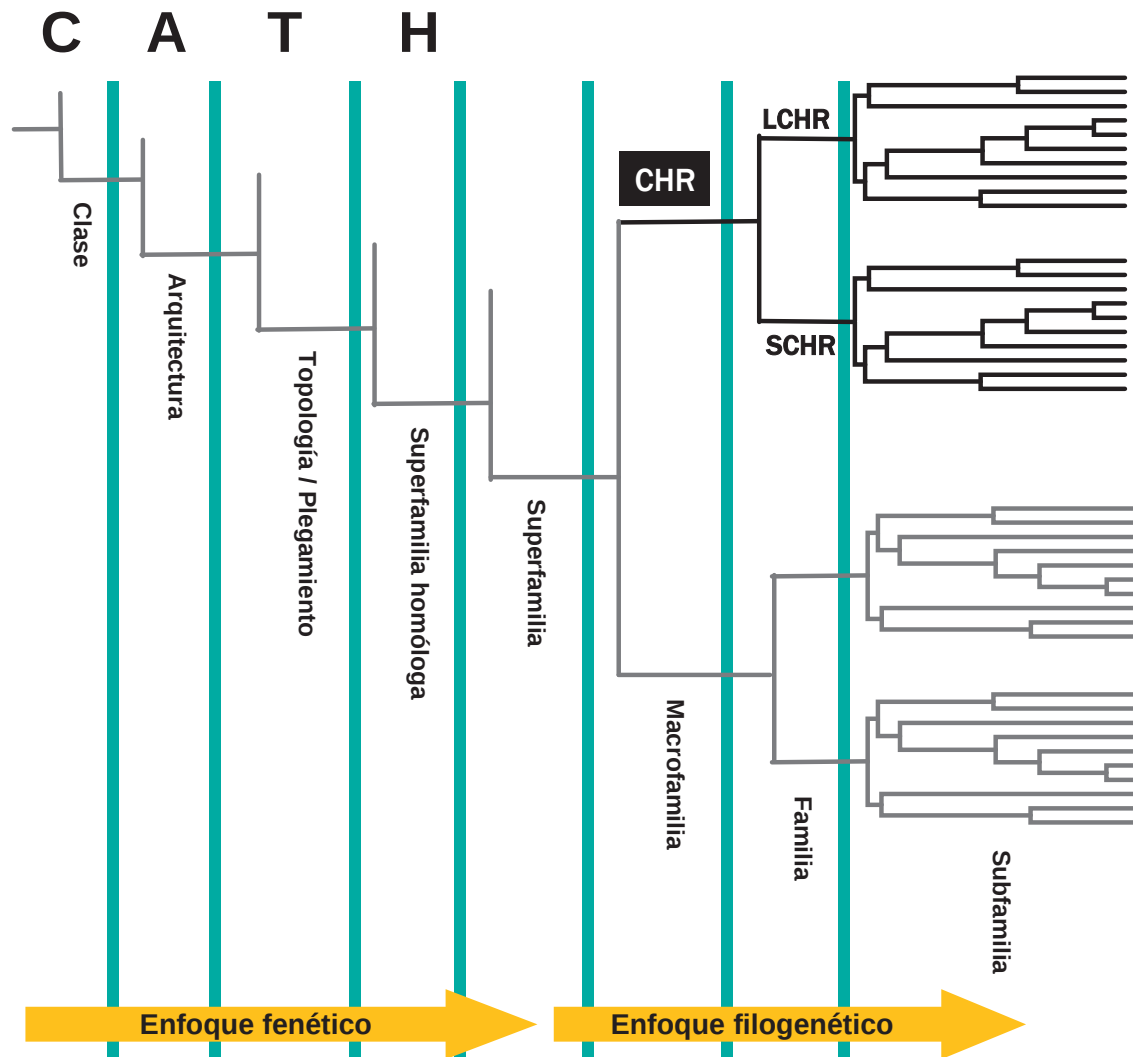


Figura 33 – Taxonomía de la familia CHR. Se muestra la ubicación de la macrofamilia CHR, la cual cuenta con dos familias parálogas, una formada por las proteínas LCHR y otra formada por las proteínas SCHR.

V.2. IMPORTANCIA DE LOS AMINOÁCIDOS CONSERVADOS DE LA MACROFAMILIA CHR

Una vez obtenida la filogenia de la macrofamilia CHR, procedimos a identificar cinco de los aminoácidos más conservados para, mediante mutagénesis dirigida, probar si éstos son relevantes para que la proteína ChrA de *P. aeruginosa* lleve a cabo correctamente su función.

Se seleccionaron las glicinas 44, 45, 92, 100, y el triptófano 63. Las mutaciones se decidieron hacer de acuerdo a los siguientes razonamientos: primero, por el aminoácido más parecido al original, con el fin de evaluar la importancia *per se* del residuo a mutar; segundo, por un aminoácido con un tamaño mayor o menor, según sea el caso del aminoácido nativo, para evaluar si el tamaño del aminoácido es vital para la función de la proteína; tercero, usar un aminoácido de igual tamaño pero con diferente polaridad al aminoácido presente en la proteína, para tratar de discernir si la polaridad puede afectar el papel del residuo seleccionado en la proteína (Figura 19).

V.2.A. Mutaciones en residuos de Triptófano

El triptófano en muchos sentidos es un aminoácido especial; es el que tiene la mayor superficie, es el más voluminoso y es el único que presenta un grupo indol, cuyo átomo de nitrógeno puede servir como donador en puentes de hidrógeno, además de favorecer el contacto con grupo polares (Samanta y cols., 2000). Se ha reportado que en las proteínas transmembranales los triptófanos se ubican preferencialmente en el periplasma (Schiffer y cols., 1992). Este residuo puede estar jugando papeles estructurales debido a sus características hidrofóbicas (dadas por su anillo aromático) y a su volumen (Reece y cols., 2002). Por ejemplo, puesto que es el aminoácido con la mayor afinidad hacia la interfase agua-membrana, puede anclar la proteína en la membrana (Ridder y cols., 2000), o bien puede ser importante en la unión a ligandos específicos (por ejemplo cannabinoides o azúcares) cuando se localiza en medio de un STM (Rhee y cols., 2000; Vázquez-Ibar y cols., 2003) y cuando se encuentra en asas acompañado de aminoácidos cargados (Chen y cols., 2001).

V.2.A.i. Triptófano 63

El W63 se encuentra localizado en el asa periplásmica P1 en la proteína ChrA (Figura 34A) y está conservado en un 82% en de los 212 dominios homólogos de ChrA. La gran conservación de este aminoácido en todas las proteínas de la macrofamilia CHR nos



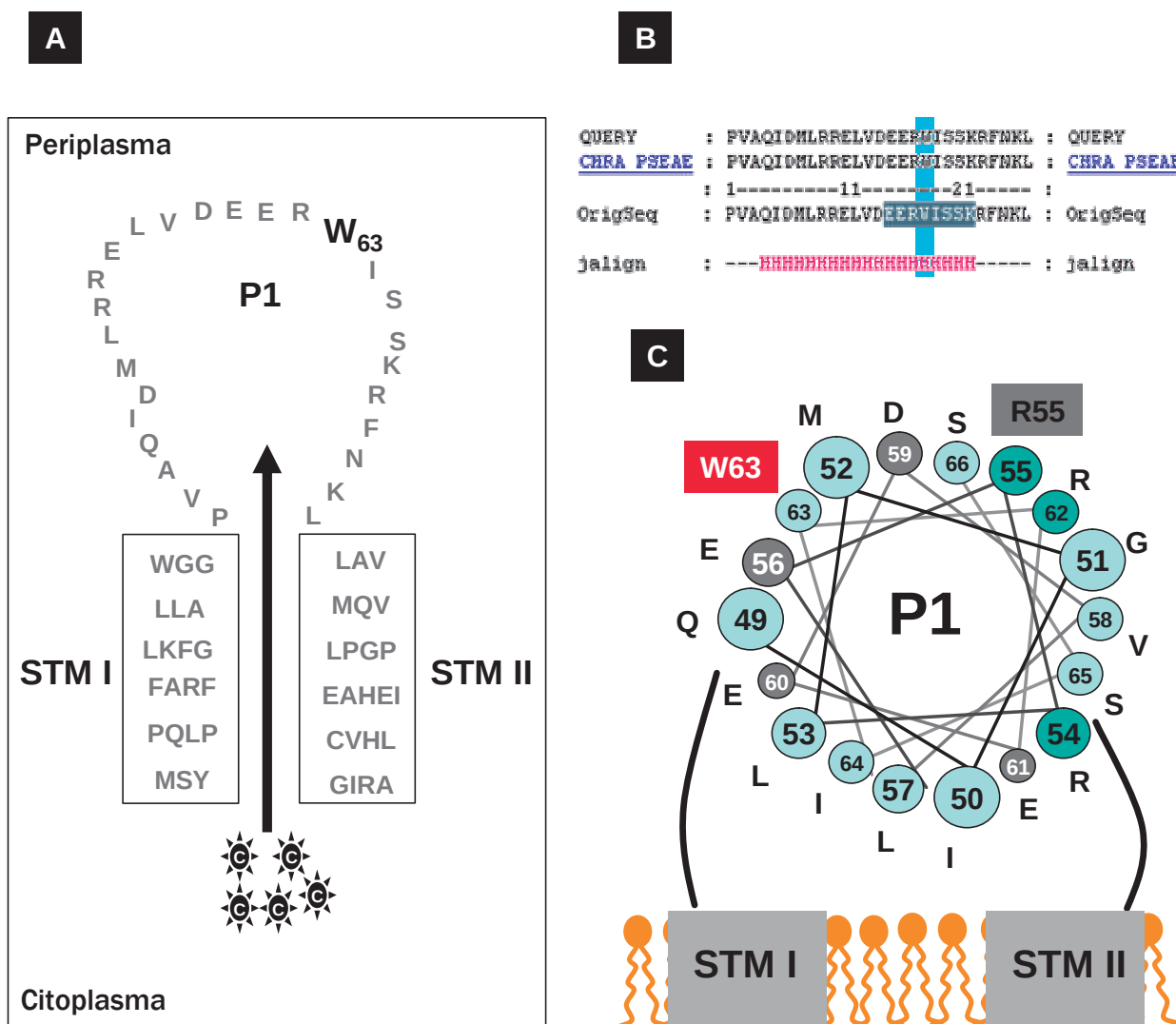


Figura 34 – Posible papel Importancia del triptófano 63 en la proteína ChrA de *P. aeruginosa*. A) Esquema del asa periplásmica 1 (P1) de la proteína ChrA y de los STM I y II, los cuales son conectados por P1. Se resalta el triptófano 63. La flecha indica la vía de expulsión del cromato (■). B) Predicción de la estructura secundaria del asa periplásmica P1, hecha con el programa jpred (Cuff y cols., 1998). Las H's indican que se trata de una hélice alfa. C) Proyección helicoidal de la posible hélice alfa que contiene al triptófano 63 (enmarcado en el cuadro negro). En fondo oscuro y con letra negra se marcan los aminoácidos con carga positiva (R) y con letra clara aquellos con carga negativa (D, E). Se enmarca también la arginina 55, otro residuo importante para la función de la proteína ChrA (Moreno, 2005).

sugiere que debe de jugar un papel importante en la proteína ChrA. Las dos mutantes obtenidas en el triptófano (W63F y W63Y) presentaron susceptibilidad a cromato (Figura 28). El papel del triptófano en ChrA no pudo ser compensado por el cambio del aminoácido más parecido a éste, la fenilalanina (la cual es aromática pero no cuenta con un grupo indol), ni por la tirosina (un aminoácido aromático y polar). Esto nos permite establecer la gran importancia del triptófano en la correcta función de la proteína ChrA.

Hay una gama de posibles explicaciones sobre la función del W63 dentro del asa periplásmica 1. P1 forma una hélice alfa extramembranal; así, se formaría una hélice periplásmica extramembranal, la cual conecta los STM's I y II (Figura 34B). Primeramente podemos especular sobre su participación en una hélice que actúe como "cerradura" del canal; es decir, que forme una estructura donde una señal indique la abertura o el cerrado del transportador. Este fenómeno se ha reportado en los quimiorreceptores Tar de *E. coli* y *S. typhimurium*, donde se ha observado que los grupos R del triptófano cercanos a la interfase membrana-agua intervienen en la actividad y sensibilidad del receptor, controlando la transición de apagado a encendido (Miller y Falke, 2004; Draheim y cols., 2005). Además, se ha visto que estructuras extramembranales de distintos canales y transportadores intervienen en los procesos de activación e inactivación de canales iónicos (Guo y cols., 1995; Bezanilla, 2000).

De la misma forma, el grupo indol del triptófano tiene la capacidad de interaccionar con residuos cargados, como se ha visto en distintas proteínas membranales (Chen y cols., 2001; Okada y cols., 2001). El W63 de ChrA se encuentra rodeado de algunos aminoácidos cuyo grupo R presenta carga positiva o negativa (Figura 34C), por lo que su función probablemente sería a través de la interacción con algunos de estos residuos, como un mecanismo de acción de abertura o cerradura del transportador. Esta hipótesis toma fuerza debido a que la mutación de la arginina 55 de ChrA, vecina del W63 (Figura 34C), causa disminución de la resistencia a cromato señalando a este residuo como esencial (Moreno, 2005).

Otra interesante posibilidad es que el triptófano actúe como estabilizador en la larga asa periplásmica donde se encuentra (Figura 34). Se ha reportado que los grupos R del triptófano, debido a sus propiedades aromáticas, anclan los segmentos transmembranales en la membrana; así, el W puede quedar interaccionando en la interfase membrana-agua (Ridder y cols., 2000). Además, en simulaciones y experimentos *in vivo* con hélices aisladas de distintas proteínas, las hélices con residuos



de triptófano en la interfase membrana-agua fueron más estables que aquellas que carecen de estos residuos (Woolf, 1997; Braun y von Heijne, 1999).

V.2.B. Mutaciones en residuos de Glicina

La principal característica estructural de la glicina es que tiene el grupo R más pequeño de los aminoácidos, un solo hidrógeno (Javadpour y cols., 1999). La glicina se encuentra con alta frecuencia en las hélices transmembranales (Landolt-Marticorena y cols., 1993), lo cual indica que probablemente tiene un papel estructural diferente al que tiene en las proteínas globulares, el cual es participar en la formación de asas, giros β y rompiendo hélices (Javadpour y cols., 1999). Se ha atribuido que las glicinas participan en las interacciones hélice-hélice de proteínas de membrana, orientando las hélices de manera correcta (Javadpour y cols., 1999; Russ y Engelman, 2000).

En este trabajo se observó que la mayoría de las mutaciones en las glicinas seleccionadas (G44L, G44S, G45A, G45L, G45S, G92A, G92L, G92S, G100A y G100S) abatieron completamente la resistencia a cromato (Figuras 26, 27, 29, 30), lo cual nos indica que las glicinas son importantes para que ChrA pueda conferir resistencia a este oxianión.

V.2.B.i. Glicinas 44 y 45

Los residuos G44 y G45, localizados en la salida del STM I (Figura 35A), están conservados en un 96 y 89%, respectivamente; aunado a que son adyacentes nos hace pensar que forman un motivo que juega un papel determinante en la proteína ChrA. Esto se demostró dado que de las seis mutantes analizadas cinco perdieron por completo su capacidad de conferir resistencia y sólo una (G44A) la perdió parcialmente (Figura 26).

En el caso de la glicina 44, la mutación por alanina conservó en parte la resistencia a cromato, sin embargo, los cambios por leucina y serina abatieron por completo la resistencia (Figura 26). Por lo tanto, podemos señalar el importante papel de este residuo en el correcto funcionamiento del transportador. Por otra parte, las tres mutantes de la glicina 45 (G45A, G45L y G45S) perdieron totalmente la resistencia a cromato, por lo que podemos concluir que la glicina 45 es un aminoácido absolutamente esencial para la correcta función de la proteína ChrA.

La glicina 44 está más conservada dentro de la macrofamilia CHR que la glicina 45 (96 y 89%, respectivamente) y, sin embargo, en la proteína ChrA de *P. aeruginosa*, la glicina 44 es capaz de soportar parcialmente un cambio sutil (Ala) sin que se pierda la



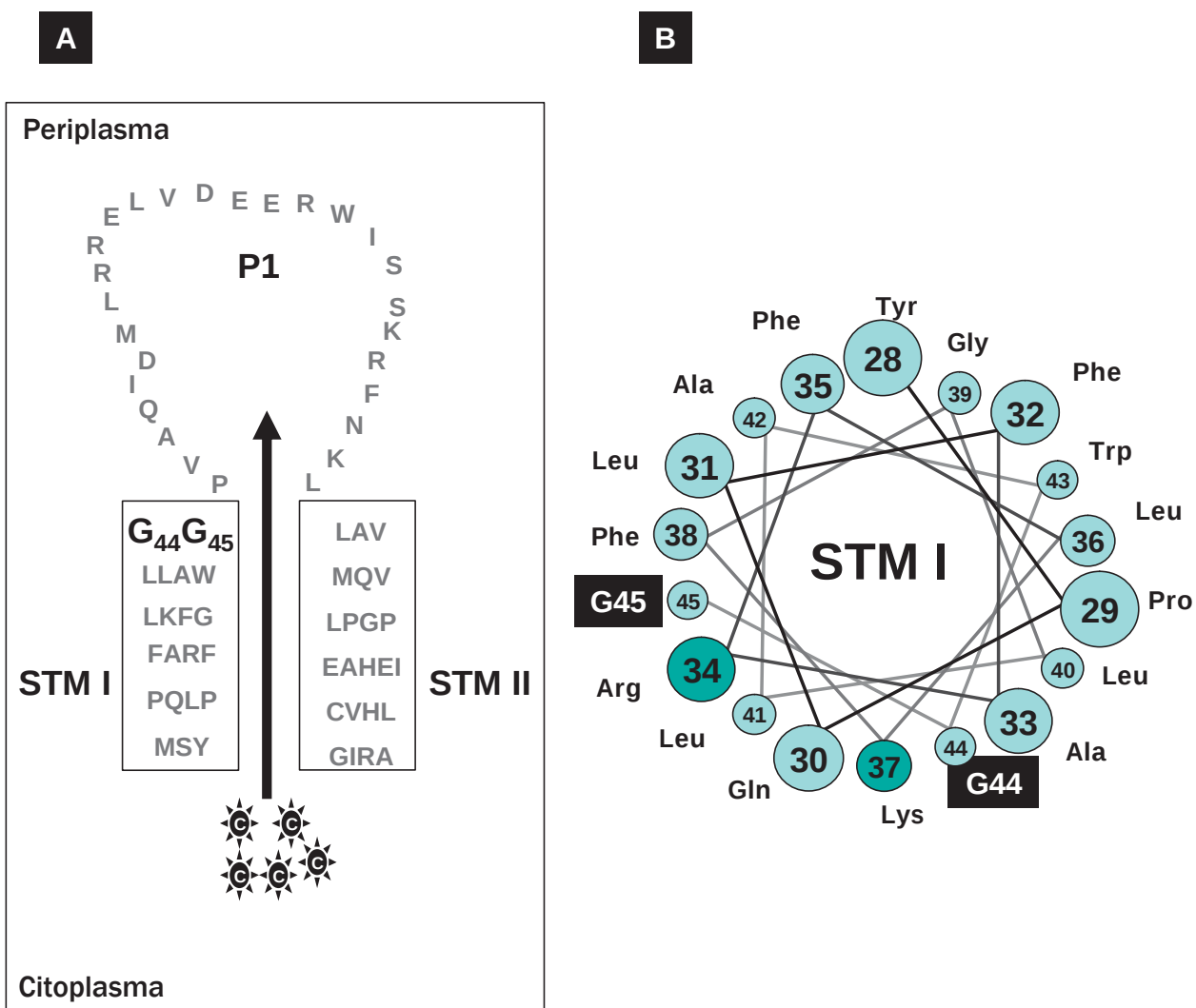


Figura 35 – Posible papel de las glicinas 44 y 45 en la proteína ChrA de *P. aeruginosa*. A) Esquema de los STM I y STM II de la proteína ChrA, los cuales son conectados por el asa P1. Se resaltan las glicinas 44 y 45. La flecha indica la vía de expulsión del cromato (▣). B) Proyección helicoidal de la hélice alfa que contiene a las glicinas 44 y 45 (enmarcadas en un cuadro negro). En fondo oscuro y con letra negra se marcan los aminoácidos con carga positiva (Arg y Lys).

función, y no así con la glicina 45. Este resultado presenta una aparente contradicción, ya que esperaríamos que las mutantes en la glicina 44 (sobre todo G44A) se comportaran al menos tan sensibles como la glicina 45. Este hecho podría deberse a múltiples factores, por ejemplo, la proyección helicoidal del STM I (Figura 35B) muestra que la G44 se encuentra cercana a una alanina (A33; volumen = 88.3 Å³) y a una lisina (K37; volumen = 168.6 Å³), por lo cual quizá al introducir un aminoácido con grupo R más voluminoso en esta región de la proteína no se alteraría demasiado la estructura de la proteína, y la G44 podría llevar a cabo su función de manera parcial. Por otra parte, la G45 se encuentra cercana a dos aminoácidos con un grupo R más voluminoso, una fenilalanina (F38; volumen = 189.9 Å³) y una arginina (R34; volumen = 177.4 Å³), lo cual quizá provocaría que al introducir otro aminoácido con un grupo R más grande que el original, se alterara la estructura de la proteína y por consiguiente la función de la misma.

Otra posible explicación es que la glicina 44 otorgue otros beneficios diferentes a la proteína, por ejemplo, que participe en la formación de estructuras que confieran estabilidad en otras condiciones, como alta temperatura, salinidad o presión osmótica, como sucede en la lisozima del fago T4, una proteína globular, en donde mutaciones puntuales en residuos de glicina no afectan la función de la proteína y sin embargo alteran su estabilidad (Nicholson y cols., 1992).

El posible papel de las glicinas dentro de los STM's de las proteínas estudiadas, se ha relacionado con una función estructural, debido a que su grupo R es sólo un hidrógeno. En muchos de los casos las glicinas se han asociado formando parte de interacciones hélice-hélice (Lemmon y Engelman, 1994; Jessen-Marshall y cols., 1995; Javadpour y cols., 1999; Green y cols., 2000; Russ y Engelman, 2000; Fleming y Engelman, 2001; Curran y Engelman, 2003; Aller y cols., 2004). Así, las glicinas G44 y G45 se encuentran en una secuencia parecida al motivo GXXXG (Figura 18; motivo GLLAWG₄₄G₄₅), el cual es importante para este tipo de interacciones (Russ y Engelman, 2000; Fleming y Engelman, 2001; Senes y cols., 2004); asimismo, en la glicoporina A, dos residuos de glicina separados por cuatro residuos son un elemento esencial para un motivo de unión entre STM's (Lemmon y Engelman, 1994).

Igualmente se ha observado el importante papel de glicinas adyacentes en otras proteínas membranales. Así, mutaciones en las glicinas adyacentes del intercambiador de Na⁺/H⁺ NHE1 y en los transportadores de hierro Nramp1 y Nramp 2, cambian las propiedades de los transportadores, por lo que G44 y G45 pueden estar involucradas en favorecer cambios conformacionales en el STM I (Su y cols., 1998; Wakabayashi y cols.,



2003). Otra explicación de la carencia de función en las mutantes G44 y G45, puede ser por la pérdida de un doblez formado por la presencia de dos glicinas y una prolina adyacentes (Figura 18) en esa zona, como se ha visto en la proteína membranal citocromo c oxidasa (Javadpour y cols., 1999).

V.2.B.ii. Glicina 92 y 100

La G92, localizada en la salida del STM II, se encuentra conservada en un 96% de los dominios de ChrA, mientras que la G100 presenta una conservación del 92% en las 212 secuencias de los dominios de ChrA y está localizada en el STM III (Figura 36A). Las tres mutantes de la glicina 92 (G92A, G92L y G92S) fueron incapaces de ser resistentes al cromato (Figura 29), por lo que podemos concluir que G92 resultó ser un residuo absolutamente esencial para la proteína ChrA.

En el caso de la glicina 100 podemos observar que tiene un papel importante dentro de la proteína ChrA, ya que la mutación por serina, un aminoácido con grupo R polar, abate por completo la resistencia a cromato. Sin embargo, la mutante por el residuo más parecido, alanina, se comportó como el control resistente (Figura 30). Estos resultados indican que la G100 es importante, mas no esencial, para la correcta función de la proteína ChrA de *P. aeruginosa*.

En la Figura 36B, la proyección helicoidal del STM II muestra que la glicina 92 y la glicina 81 (otro residuo esencial para la función de la proteína ChrA (Aguilar, 2005)), quedan en una misma cara de la proteína, quizá proveyendo una adecuada superficie de contacto para la interacción con otros STM's, como ya se ha mencionado antes (Lemmon y Engelman, 1994; Jessen-Marshall y cols., 1995; Javadpour y cols., 1999; Green y cols., 2000; Russ y Engelman, 2000; Fleming y Engelman, 2001; Curran y Engelman, 2003; Aller y cols., 2004). Es posible que la glicina 92 pueda estar promoviendo este tipo de interacciones dentro de la proteína ChrA. Además, se ha observado que la mutación en una glicina que sale del STM II causa pérdida de función del transportador de hidroxibenzoato PcaK de la MFS (Ditty y Harwood, 1999), por lo que la G92 puede ser requerida para una estructura de giro al inicio del asa C1.

En nuestro laboratorio se ha visto que la prolina 80, la glicina 81 y prolina 82, que forman el motivo PGP, con un 87% de identidad en la macrofamilia CHR, y que se encuentra también dentro del STM II, resultaron ser esenciales para la correcta función de la ChrA (Aguilar, 2005; Cortés, 2005). Se ha observado también que la función del STM II en el transportador LacY, proteína modelo de la MFS, es la de participar en los cambios



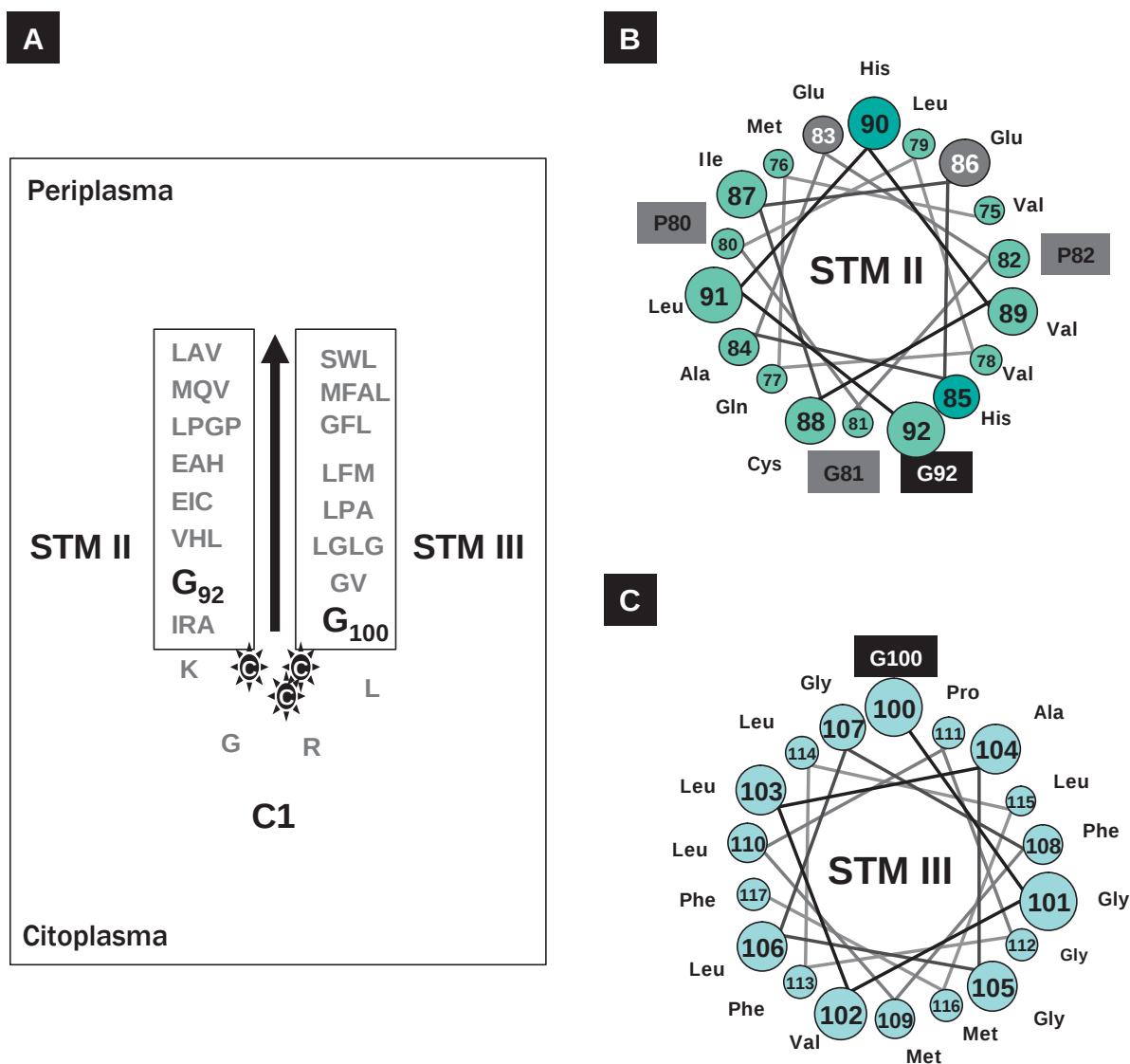


Figura 36 – Posible papel de las glicinas 92 y 100 en la proteína ChrA de *P. aeruginosa*. A) Esquema de los STM II y STM III de la proteína ChrA, los cuales son conectados por C1. Se resaltan las glicinas 92 y 100. La flecha indica la vía de expulsión del cromato (→). B) Proyección helicoidal de la hélice alfa que contiene a la glicina 92 (enmarcada en un cuadro negro). Se enmarcan la glicina 81 y las prolinas 80 y 82, otros residuos también esenciales para la función de la proteína ChrA (Aguilar, 2005; Cortés, 2005). En fondo oscuro y con letra negra se marcan los aminoácidos con carga positiva (His) y con letra clara aquellos con carga negativa (Glu). C) Proyección helicoidal de la hélice alfa que contiene a la glicina 100 (enmarcada en un cuadro negro).

conformacionales de la proteína, que sirven para que esta reconozca a la lactosa (Jessen-Marshall y Brooker, 1996; Green y cols., 2000). Por lo tanto, otro probable papel de la glicina 92 dentro de la proteína ChrA, es que actúe generando los cambios necesarios para que, junto con la glicina 82 y las prolinas 80 y 82, el STM II tome la forma correcta y pueda expulsar al cromato.

La glicina 100 se encuentra en un motivo muy conservado dentro de la rama del árbol filogenético donde se agrupó la proteína ChrA de *P. aeruginosa*, G₁₀₀GVLAG (Figura 18). Aunque esta secuencia sólo está conservada en dicho grupo (Figura 18), es parecida al motivo GXXXG, al que se ha atribuido gran importancia en las uniones entre STM's (Russ y Engelman, 2000; Fleming y Engelman, 2001; Senes y cols., 2004). Además, como vemos en la Figura 36C, la glicina 100 queda en la misma cara con la glicina 107, lo cual posiblemente genere una superficie de contacto adecuado para la interacción con otro STM, por lo que es muy probable que el papel de la glicina 100 sea la de promover la interacción del STM III con otro STM.

Las glicinas 92 y 100 forman parte del motivo G₉₂IRAKGRLG₁₀₀, parecido al motivo GXXX(D/E)(R/K)XG de la permeasa de lactosa, llamado motivo A y que se encuentra conservado en 17 familias de la MFS (Pao y cols., 1998). Se ha visto que mutaciones en las glicinas correspondientes a G92 y G100 provocan la pérdida de función en ese transportador (Jessen-Marshall y cols., 1995), por lo que estas glicinas podrían también jugar un papel importante en este motivo.

V.2.C. Modelo estructural de los STM I, II y III de ChrA de *P. aeruginosa*

La macrofamilia CHR es cercana a la superfamilia MFS, de cuyos al menos tres miembros se conoce su estructura tridimensional (Hirai y cols., 2002; Abramson y cols., 2003; Huang y cols., 2003), permitiendo confirmar los estudios bioquímicos previos donde se localizaron 12 STM's. En el caso de ChrA de *P. aeruginosa*, los análisis bioquímicos recientemente dirigidos a conocer la topología de la esta proteína han permitido establecer que consta de 13 STM's (Jiménez, 2005). A nivel estructural, estos resultados marcan una clara diferencia entre los dos grupos de proteínas; sin embargo, la secuencia del extremo amino es muy similar en ambos grupos, ya que conserva la orientación, contienen un motivo muy similar (motivo A). Esto se ve confirmado por la estructura topológica obtenida en nuestro grupo de trabajo, donde se ve claramente que el extremo amino de ChrA es muy similar en estructura al extremo amino de las proteínas de la superfamilia MFS (Jiménez, 2005) y al parecer están formados por 6 STM's.



Por lo tanto, de acuerdo a los resultados obtenidos en este trabajo, proponemos un modelo estructural y funcional basado a su vez en un modelo general propuesto a partir de las características de las proteínas de la MFS cuya estructura ya ha sido resuelta: LacY, GlpT y OxiT (Hirai y cols., 2003). Como se muestra en la Figura 37, el STM I estaría interaccionando con dos STM's más, con el STM IV, usando la superficie de contacto de la glicina 44, y probablemente el STM V (Hirai y cols., 2003), a través de la glicina 45. Este STM I estaría probablemente interaccionando con el cromato debido a la presencia de aminoácidos cargados (R34 y K37) en una de sus caras. El STM III sería un segmento que atravesaría perpendicularmente la membrana y serviría de soporte estructural (Hirai y cols., 2003); esto se reafirma debido a la presencia de residuos hidrofóbicos, los cuales son ideales para interaccionar con los fosfolípidos de la membrana. La glicina 100 estaría ayudando a la interacción con el STM II (Figura 37), además de ayudar a los cambios estructurales necesarios para expulsar el cromato.

El STM II estaría interaccionando con el STM III (Figura 37) (Hirai y cols., 2003), a través de las glicinas 92 y 81. Este STM II, como en la MFS, no es perpendicular a la membrana, sino tendría una curvatura característica (la cual podrían proveer las prolinas 80 y 82), lo que ayudaría a generar los cambios estructurales necesarios para la expulsión del cromato desde el citoplasma de la célula (Hirai y cols., 2003).



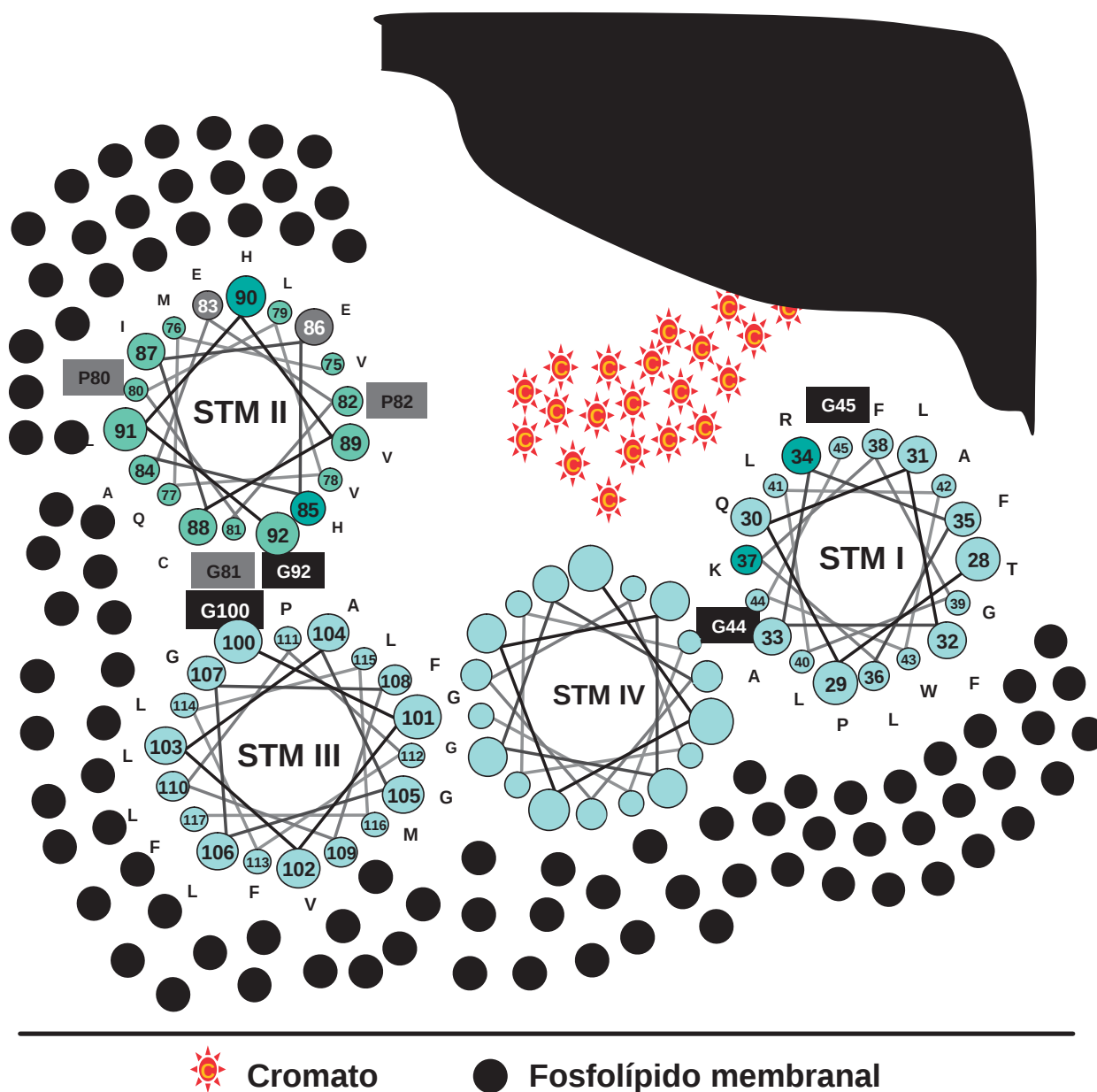


Figura 37 – Modelo propuesto de la interacción de los segmentos transmembranales I, II y III de la proteína ChrA de *P. aeruginosa*. El esquema se construyó de acuerdo a los modelos resueltos de las proteínas de la familia MFS. Las cabezas de los fosfolípidos (●) de la membrana rodean a las hélices que representan los STM's de ChrA. Se marcan también las moléculas de cromato (⌘) en la posible región del canal de expulsión.

VI. CONCLUSIONES

1. En este trabajo se identificaron un total de 135 secuencias de proteínas pertenecientes a la familia de proteínas transportadoras CHR y se llevó a cabo el primer análisis filogenético de esta familia, identificándola como la macrofamilia CHR.
2. La presencia de proteínas CHR (de cadena larga) en los tres dominios de los seres vivos (*Eubacteria*, *Archaea* y *Eukaryota*) permite suponer que se trata de una macrofamilia antigua y cuyo origen, por duplicación de monodominios CHR, es también muy antiguo.
3. La macrofamilia CHR está conformada por dos tipos de proteínas: (1) las de cadena corta o monodominio (denominadas SCHR), cuyo tamaño es de alrededor de 200 aa; y (2) las de cadena larga o bidominio (denominadas LCHR), con un tamaño de alrededor de 400 aa, cuya historia evolutiva probablemente ha estado sujeta a varios eventos de duplicación, fusión y/o sustitución de dominios, por lo que es posible que sus miembros participen en funciones adicionales al transporte de cromatos.
4. La distribución más limitada de las SCHR (de cadena corta) permite suponer que se trata de un grupo de proteínas parálogas derivadas también de una CHR monodominio ancestral; existen además LCHR quiméricas (en las cuales uno de sus dominios se sustituyó por una SCHR), por lo que en consecuencia, su origen es probablemente más reciente que el resto de las LCHR *bona fide* (de cadena larga).
5. Dentro de las LCHR, además de las proteínas quiméricas, existen tres grupos o subfamilias principales: la subfamilia CHRA, presente exclusivamente en bacterias y que contiene entre sus miembros a las únicas dos proteínas ChrA caracterizadas funcionalmente como transportadores de cromatos; la subfamilia de las CHR de hongos, que conforman el único grupo de proteínas de esta familia de transportadores presente en eucariontes y la subfamilia de las proteínas emparentadas a las CHR de hongos, las cuales están presentes en diversos tipos de eubacterias.
6. Dentro de la macrofamilia CHR se encontraron, en este trabajo, cinco aminoácidos muy bien conservados, las glicinas 44, 45 92 y 100 y el triptófano 63. Se predijo que al mutarlos por distinto aminoácidos semejantes en mayor o menor medida, abatirían los niveles de resistencia a cromato, mostrando así su importancia para la estructura o función de la proteína ChrA de *P. aeruginosa*.
7. Observamos que las glicinas 44 y 100 son importantes para que la proteína ChrA confiera resistencia a cromato, dado que un cambio por una aa polar (Ser) provocó susceptibilidad, pero permitieron el cambio por un aa parecido (Ala).



8. El triptófano 63 es importante ya que el cambio por aa muy similar (Phe) en esta posición abatió la resistencia a cromato.
9. Las glicinas 45 y 92 son esenciales para conferir resistencia a cromato dado que no permitieron ningún cambio sutil en esa posición.
10. Estos resultados resaltan la importancia de estos residuos dentro de la función de la proteína ChrA de *P. aeruginosa*, y también apoyan nuestra teoría de que el extremo amino de la proteína ChrA es el encargado de reconocer y llevar a cabo el transporte del cromato, mientras que el extremo carboxilo es posiblemente el encargado de dar estabilidad al transportador.

El análisis filogenético del transportador ChrA nos permitió predecir e identificar con éxito aminoácidos esenciales para su función, lo cual se demostró mediante la mutagénesis dirigida hecha en este trabajo. Este estudio nos ayuda a conocer más a fondo a nuestro modelo de estudio, la proteína transportadora de cromato ChrA.



VII. ANEXOS

VII.1. ANEXO I – GRUPO DE DATOS DE SECUENCIAS DE PROTEÍNAS USADAS

En este trabajo se utilizaron un grupo de 135 secuencias, las cuales se listan a continuación.

Tabla 4 – Secuencias de proteínas usadas para reconstruir la filogenia de la proteína ChrA de *P. aeruginosa*

No. de acceso	Proteína	Organismo	No. de aa ¹
P14285	CHRA	<i>Pseudomonas aeruginosa</i>	416
P17551	CHRA	<i>Alcaligenes eutrophus</i>	401
AAO16588	chrA	<i>Acetobacter aceti</i>	406
NP_357519	AGR_L_3461	<i>Agrobacterium tumefaciens str. C58</i>	409
XP_306818	XM_306818.1	<i>Anopheles gambiae str. PEST</i>	384
EAA62587	AN5427.2	<i>Aspergillus nidulans FGSC A4</i>	523
EAA63309	AN3341.2	<i>Aspergillus nidulans FGSC A4</i>	510
ZP_00090393	Avin2082	<i>Azotobacter vinelandii</i>	481
CAB40623	chrA	<i>Bacillus cereus ATCC 10987</i>	393
NP_241440	chrA	<i>Bacillus halodurans C-125</i>	397
NP_241735	BH0869	<i>Bacillus halodurans C-125</i>	194
NP_241736	BH0870	<i>Bacillus halodurans C-125</i>	180
CAB07796	ywrA	<i>Bacillus subtilis</i>	178
CAB07797	ywrB	<i>Bacillus subtilis</i>	197
NP_810648	BT1735	<i>Bacteroides thetaiotaomicron VPI-5482</i>	185
NP_810649	BT1736	<i>Bacteroides thetaiotaomicron VPI-5482</i>	182
NP_811700	BT2788	<i>Bacteroides thetaiotaomicron VPI-5482</i>	174
NP_811701	BT2789	<i>Bacteroides thetaiotaomicron VPI-5482</i>	185
NP_968697	chrA	<i>Bdellovibrio bacteriovorus HD100</i>	378
NP_887842	BB1296	<i>Bordetella bronchiseptica RB50</i>	441
NP_887862	BB1316	<i>Bordetella bronchiseptica RB50</i>	190
NP_887863	BB1317	<i>Bordetella bronchiseptica RB50</i>	173
NP_212585	BB0451	<i>Borrelia burgdorferi B31</i>	177
NP_212586	BB0452	<i>Borrelia burgdorferi B31</i>	234
AAF22878	NA ²	<i>Bradyrhizobium japonicum</i>	461
NP_771692	blr5052	<i>Bradyrhizobium japonicum USDA 110</i>	461
NP_773156	bll6516	<i>Bradyrhizobium japonicum USDA 110</i>	176
NP_773157	bll6517	<i>Bradyrhizobium japonicum USDA 110</i>	171
ZP_00220917	Bucepa02004516	<i>Burkholderia cepacia R1808</i>	405



Tabla 4.- continuación.

ZP_00221415	Bucepa02003971	<i>Burkholderia cepacia</i> R1808	441
ZP_00221417	Bucepa02003973	<i>Burkholderia cepacia</i> R1808	382
ZP_00223004	Bucepa02002157	<i>Burkholderia cepacia</i> R1808	402
ZP_00223383	Bucepa02001813	<i>Burkholderia cepacia</i> R1808	178
ZP_00223384	Bucepa02001814	<i>Burkholderia cepacia</i> R1808	179
ZP_00224096	Bucepa02001048	<i>Burkholderia cepacia</i> R1808	175
ZP_00224097	Bucepa02001049	<i>Burkholderia cepacia</i> R1808	204
ZP_00224938	Bucepa02000074	<i>Burkholderia cepacia</i> R1808	392
ZP_00212348	Bcepa02006395	<i>Burkholderia cepacia</i> R18194	402
ZP_00212995	Bcepa02005793	<i>Burkholderia cepacia</i> R18194	179
ZP_00214008	Bcepa02004607	<i>Burkholderia cepacia</i> R18194	388
ZP_00216900	Bcepa02002078	<i>Burkholderia cepacia</i> R18194	179
ZP_00216901	Bcepa02002079	<i>Burkholderia cepacia</i> R18194	203
ZP_00027301	Bcep0052	<i>Burkholderia fungorum</i>	178
ZP_00027302	Bcep0053	<i>Burkholderia fungorum</i>	123
ZP_00027400	Bcep0155	<i>Burkholderia fungorum</i>	402
ZP_00031598	Bcep4438	<i>Burkholderia fungorum</i>	449
ZP_00031695	Bcep4535	<i>Burkholderia fungorum</i>	222
ZP_00031696	Bcep4536	<i>Burkholderia fungorum</i>	176
ZP_00032121	Bcep4968	<i>Burkholderia fungorum</i>	470
NP_419913	CC1097	<i>Caulobacter crescentus</i> CB15	426
ZP_00017652	Chlo0625	<i>Chloroflexus aurantiacus</i>	394
NP_900592	chrA	<i>Chromobacterium violaceum</i> ATCC 12472	393
NP_900787	CV1117	<i>Chromobacterium violaceum</i> ATCC 12472	185
NP_900788	CV1118	<i>Chromobacterium violaceum</i> ATCC 12472	171
NP_900941	CV1271	<i>Chromobacterium violaceum</i> ATCC 12472	437
NP_563164	CPE2248	<i>Clostridium perfringens</i> str. 13	178
NP_563165	CPE2249	<i>Clostridium perfringens</i> str. 13	173
NP_781330	CTC00660	<i>Clostridium tetani</i> E88	181
NP_781331	CTC00661	<i>Clostridium tetani</i> E88	179
NP_738956	CE2346	<i>Corynebacterium efficiens</i> YS-314	375
NP_601647	NCgl2363	<i>Corynebacterium glutamicum</i> ATCC 13032	376
ZP_00179154	Cwat020009	<i>Crocospaera watsonii</i> WH 8501 (<i>Synechocystis</i> sp. WH 8501)	389
ZP_00116649	Chut0006	<i>Cytophaga hutchinsonii</i>	385
ZP_00118906	Chut2301	<i>Cytophaga hutchinsonii</i>	400
ZP_00150272	Daro286201	<i>Dechloromonas aromatica</i> RCB	149
ZP_00150273	Daro286301	<i>Dechloromonas aromatica</i> RCB	172
NP_296134	DR2413	<i>Deinococcus radiodurans</i> R1	400



Tabla 4.- continuación.

ZP_00099495	Desu2638	<i>Desulfotobacterium hafniense</i>	157
AAS94909	DVU0426	<i>Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough</i>	450
YP_009133	DVUA0093	<i>Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough</i>	445
ZP_00182363	Exigu022435	<i>Exiguobacterium sp. 255-15</i>	385
NP_603609	FN0712	<i>Fusobacterium nucleatum subsp. nucleatum ATCC 25586</i>	186
NP_603610	FN0713	<i>Fusobacterium nucleatum subsp. nucleatum ATCC 25586</i>	176
ZP_00143575	FNV1807	<i>Fusobacterium nucleatum subsp. vincentii ATCC 49256</i>	176
ZP_00081244	Gmet1993	<i>Geobacter metallireducens</i>	192
ZP_00081245	Gmet1994	<i>Geobacter metallireducens</i>	215
XP_384019	FG03843.1	<i>Gibberella zeae PH-1 (anamorph: Fusarium graminearum)</i>	906
NP_923511	gll0565	<i>Gloeobacter violaceus PCC 7421</i>	423
ZP_00227130	Krad06001835	<i>Kineococcus radiotolerans SRS30216</i>	404
EAA57562	MG10634.4	<i>Magnaporthe grisea 70-15</i>	418
ZP_00051015	Magn3691	<i>Magnetospirillum magnetotacticum</i>	176
ZP_00052808	Magn027293	<i>Magnetospirillum magnetotacticum</i>	188
ZP_00208803	Magn026383	<i>Magnetospirillum magnetotacticum</i>	419
NP_106077	mlr5407	<i>Mesorhizobium loti MAFF303099</i>	463
ZP_00193416	MBNC02003383	<i>Mesorhizobium sp. BNC1</i>	421
AAB98712	MJ0718	<i>Methanococcus jannaschii</i>	402
ZP_00171795	Mflag022462	<i>Methylobacillus flagellatus KT</i>	404
ZP_00065378	Mdeg0749	<i>Microbulbifer degradans 2-40</i>	387
NP_326558	MYPU_7270	<i>Mycoplasma pulmonis UAB CTIP</i>	219
NP_326559	MYPU_7280	<i>Mycoplasma pulmonis UAB CTIP</i>	201
CAD21291	G15G9.040	<i>Neurospora crassa</i>	584
ZP_00105966	Npun0335	<i>Nostoc punctiforme</i>	406
NP_485153	all1110	<i>Nostoc sp. PCC 7120</i>	402
ZP_00093789	Saro0797	<i>Novosphingobium aromaticivorans</i>	495
NP_693525	OB2604	<i>Oceanobacillus iheyensis HTE831</i>	175
NP_693526	OB2605	<i>Oceanobacillus iheyensis HTE831</i>	187
NP_694187	OB3265	<i>Oceanobacillus iheyensis HTE831</i>	175
NP_694188	OB3266	<i>Oceanobacillus iheyensis HTE831</i>	193
NP_694198	OB3276	<i>Oceanobacillus iheyensis HTE831</i>	400
NP_892829	PMM0711	<i>Prochlorococcus marinus subsp. pastoris str. CCMP1986</i>	408
NP_252979	PA4289	<i>Pseudomonas aeruginosa PAO1</i>	401
ZP_00084601	Pflu1843	<i>Pseudomonas fluorescens PfO-1</i>	451
NP_744701	PP2556	<i>Pseudomonas putida KT2440</i>	450
ZP_00170113	Raut021274	<i>Ralstonia eutropha JMP134</i>	405
ZP_00021285	Reut0185	<i>Ralstonia metallidurans</i>	220



Tabla 4.- continuación.

ZP_00021286	Reut0186	<i>Ralstonia metallidurans</i>	176
ZP_00024026	Reut2992	<i>Ralstonia metallidurans</i>	407
ZP_00024348	Reut3320	<i>Ralstonia metallidurans</i>	411
NP_519100	RSc0979	<i>Ralstonia solanacearum</i> GMI1000	193
NP_519101	RSc0980	<i>Ralstonia solanacearum</i> GMI1000	181
NP_522116	chrA	<i>Ralstonia solanacearum</i> GMI1000	401
ZP_00004819	Rsph0707	<i>Rhodobacter sphaeroides</i>	486
ZP_00010183	Rpal1868	<i>Rhodopseudomonas palustris</i>	429
ZP_00012995	Rpal4722	<i>Rhodopseudomonas palustris</i>	198
NP_948836	RPA3498	<i>Rhodopseudomonas palustris</i> CGA009	198
ZP_00015256	Rrub2277	<i>Rhodospirillum rubrum</i>	416
ZP_00187419	Rxyl155901	<i>Rubrobacter xylanophilus</i> DSM 9941	150
NP_716615	SO0986	<i>Shewanella oneidensis</i> MR-1	390
NP_384859	chrA	<i>Sinorhizobium meliloti</i> 1021	480
ZP_00163524	Selo207101	<i>Synechococcus elongatus</i> PCC 7942	383
C56274	SrpC	<i>Synechococcus</i> sp.	393
NP_897416	SYNW1323	<i>Synechococcus</i> sp. WH 8102	394
BAA18657	chrA	<i>Synechocystis</i> sp. PCC 6803	399
NP_942194	slr5038	<i>Synechocystis</i> sp. PCC 6803	412
NP_229615	TM1818	<i>Thermotoga maritima</i> MSB8	166
NP_229616	TM1819	<i>Thermotoga maritima</i> MSB8	190
NP_972347	TDE1743	<i>Treponema denticola</i> ATCC 35405	187
NP_972348	TDE1744	<i>Treponema denticola</i> ATCC 35405	195
G71379	TP0146	<i>Treponema pallidum</i> subsp. <i>pallidum</i> (syphilis treponeme)	456
ZP_00073611	Tery2879	<i>Trichodesmium erythraeum</i> IMS101	416
NP_598135	chrA	<i>uncultured bacterium</i>	410
EAK83335	UM02213.1	<i>Ustilago maydis</i> 521	581
NP_231969	VC2339	<i>Vibrio cholerae</i> O1 biovar <i>el tor</i> str. N16961	380
NP_798786	VP2407	<i>Vibrio parahaemolyticus</i> RIMD 2210633	380
NP_935442	VV2649	<i>Vibrio vulnificus</i> YJ016	383

¹aa – aminoácidos; ²NA – No asignado.



VII.2. ANEXO II - NO TODO LO QUE BRILLA ES ORO

Otro hecho muy interesante es la presencia de una proteína similar a ChrA (25% de identidad con ChrA de *P. aeruginosa*) en el mosquito *Anopheles gambiae* (No. acceso XP_306818), ya que éste se encuentra filogenéticamente muy lejano a los demás organismos encontrados. De la misma manera que ocurre en hongos, se esperaba encontrar alguna proteína homóloga en la mosca modelo *Drosophila melanogaster*, pero no se identificó ninguna secuencia de este organismo ni dentro del grupo inicial de 813 secuencias.

Este comportamiento si bien no es común, no nos extraña demasiado ya que en otros análisis filogenéticos ya se han encontrado proteínas de *A. gambiae* en grupos donde no se espera encontrarla, y aunque es un fenómeno muy raro, no podemos descartar tampoco que haya ocurrido un evento transferencia horizontal de genes, como el que ocurrió con el dominio transmembranal HTTM (Schultz, 2004). Ahora bien, este tipo de distribución de secuencias se puede deber a fenómenos ya descritos: la transferencia o pérdida de genes (Salzberg y cols., 2001), o inclusive, problemas de contaminación en las muestras del DNA secuenciado (Holt y cols., 2002; Gómez y cols., 2005). Por ejemplo, varias de las secuencias que se publicaron en las base de datos (incluyendo las del GenkBank) han sido removidas una vez que se han descubierto estos problemas, incluyendo la secuencia usada en este trabajo. Esta observación recalca el problema del curado y actualización de las bases de datos, problema mencionado anteriormente.



VII.3. ANEXO III – CÓDIGO FUENTE DE LOS SCRIPTS

Para programar los script se utilizó el editor gVIM; se corrieron usando la sintaxis y el compilador de Perl v5.0. Los scripts se ejecutaron en el sistema operativo Linux Mandrake 9.1 con el fin de hacer más eficiente el proceso.

Script *extract1.pl*

```
#!/usr/bin/perl
#programa para parsear datos de archivos gb en datos
#para el análisis
#
$ARGV[1] || die "use extract.pl archivo.extrada nombre_proyecto\n";
open (IN,"$ARGV[0]") || die "no se puede abrir archivo $ARGV[0]";
open (OUT,">$ARGV[1].rep") || die "no se puede abrir archivo $ARGV[1].rep";
open (SAL,">$ARGV[1].bib") || die "no se puede abrir archivo $ARGV[1].bib";
open (EXT,">$ARGV[1].ext") || die "no se puede abrir archivo $ARGV[1].ext";
open (DB,">$ARGV[1].db") || die "no se puede abrir archivo $ARGV[1].db";
open (FAS,">$ARGV[1].txt") || die "no se puede abrir archivo $ARGV[1].txt";
$k=0;
$j=0;
$m=0;
while (<IN>)
{
    chop($_);
    #print "$_";
    if ($_ =~ /\bACCESSION\b/)
    {
        @linea=split(/\s/, $_);
        $acc=$linea[$#linea];
        @linea="";
    }
    if (($_ =~ /\b([a-z]+) aa/) and ($k==0))
    {
        $aa=$1;
        $aa =~ s/\s/_/;
    }
    if (($_ =~ /\b([a-z]+) aa/) and ($k==0))
    {
        $aa_num=$1;
    }
    if ($_ =~ /\bSOURCE\b/)
    {
        $_ =~ s/SOURCE //;
        $org=$_;
        @orgs=split(/\s/, $org);
        if ($orgs[0] =~ /(.*?) /)
        {
            $gen=$1;
        }
        if ($orgs[1] =~ /(.*?) /)
        {
            $esp=$1;
        }
        $org2=$gen.$esp;
        $org2 =~ tr/a-z/A-Z/;
    }
    $k=1 if ($_ =~ /\bORIGIN\b/);
    if ($k==1)
    {
        $_ =~ s/\s//g;
        if ($_ =~ /\b([a-z]+) /)
        {
            $fasta= $fasta.$1;
        }
        $fasta =~ tr/a-z/A-Z/;
    }
    $j=1 if (($_ =~ /\bCDS\b/) or ($_ =~ / gene /));
    if ($j==1)
    {
        if (($_ =~ /\b/gene="(.*?) /) or ($_ =~ /\b/locus_tag="(.*?) /) or ($_ =~ /\b/note="(.*?) /))
        {
            $gene=$1;
            $gene =~ s/similar to /s-/;
        }
    }
}
```



```

        chop($gene);
        $j++;
    }
}
$m=1 if ($_ =~ /\bAUTHORS\b/);
$m=0 if ($_ =~ /\bFEATURES\b/);

if ($m==1)
{
    $bib=$bib."\n".$_;
}
if ($_ =~ / ORGANISM/)
{
    $ww=1;
    $reino="";
}
$ww=0 if (($_ =~ /\bREFERENCE\b/) or ($_ =~ /\bCOMMENT\b/));
if ($ww == 2)
{
    $reino=$reino._;
    $reino =~ s/          //;
    $reino =~ s/ //;
}
$ww++ if ($ww == 1);
if ($_ =~ /\^\^\^\/)
{
    print OUT "$gene-$org2-$acc-$aa\n";
    print SAL "$gene\t$org\t$aa\n>$acc\n$fasta\n$bib\n\\\n\n";
    print EXT "GEN: $gene ($aa)\nORGANISMO: $org\n$reino\n>$acc\n$fasta\n\\\n\n";
    print DB "ID\t$acc\t$gene\t$org\t$aa_num\t$fasta\n\n";
    print FAS ">$org2-$acc-$aa\n$fasta\\\n\n";
    $acc="";
    $org="";
    $org2="";
    $fasta="";
    $gene="";
    $k=0;
    $bib="";
    $j=0;
}
$sk=1 if ($_ =~ /\bORIGIN\b/)
}
close OUT;
close SAL;
close EXT;
close DB;
close FAS;

open (IN, "$ARGV[1].rep") || die "no se puede abrir archivo $ARGV[0]";
open (OUT, ">$ARGV[1].temp") || die "no se puede abrir archivo $ARGV[1].rp1";
while (<IN>)
{
    chop($_);
    @linea=split(/-/,$_);
    shift (@linea);
    $temp="";
    $temp=$linea[0]."-".$linea[1]."-".$linea[2];
    push(@ord, $temp);
}
@ord2=sort(@ord);
foreach $a (@ord2)
{
    print OUT "$a\n";
}

system("cat $ARGV[1].temp > $ARGV[1].rep");
system("rm -f $ARGV[1].temp");

close IN;
close OUT;

```



Script *parse_blast1.pl*

```
#!/usr/bin/perl
#programa para parsear blast y analizar los datos

$ARGV[1] || die "use parse_blast.pl archivo.raiz nombre_proyecto\n";
open (IN, "$ARGV[0]") || "no se puede abrir archivo $ARGV[0]";
open (OUT, ">$ARGV[1]") || "no se puede abrir archivo $ARGV[1]";
$mk=0;
while (<IN>)
{
    chop($_); $ck=0;
    if ($_ =~ /S2/)
    {
        $jk=0;
        foreach $line (@ficha)
        {
            if ($line =~ /==> (.*?) <==/)
            {
                print OUT "$line\n";
            }
            if ($line =~ /^>(.*?) /)
            {
                $clv=$clave; $clave=$1;
                if ($pk==0)
                {
                    $var = $var."\t$e\t$iden\t$pos\t$inicio\t$final\t";
                }
                push(@blast,$var) if ($ck > 0);
                $ck=0; $dk=0; $jk++;
                $var=$clave."\t" if ($ck == 0);
            }
            if ($line =~ /Score/)
            {
                if ($ck>0)
                {
                    $var = $var."\t$e\t$iden\t$pos\t$inicio\t$final\t";
                }
                $ck++; $pk=0; $dk=0;
            }
            if ($line =~ / Expect = (.*?) /)
            {
                $e=$1;
            }
            if ($line =~ /Identities = (.*?) \ (.*?) , Positives = (\d+\/\d+)\s+ /)
            {
                $iden= $1; $pos= $2;
            }
            if ($line =~ /Sbjct: (\d+) .* (\d+) /)
            {
                $inicio = $1 if ($dk == 0);
                $final= $2;
                $dk++;
            }
            if ($line =~ / Database: /)
            {
                $var = $var."\t$e\t$iden\t$pos\t$inicio\t$final\t";
                push(@blast,$var);
                $sss=shift(@blast) if ($mk ne 0);
                $sss=shift(@blast);
                foreach $r (@blast)
                {
                    print OUT "$r\n";
                }
                $mk++;
                @blast="";
            }
        }
        @ficha="";
    }
    push(@ficha,$_);
}
close IN;
close OUT;
```



Script *reporte_blast.pl*

```
#!/usr/bin/perl
#
#obtener un reporte de resultado de blast a partir de un archivo solo con scores

$ARGV[2] || die "use ./reporte_blast entrada grupos salida\n";
open (IN,"$ARGV[0]") || "no se puede abrir archivo $ARGV[0]";
open (REF,"$ARGV[1]") || "no se puede abrir archivo $ARGV[1]";
open (OUT,">$ARGV[2]") || "no se puede abrir archivo $ARGV[2]";
open (SAL,">$ARGV[2].rep") || "no se puede abrir archivo $ARGV[2]";
@gr="";
$i=0;
while (<REF>)
{
    chop($_);
    if ($_ =~ /grupo/)
    {
        $grupo = "@gr" if ($i>=1);
        $i++;
        @gr="";
        next;
    }
    push(@gr,"$ ");
}
open (DOS,">$ARGV[2].rp2") || "no se puede abrir archivo $ARGV[2]";
while(<IN>)
{
    chop($_);
    if (($_ =~ /\s+(\w.+\\d+)-\\d+_aa/) or ($_ =~ /==> (\w.+\\d+)_\\d+_aa/))
    {
        $acc=$1;
        for ($j=1; $j<=$i-1; $j++)
        {
            @gr=split(/\s/, $grupo{$j});
            foreach $var (@gr)
            {
                if ($var =~ /$acc/)
                {
                    $_=$_."t--".$j;
                }
            }
        }
        print DOS "$_\\n";
    }
}
close DOS;
close IN;
open (DOS,"$ARGV[2].rp2") || "no se puede abrir archivo $ARGV[0]";
$l=0;
$check=0;
while (<DOS>)
{
    chop($_);
    #reconocimiento del numero de acceso de blast y el grupo al que pertenece
    #en $acc y $per y @gr
    if ($_ =~ /==> (\w.+\\d+)_\\d+_aa/)
    {
        $acc1=$acc if ($l != 0);
        $acc=$1;
        if ($l != 0)
        {
            print SAL "$acc1--$per--$k\\n";
            print SAL "$rept\\n";
            $k=0;
            $rept="";
        }
        $l++;
        for ($j=1; $j<=$i-1; $j++)
        {
            @gr=split(/\s/, $grupo{$j});
            foreach $var (@gr)
            {
                if ($var =~ /$acc/)
                {
                    $per=$j;
                }
            }
        }
    }
    @gr=split(/\s/, $grupo{$per});
}
```



```

if ($_ =~ /\w.+\.d+\.aa/)
{
    $bl=$1;
    foreach $var (@gr)
    {
        if ($var =~ $bl)
        {
            $k++;
            $check=1;
            $rept=$rept.$_."\\n";
        }
        else
        {
            #print OUT "$_\\n";
        }
    }
    print OUT "$_\\n" if ($check == 0);
    print OUT "$_ - * \\n" if ($check == 1);
    $check=0;
}
print SAL "$acc--$per--$k\\n";
print SAL "$rept\\n";
close IN;
close OUT;

```



Script *ordena_blast.pl*

```
#!/usr/bin/perl

$ARGV[1] || die "use ordena_blast_grp.pl entrada salida\n";
open (IN,"$ARGV[0]") || "no se puede abrir archivo $ARGV[0]";
open (OUT,">$ARGV[1]") || "no se puede abrir archivo $ARGV[1]";
$grp=0;
while(<IN>)
{
    #chop($_);
    if ($_ =~ /==> (\w.+d+)_\d+_aa.+-(\d+)/)
    # if ($_ =~ /==> (\w.+d+)_\d+_aa.+(\d+)/)
    {
        $pert=$nvo;
        $nvo=$2;
        $gr{$pert} = "$gr{$pert}".$ficha";
        @ficha = "";
        $grp = $pert if ($grp < $pert);
    }

    push(@ficha,"$_");
}
$pert=$nvo;
$gr{$pert} = "$gr{$pert}".$ficha";
@ficha = "";
print "$grp\n";
for ($j=1; $j<=$grp; $j++)
{
    @grupo=split(/\n/, $gr{$j});
    foreach $var (@grupo)
    {
        $var =~ s/ //;
        print OUT "$var\n";
    }
}
close IN;
close OUT;
```



Script *linea_ord.pl*

```
#!/usr/bin/perl
#
$ARGV[1] || die "use linea_ord.pl entrada salida\n";
open (IN,"$ARGV[0]") || die "no se puede abrir archivo $ARGV[0]";
open (OUT,">$ARGV[1]") || die "no se puede abrir archivo $ARGV[1]";
while(<IN>)
{
    chop($_);
    if ($_ =~ />=> ((\w.+d+)_d+_aa.+(d+))/)
    {
        print OUT "$1\n";
        #print "$1\n";
    }
    else
    {
        @lin="";
        @linea=split(/\t/, $_);
        $linea[1] = $linea[$#linea];
        $linea[$#linea] = "";
        foreach $x (@linea)
        {
            if ($x eq "")
            {
            }
            else
            {
                push(@lin,$x)
            }
        }
        shift(@lin);
        for ($i=0;$i<=$#lin;$i++)
        {
            print OUT "$lin[$i]\t" if ($i != $#lin);
            print OUT "$lin[$i]\n" if ($i == $#lin);
        }
    }
}
close IN;
close OUT;
```



VIII. BIBLIOGRAFÍA

- Abramson, J., I. Smirnova, V. Kasho, G. Verner, H. R. Kaback y S. Iwata (2003). "Structure and mechanism of the lactose permease of *Escherichia coli*." *Science* 301: 610-5.
- Ackerley, D. F., C. F. González, C. H. Park, R. Blake, 2nd, M. Keyhan y A. Matin (2004). "Chromate-reducing properties of soluble flavoproteins from *Pseudomonas putida* and *Escherichia coli*." *Appl Environ Microbiol* 70: 873-82.
- Aguilar, M. E. (2005). *Análisis de la función de residuos esenciales de la proteína ChrA de Pseudomonas aeruginosa*. Tesis de Maestría. IIQB. UMSNH. Morelia, Mich. MÉXICO. 122p.
- Aguilera, S., M. E. Aguilar, M. P. Chávez, J. E. López-Meza, M. Pedraza-Reyes, J. Campos-García y C. Cervantes (2004). "Essential residues in the chromate transporter ChrA of *Pseudomonas aeruginosa*." *FEMS Microbiol Lett* 232: 107-12.
- Alcedo, J. A. y K. E. Wetterhahn (1990). "Chromium toxicity and carcinogenesis." *Int Rev Exp Pathol* 31: 85-108.
- Aller, S. G., E. T. Eng, C. J. De Feo y V. M. Unger (2004). "Eukaryotic CTR copper uptake transporters require two faces of the third transmembrane domain for helix packing, oligomerization, and function." *J Biol Chem* 279: 53435-41.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller y D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25: 3389-402.
- Álvarez, A. H., R. Moreno-Sánchez y C. Cervantes (1999). "Chromate efflux by means of the ChrA chromate resistance protein from *Pseudomonas aeruginosa*." *J Bacteriol* 181: 7398-400.
- Arias, Y. M. y B. M. Tebo (2003). "Cr(VI) reduction by sulfidogenic and nonsulfidogenic microbial consortia." *Applied and Environmental Microbiology* 69: 1847-53.
- ATSDR (2000). *ToxFAQs™ for Chromium*. Update. Atlanta, GA; USA., Department of Health and Human Services, Public Health Service.
- Attwood, T. K. (2000). "Genomics. The Babel of bioinformatics." *Science* 290: 471-3.
- Baldauf, S. L. (2003). "Phylogeny for the faint of heart: a tutorial." *Trends Genet* 19: 345-51.
- Ballatori, N. (2002). "Transport of toxic metals by molecular mimicry." *Environ Health Perspect* 110 Suppl 5: 689-94.
- Bapteste, E. y C. Brochier (2004). "On the conceptual difficulties in rooting the tree of life." *Trends Microbiol* 12: 9-13.
- Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats y S. R. Eddy (2004). "The Pfam protein families database." *Nucleic Acids Res* 32 Database issue: D138-41.
- Bezanilla, F. (2000). "The voltage sensor in voltage-dependent ion channels." *Physiol Rev* 80: 555-92.
- Birnboim, H. C. y J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." *Nucleic Acids Res* 7: 1513-23.
- Braun, P. y G. von Heijne (1999). "The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane." *Biochemistry* 38: 9778-82.
- Brown, J. R. (2001). *Universal Tree of Life*. ENCYCLOPEDIA OF LIFE SCIENCES, Nature Publishing Group.
- Campos-García, J., G. Martínez-Cadena, R. Álvarez-González y C. Cervantes (1997). "Purification and partial characterization of a chromate reductase from *Bacillus*." *Rev Latinoam Microbiol* 39: 73-81.
- Cervantes, C., J. Campos-García, S. Devars, F. Gutierrez-Corona, H. Loza-Tavera, J. C. Torres-Guzman y R. Moreno-Sánchez (2001). "Interactions of chromium with microorganisms and plants." *FEMS Microbiol Rev* 25: 335-47.
- Cervantes, C., H. Ohtake, L. Chu, T. K. Misra y S. Silver (1990). "Cloning, nucleotide sequence, and expression of the chromate resistance determinant of *Pseudomonas aeruginosa* plasmid pUM505." *J Bacteriol* 172: 287-91.
- Cervantes, C. y S. Silver (1992). "Plasmid chromate resistance and chromate reduction." *Plasmid* 27: 65-71.
- Chang, B. S. y M. J. Donoghue (2000). "Recreating ancestral proteins." *Trends Ecol Evol* 15: 109-114.
- Chen, N., R. A. Vaughan y M. E. Reith (2001). "The role of conserved tryptophan and acidic residues in the human dopamine transporter as characterized by site-directed mutagenesis." *J Neurochem* 77: 1116-27.
- Cortés, R. (2005). *Análisis de las prolinas del segmento transmembranal II de la proteína ChrA*. Tesis de Licenciatura. Escuela de Químico-Farmacología. UMSNH. Morelia, Mich. MÉXICO. 61p.
- Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay y G. J. Barton (1998). "JPred: a consensus secondary structure prediction server." *Bioinformatics* 14: 892-3.
- Curran, A. R. y D. M. Engelman (2003). "Sequence motifs, polar interactions and conformational changes in helical membrane proteins." *Curr Opin Struct Biol* 13: 412-7.
- Dhakephalkar, D. K., J. V. Bhide y K. M. Paknikar (1996). "Plasmid mediated chromate resistance and reduction in *Pseudomonas mendocina* MCM B-180." *Biotechnology Letters* 18: 1119-1122.
- Ditty, J. L. y C. S. Harwood (1999). "Conserved cytoplasmic loops are important for both the transport and chemotaxis functions of PcaK, a protein from *Pseudomonas putida* with 12 membrane-spanning regions." *J Bacteriol* 181: 5068-74.
- Dobzhansky, T. (1973). "Nothing in Biology Makes Sense Except in the Light of Evolution." *American Biology Teacher* 35: 125-129.
- Dopazo, J. y J. M. Carazo (1997). "Phylogenetic reconstruction using an unsupervised growing neural network that adopts the topology of a phylogenetic tree." *J Mol Evol* 44: 226-33.
- Draheim, R. R., A. F. Bormans, R. Z. Lai y M. D. Manson (2005). "Tryptophan residues flanking the second transmembrane helix (TM2) set the signaling state of the Tar chemoreceptor." *Biochemistry* 44: 1268-77.
- Dreyfuss, J. (1964). "Characterization of a Sulfate- and Thiosulfate-Transporting System in *Salmonella Typhimurium*." *J Biol Chem* 239: 2292-7.
- Efstathiou, J. D. y L. L. McKay (1977). "Inorganic salts resistance associated with a lactose-fermenting plasmid in *Streptococcus lactis*." *J Bacteriol* 130: 257-65.
- Enderle, P. J. y M. A. Farwell (1998). "Electroporation of freshly plated *Escherichia coli* and *Pseudomonas aeruginosa* cells." *Biotechniques* 25: 954-6, 958.



- Fisher, C. L. y G. K. Pei (1997). "Modification of a PCR-based site-directed mutagenesis method." *Biotechniques* 23: 570-1, 574.
- Fleming, K. G. y D. M. Engelman (2001). "Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants." *Proc Natl Acad Sci U S A* 98: 14340-4.
- Fournier, M., Z. Dermoun, M. C. Durand y A. Dolla (2004). "A new function of the *Desulfovibrio vulgaris* Hildenborough [Fe] hydrogenase in the protection against oxidative stress." *J Biol Chem* 279: 1787-93.
- Gaberell, M., Y. P. Chin, S. J. Hug y B. Sulzberger (2003). "Role of dissolved organic matter composition on the photoreduction of Cr(VI) to Cr(III) in the presence of iron." *Environ Sci Technol* 37: 4403-9.
- Gómez, S. M., K. Eiglmeier, B. Segurens, P. Dehoux, A. Couloux, C. Scarpelli, P. Wincker, J. Weissenbach, P. T. Brey y C. W. Roth (2005). "Pilot *Anopheles gambiae* full-length cDNA study: sequencing and initial characterization of 35,575 clones." *Genome Biol* 6: R39.
- González, M. V. (2004). *Participación de Helicasas en la Toxicidad de cromato en Pseudomonas aeruginosa*. Tesis de licenciatura. Escuela de Químico Farmacobiología. UMSNH. Morelia, Mich. MÉXICO. 81p.
- Green, A. L., E. J. Anderson y R. J. Brooker (2000). "A revised model for the structure and function of the lactose permease. Evidence that a face on transmembrane segment 2 is important for conformational changes." *J Biol Chem* 275: 23240-6.
- Guo, X. W., P. R. Smith, B. Cognon, D. D'Arcangelis, E. Dolginova y C. A. Mannella (1995). "Molecular design of the voltage-dependent, anion-selective channel in the mitochondrial outer membrane." *J Struct Biol* 114: 41-59.
- Hall, B. G. (2001). *Phylogenetic Trees Made Easy, Second Edition*. Sunderland, Sinauer Associates, Inc.
- Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucleic Acids Symp Ser* 41: 95-98.
- Harrison, R. M. (2001). *Variation, Within Species: Introduction*. *ENCYCLOPEDIA OF LIFE SCIENCES*, Nature Publishing Group.
- Hirai, T., J. A. Heymann, P. C. Maloney y S. Subramaniam (2003). "Structural model for 12-helix transporters belonging to the major facilitator superfamily." *J Bacteriol* 185: 1712-8.
- Hirai, T., J. A. Heymann, D. Shi, R. Sarker, P. C. Maloney y S. Subramaniam (2002). "Three-dimensional structure of a bacterial oxalate transporter." *Nat Struct Biol* 9: 597-600.
- Holloway, B. W. (1969). "Genetics of *Pseudomonas*." *Bacteriol Rev* 33: 419-43.
- Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab, D. R. Nusskern, P. Wincker, A. G. Clark, J. M. Ribeiro, R. Wides, S. L. Salzberg, B. Loftus, M. Yandell, W. H. Majoros, D. B. Rusch, Z. Lai, C. L. Kraft, J. F. Abril, V. Anthouard, P. Arensburger, P. W. Atkinson, H. Baden, V. de Berardinis, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chaturverdi, G. K. Christophides, M. A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C. A. Evans, M. Flanagan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M. E. Hillenmeyer, S. L. Hladun, J. R. Hogan, Y. S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J. J. Lin, N. F. Lobo, J. R. López, J. A. Malek, T. C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S. D. Murphy, D. A. O'Brochta, C. Pfannkoch, R. Qi, M. A. Regier, K. Remington, H. Shao, M. V. Sharakhova, C. D. Sitter, J. Shetty, T. J. Smith, R. Strong, J. Sun, D. Thomasova, L. Q. Ton, P. Topalis, Z. Tu, M. F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K. J. Woodford, J. R. Wortman, M. Wu, A. Yao, E. M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S. C. Zhu, I. Zhimulev, M. Coluzzi, A. della Torre, C. W. Roth, C. Louis, F. Kalush, R. J. Mural, E. W. Myers, M. D. Adams, H. O. Smith, S. Broder, M. J. Gardner, C. M. Fraser, E. Birney, P. Bork, P. T. Brey, J. C. Venter, J. Weissenbach, F. C. Kafatos, F. H. Collins y S. L. Hoffman (2002). "The genome sequence of the malaria mosquito *Anopheles gambiae*." *Science* 298: 129-49.
- Hryniewicz, M., A. Sirko, A. Palucha, A. Bock y D. Hulanicka (1990). "Sulfate and thiosulfate transport in *Escherichia coli* K-12: identification of a gene encoding a novel protein involved in thiosulfate binding." *J Bacteriol* 172: 3358-66.
- Huang, Y., M. J. Lemieux, J. Song, M. Auer y D. N. Wang (2003). "Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*." *Science* 301: 616-20.
- Ish-Horowitz, D. y J. F. Burke (1981). "Rapid and efficient cosmid cloning." *Nucleic Acids Res* 9: 2989-98.
- Jamison, D. C. (2003). "Open bioinformatics." *Bioinformatics* 19: 679-80.
- Javadpour, M. M., M. Eilers, M. Groesbeek y S. O. Smith (1999). "Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association." *Biophysical Journal* 77: 1609-18.
- Jessen-Marshall, A. E. y R. J. Brooker (1996). "Evidence that transmembrane segment 2 of the lactose permease is part of a conformationally sensitive interface between the two halves of the protein." *J Biol Chem* 271: 1400-4.
- Jessen-Marshall, A. E., N. J. Paul y R. J. Brooker (1995). "The conserved motif, GXXX(D/E)(R/K)XG[X](R/K)(R/K), in hydrophilic loop 2/3 of the lactose permease." *J Biol Chem* 270: 16251-7.
- Jiménez, R. (2005). *Determinación de la topología membranar de la proteína ChrA de Pseudomonas aeruginosa*. I Tesis de Maestría. IIQB. UMSNH. Morelia, Mich. MÉXICO. 102p.
- Juhnke, S., N. Peitzsch, N. Hubener, C. Grosse y D. H. Nies (2002). "New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34." *Arch Microbiol* 179: 15-25.
- Kada, T., K. Tutikawa y Y. Sadaie (1972). "In vitro and host-mediated "rec-assay" procedures for screening chemical mutagens; and phloxine, a mutagenic red dye detected." *Mutat Res* 16: 165-74.
- Koonin, E. V., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, D. M. Krylov, K. S. Makarova, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, I. B. Rogozin, S. Smirnov, A. V. Sorokin, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin y D. A. Natale (2004). "A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes." *Genome Biol* 5: R7.
- Kumar, S., K. Tamura, I. B. Jakobsen y M. Nei (2001). "MEGA2: molecular evolutionary genetics analysis software." *Bioinformatics* 17: 1244-5.
- Laage, R. y D. Langosch (2001). "Strategies for prokaryotic expression of eukaryotic membrane proteins." *Traffic* 2: 99-104.
- Landolt-Marticorena, C., K. A. Williams, C. M. Deber y R. A. Reithmeier (1993). "Non-random distribution of amino acids in the transmembrane segments of



- human type I single span membrane proteins." *J Mol Biol* 229: 602-8.
- LaVelle, J. M. (1986). "Potassium chromate potentiates frameshift mutagenesis in *E. coli* and *S. typhimurium*." *Mutat Res* 171: 1-10.
- Lemmon, M. A. y D. M. Engelman (1994). "Specificity and promiscuity in membrane helix interactions." *FEBS Lett* 346: 17-20.
- Lerat, E., V. Daubin y N. A. Moran (2003). "From Gene Trees to Organismal Phylogeny in Prokaryotes: The Case of the γ -Proteobacteria." *PLoS Biology* 1: 101-109.
- Lo Conte, L., B. Ailey, T. J. Hubbard, S. E. Brenner, A. G. Murzin y C. Chothia (2000). "SCOP: a structural classification of proteins database." *Nucleic Acids Res* 28: 257-9.
- Mansilla, M. C. y D. de Mendoza (2000). "The *Bacillus subtilis* *cysP* gene encodes a novel sulphate permease related to the inorganic phosphate transporter (Pit) family." *Microbiology* 146 (Pt 4): 815-21.
- McGrath, S. P. y S. Smith (1990). *Chromium and nickel. Heavy Metals in Soils*. A. B. J. Glasgow, Blackie. 7: 125-150.
- McLean, J. y T. J. Beveridge (2001). "Chromate reduction by a pseudomonad isolated from a site contaminated with chromated copper arsenate." *Applied and Environmental Microbiology* 67: 1076-84.
- Meinel, T., A. Krause, H. Luz, M. Vingron y E. Staub (2005). "The SYSTERS Protein Family Database in 2005." *Nucleic Acids Res* 33 Database Issue: D226-9.
- Miller, A. S. y J. J. Falke (2004). "Side chains at the membrane-water interface modulate the signaling state of a transmembrane receptor." *Biochemistry* 43: 1763-70.
- Miranda, A. T. (2003). *Genes de resistencia a cromato en el cromosoma de Pseudomonas aeruginosa*. Tesis de Maestría. IIQB. UMSNH. Morelia, Mich. MÉXICO. 120p.
- Moreno, M. G. (2005). *Importancia de residuos básicos de la proteína ChrA de Pseudomonas aeruginosa*. Tesis de Maestría. IIQB. UMSNH. Morelia, Mich. MÉXICO. En proceso.
- Nei, M. y S. Kumar (2000). *Molecular Evolution and Phylogenetics*. New York., Oxford University Press.
- Nicholson, H., D. E. Tronrud, W. J. Becktel y B. W. Matthews (1992). "Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme." *Biopolymers* 32: 1431-41.
- Nies, A., D. H. Nies y S. Silver (1990). "Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from *Alcaligenes eutrophus*." *J Biol Chem* 265: 5648-53.
- Nies, D. H. (2003). "Efflux-mediated heavy metal resistance in prokaryotes." *FEMS Microbiol Rev* 27: 313-39.
- Nies, D. H., S. Koch, S. Wachi, N. Peitzsch y M. H. Saier, Jr. (1998). "CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate antiporters." *J Bacteriol* 180: 5799-802.
- Ohtake, H., C. Cervantes y S. Silver (1987). "Decreased chromate uptake in *Pseudomonas fluorescens* carrying a chromate resistance plasmid." *J Bacteriol* 169: 3853-6.
- Ohtake, H., K. Komori, C. Cervantes y K. Toda (1990). "Chromate-resistance in a chromate-reducing strain of *Enterobacter cloacae*." *FEMS Microbiol Lett* 55: 85-8.
- Okada, A., T. Miura y H. Takeuchi (2001). "Protonation of histidine and histidine-tryptophan interaction in the activation of the M2 ion channel from influenza A virus." *Biochemistry* 40: 6053-60.
- Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells y J. M. Thornton (1997). "CATH--a hierarchic classification of protein domain structures." *Structure* 5: 1093-108.
- Page, R. D. (1996). "TreeView: an application to display phylogenetic trees on personal computers." *Comput Appl Biosci* 12: 357-8.
- Pao, S. S., I. T. Paulsen y M. H. Saier, Jr. (1998). "Major facilitator superfamily." *Microbiol Mol Biol Rev* 62: 1-34.
- Pardee, A. B., L. S. Prestidge, M. B. Whipple y J. Dreyfuss (1966). "A binding site for sulfate and its relation to sulfate transport into *Salmonella typhimurium*." *J Biol Chem* 241: 3962-9.
- Penny, D. (2004). "Evolutionary trees for all." *Trends Ecol Evol* 19: Editorial.
- Petrilli, F. L. y S. De Flora (1977). "Toxicity and mutagenicity of hexavalent chromium on *Salmonella typhimurium*." *Applied and Environmental Microbiology* 33: 805-9.
- Philippe, H. y C. J. Douady (2003). "Horizontal gene transfer and phylogenetics." *Curr Opin Microbiol* 6: 498-505.
- Pimentel, B. E., R. Moreno-Sánchez y C. Cervantes (2002). "Efflux of chromate by *Pseudomonas aeruginosa* cells expressing the ChrA protein." *FEMS Microbiol Lett* 212: 249-54.
- Quievryn, G., E. Peterson, J. Messer y A. Zhitkovich (2003). "Genotoxicity and mutagenicity of chromium(VI)/ascorbate-generated DNA adducts in human and bacterial cells." *Biochemistry* 42: 1062-70.
- Reece, S., C. P. Simmons, R. J. Fitzhenry, M. Batchelor, C. Hale, S. Matthews, A. D. Phillips, G. Dougan y G. Frankel (2002). "Mutagenesis of conserved tryptophan residues within the receptor-binding domain of intimin: influence on binding activity and virulence." *Microbiology* 148: 657-65.
- Rehm, B. H. (2001). "Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification." *Appl Microbiol Biotechnol* 57: 579-92.
- Rhee, M. H., I. Nevo, M. L. Bayewitch, O. Zagoory y Z. Vogel (2000). "Functional role of tryptophan residues in the fourth transmembrane domain of the CB(2) cannabinoid receptor." *J Neurochem* 75: 2485-91.
- Ridder, A. N., S. Morein, J. G. Stam, A. Kuhn, B. de Kruijff y J. A. Killian (2000). "Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins." *Biochemistry* 39: 6521-8.
- Riveros-Rosas, H. (2004). *Diversidad, taxonomía y evolución de las deshidrogenasas/reductasas de cadena media*. Tesis doctoral. División de Estudios de Posgrado e Investigación, Facultad de Medicina. UNAM. México D.F. MÉXICO. 129p.
- Riveros-Rosas, H., A. Julian-Sánchez, R. Villalobos-Molina, J. P. Pardo y E. Pina (2003). "Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily." *Eur J Biochem* 270: 3309-34.
- Rubin, R. A., S. B. Levy, R. L. Heinrikson y F. J. Kezdy (1990). "Gene duplication in the evolution of the two complementing domains of gram-negative bacterial tetracycline efflux proteins." *Gene* 87: 7-13.
- Russ, W. P. y D. M. Engelman (2000). "The GxxxG motif: a framework for transmembrane helix-helix association." *J Mol Biol* 296: 911-9.
- Saier, M. H., Jr. (1994). "Computer-aided analyses of transport protein sequences: gleaned evidence



- concerning function, structure, biogenesis, and evolution." *Microbiol Rev* 58: 71-93.
- Saier, M. H., Jr. (2000). "A functional-phylogenetic classification system for transmembrane solute transporters." *Microbiol Mol Biol Rev* 64: 354-411.
- Saier, M. H., Jr. (2003). "Tracing pathways of transport protein evolution." *Mol Microbiol* 48: 1145-56.
- Salzberg, S. L., O. White, J. Peterson y J. A. Eisen (2001). "Microbial genes in the human genome: lateral transfer or gene loss?" *Science* 292: 1903-6.
- Samanta, U., D. Pal y P. Chakrabarti (2000). "Environment of tryptophan side chains in proteins." *Proteins* 38: 288-300.
- Schiffer, M., C. H. Chang y F. J. Stevens (1992). "The functions of tryptophan residues in membrane proteins." *Protein Eng* 5: 213-4.
- Schultz, J. (2004). "HTTM, a horizontally transferred transmembrane domain." *Trends Biochem Sci* 29: 4-7.
- Senes, A., D. E. Engel y W. F. DeGrado (2004). "Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs." *Current Opinion in Structural Biology* 14: 465-79.
- Silver, S. y M. Walderhaug (1992). "Gene regulation of plasmid- and chromosome-determined inorganic ion transport in bacteria." *Microbiol Rev* 56: 195-228.
- Sirko, A., M. Hryniewicz, D. Hulanicka y A. Bock (1990). "Sulfate and thiosulfate transport in *Escherichia coli* K-12: nucleotide sequence and expression of the *cysTWAM* gene cluster." *J Bacteriol* 172: 3351-7.
- Stein, L. (2004). *Using Perl to facilitate biological analysis. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. B. F. F. O. Andreas D. Baxeavanis: 413.
- Stein, L. D. (1996). "How Perl Saved the Human Genome Project." *Perl Journal* 1: 413-49.
- Stratagene QuikChange® Site-Directed Mutagenesis Kit 1: 20.
- Su, M. A., C. C. Trenor, J. C. Fleming, M. D. Fleming y N. C. Andrews (1998). "The G185R mutation disrupts function of the iron transporter *Nramp2*." *Blood* 92: 2157-63.
- Summers, A. O. y G. A. Jacoby (1978). "Plasmid-determined resistance to boron and chromium compounds in *Pseudomonas aeruginosa*." *Antimicrob Agents Chemother* 13: 637-40.
- Suzuki, T., N. Miyata, H. Horitsu, K. Kawai, K. Takamizawa, Y. Tai y M. Okazaki (1992). "NAD(P)H-dependent chromium (VI) reductase of *Pseudomonas ambigua* G-1: a Cr(V) intermediate is formed during the reduction of Cr(VI) to Cr(III)." *J Bacteriol* 174: 5340-5.
- Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin y D. A. Natale (2003). "The COG database: an updated version includes eukaryotes." *BMC Bioinformatics* 4: 41.
- Telford, M. J. y G. E. Budd (2003). "The place of phylogeny and cladistics in Evo-Devo research." *Int J Dev Biol* 47: 479-90.
- Thompson, J. D., D. G. Higgins y T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res* 22: 4673-80.
- Vankova, S., J. Kupec y J. Hoffmann (1999). "Toxicity of chromium to activated sludge." *Ecotoxicol Environ Saf* 42: 16-21.
- Vázquez-Ibar, J. L., L. Guan, M. Svrakic y H. R. Kaback (2003). "Exploiting luminescence spectroscopy to elucidate the interaction between sugar and a tryptophan residue in the lactose permease of *Escherichia coli*." *Proc Natl Acad Sci U S A* 100: 12706-11.
- Viamajala, S., B. M. Peyton, R. K. Sani, W. A. Apel y J. N. Petersen (2004). "Toxic effects of chromium(VI) on anaerobic and aerobic growth of *Shewanella oneidensis* MR-1." *Biotechnol Prog* 20: 87-95.
- Wakabayashi, S., T. Hisamitsu, T. Pang y M. Shigekawa (2003). "Mutations of Arg440 and Gly455/Gly456 oppositely change pH sensing of Na⁺/H⁺ exchanger 1." *J Biol Chem* 278: 11828-35.
- Wang, P. C., T. Mori, K. Toda y H. Ohtake (1990). "Membrane-associated chromate reductase activity from *Enterobacter cloacae*." *J Bacteriol* 172: 1670-2.
- West, S. E., H. P. Schweizer, C. Dall, A. K. Sample y L. J. Runyen-Janecky (1994). "Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*." *Gene* 148: 81-6.
- Wickware, P. (2004). "The Language of Bioinformatics." *The Scientist* 14: 40.
- Woese, C. (1998). "The universal ancestor." *Proc Natl Acad Sci U S A* 95: 6854-9.
- Woese, C. R. (2000). "Interpreting the universal phylogenetic tree." *Proc Natl Acad Sci U S A* 97: 8392-6.
- Woese, C. R. y G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." *Proc Natl Acad Sci U S A* 74: 5088-90.
- Woese, C. R., O. Kandler y M. L. Wheelis (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." *Proc Natl Acad Sci U S A* 87: 4576-9.
- Wolf, Y. I., I. B. Rogozin, N. V. Grishin y E. V. Koonin (2002). "Genome trees and the tree of life." *Trends Genet* 18: 472-9.
- Woolf, T. B. (1997). "Molecular dynamics of individual alpha-helices of bacteriorhodopsin in dimyristol phosphatidylcholine. I. Structure and dynamics." *Biophys J* 73: 2376-92.
- Yanisch-Perron, C., J. Vieira y J. Messing (1985). "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors." *Gene* 33: 103-19.
- Yokoyama, S. (2002). "Evaluating adaptive evolution." *Nat Genet* 30: 350-1.
- Zuckerkandl, E. y L. Pauling (1965). "Molecules as documents of evolutionary history." *J Theor Biol* 8: 357-66.

