



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**Maestría y Doctorado en Ciencias Bioquímicas**

Regulación del gen *E2348C\_1013* de *Escherichia coli* enteropatógena

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias

PRESENTA:

Álvaro Damián Morales Ibarra

TUTOR PRINCIPAL

Dr. José Luis Puente García  
Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR

Dr. Enrique Merino Pérez  
Instituto de Biotecnología, UNAM

Dra. Bertha González Pedrajo  
Instituto de Fisiología celular, UNAM

Cuernavaca, Morelos. Octubre, 2019.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## AGRADECIMIENTOS

La palabra “gracias” proviene del latín “*gratia*” cuyo significado etimológico correspondería a “honra o alabanza dedicada a otro sin más razón”, aunque en la actualidad su uso más común sea para expresar reconocimiento por un favor recibido. Para los fines de esta sección, me gustaría utilizar esta palabra con un significado ni tan poético y romántico como la primera definición, ni seco y frío como su uso popular. Siendo así, cada que se exprese gratitud, se transmita asimismo un sentir de admiración, afecto, apreciación y bendición a cada sujeto posteriormente mencionado por su apoyo, paciencia, palabras, tiempo o simplemente por su compañía; definitivamente cada uno de ellos tuvo su importancia para poder realizar esta meta e influyó para mi formación como académico y persona.

Gracias al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada (No. 845328) durante mis estudios de maestría, así como a la Dirección General de Asuntos de Personal Académico (DGAPA-IN213516) y a CONACyT CB-239659 y FC-2015-2/950 por los recursos otorgados para la realización de este proyecto. También al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la UNAM por el apoyo económico para presentar este trabajo en el XXXII Congreso Nacional de Bioquímica de la Sociedad Mexicana de Bioquímica A.C. en Ixtapuzihuatanejo, Guerrero.

Gracias al Instituto de Biotecnología (IBt) y a toda su comunidad por demostrar entrega, dedicación y solidaridad en cada momento, en especial al M. en T.I. Juan Manuel Hurtado Ramírez y al Ing. Roberto Pablo Rodríguez Bahena de la Unidad de Cómputo del IBt por su apoyo técnico a lo largo del programa.

Gracias al Grupo del Dr. José Luis Puente García, entre ellos al Biol. Francisco Javier Santana Estrada, Dra. Lucía Perezgasga Ciscomani y Dra. Alejandra Vázquez Ramos por su apoyo técnico y en el manejo de recursos en el laboratorio.

Al Dr. Puente en sí por aceptarme como su alumno, por su guía, enseñanzas y tiempo dedicado. A la Dra. Liliana Medina Aparicio por apoyo anímico, consejos y amistad. A mi compañera M.C. Stephanie Ortiz Jiménez por tanta paciencia, solidaridad, apoyo, risas y desahogos.

Gracias a E.E.E.P.A.J.A. (Édgar, Eduardo, Eliseo, Pako, Adrián, Jorge, Álvaro); acrónimo asignado a mi grupo de amigos de la generación. Definitivamente mi estadía en el IBt y en el programa no hubiera sido la misma sin ellos. Un grupo tan diverso, pero que se pudo integrar de manera esporádica y bonita. Desde reunirnos para desayunar, estudiar temas para exámenes, o discutir sobre la mejor forma de cortar un pastel con criterios científicos\*, siempre ha sido enriquecedor académicamente o personalmente su compañía y amistad. Gracias por volver al IBt un segundo hogar y no sólo un lugar de trabajo.

A todos los mencionados anteriormente, y a otros que no pues no cabrían en un espacio tan pequeño, me gustaría dedicarles más palabras que pudiesen expresar mejor los sentimientos que evocan en mí y por los que hacen menester su mención aquí; todo lo tengo que expresar únicamente con la palabra “gracias”, y a pesar de ser una sola palabra de siete letras, se entiende el mensaje.

\* F.G. Cutting a round cake on scientific principles. 1906. Nature. 75(173). <https://doi.org/10.1038/075173c0>

## DEDICATORIA

A mi esposa, la Lic. María de la Luz García Esquivel, quien nunca ha dudado en mostrarme su apoyo sin importar la situación. Gracias por la paciencia, por tanto amor, por los consejos y por esa capacidad casi inhumana que tiene para darme paz. Anteriormente dije que el IBt se convirtió en mi segundo hogar, y eso es porque en donde sea que ella esté, será el primero.

A mi familia, María de la Luz Ibarra Robles (Mamá), Gustavo Alfredo Morales Gómez (Papá), Gustavo Alfredo Morales Ibarra (Hermano) y Luis Ángel Morales Ibarra (Hermano). A mis papás por su apoyo incondicional, consejos y porque en los valores familiares, individuales y sociales que ellos me inculcaron, reside la persona que soy. A mis hermanos por darme un buen ejemplo de pasión, compromiso y dedicación. No creo ser capaz de poder describir qué tan orgulloso estoy de pertenecer a esta familia, y espero (intento) regresar un poco del sentimiento que ellos evocan en mí.

A todos mis amigos que me apoyan a la distancia y nunca han perdido la fe en mí. Demetrio, Sergio, David, Raúl, Cristy, etc. Es un honor ser su amigo, gracias por tanto, y espero compartir este logro y más por mucho tiempo.

## ÍNDICE

|                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------|----|
| RESUMEN .....                                                                                                                                    | 1  |
| ABSTRACT .....                                                                                                                                   | 3  |
| INTRODUCCIÓN .....                                                                                                                               | 5  |
| ANTECEDENTES .....                                                                                                                               | 8  |
| <b>La unión de GrlA a la secuencia ATGT en la región espaciadora del promotor de <i>ler</i> es esencial para su actividad regulatoria.</b> ..... | 8  |
| <i>E1013</i> es regulado por GrlA .....                                                                                                          | 10 |
| HIPÓTESIS .....                                                                                                                                  | 14 |
| OBJETIVO .....                                                                                                                                   | 14 |
| Objetivos específicos .....                                                                                                                      | 14 |
| METODOLOGÍA .....                                                                                                                                | 15 |
| Cepas, plásmidos y medios de crecimiento .....                                                                                                   | 15 |
| Generación de fusiones transcripcionales al gen reportero <i>cat</i> .....                                                                       | 15 |
| Preparación de extractos para ensayo de CAT .....                                                                                                | 16 |
| Cuantificación de proteína .....                                                                                                                 | 17 |
| Cuantificación de la actividad CAT .....                                                                                                         | 17 |
| RESULTADOS Y DISCUSIÓN .....                                                                                                                     | 23 |
| <b>La región -43+26 del gen <i>E1013</i> contiene los elementos necesarios para la activación mediada por GrlA</b> .....                         | 23 |
| <b>La A del motivo ATGT del promotor de <i>E1013</i> es esencial para la activación vía GrlA</b> .....                                           | 24 |
| <b>Promotores posiblemente regulados por GrlA están altamente asociados a la secuencia NRIR en patógenos <i>A/E</i></b> .....                    | 26 |
| <b>La región correspondiente a la región líder no codificante de <i>E1013</i> ejerce</b>                                                         |    |

|                                                                                                                       |    |
|-----------------------------------------------------------------------------------------------------------------------|----|
| un efecto represor sobre su expresión.....                                                                            | 38 |
| La expresión de <i>E1013</i> es independiente de la presencia de H-NS.....                                            | 39 |
| PerC activa modestamente al promotor de <i>E1013</i> .....                                                            | 41 |
| <i>E1013</i> se expresa en bajos niveles en DMEM .....                                                                | 43 |
| <i>E1013</i> es activo en medio mínimo N con glicerol .....                                                           | 44 |
| La expresión del <i>LEE</i> y <i>E1013</i> se estimula en medios mínimos con glicerol<br>como fuente de carbono ..... | 47 |
| <b>CONCLUSIONES</b> .....                                                                                             | 52 |
| <b>PERSPECTIVAS</b> .....                                                                                             | 54 |
| <b>BIBLIOGRAFÍA</b> .....                                                                                             | 55 |
| <b>ANEXOS</b> .....                                                                                                   | 63 |
| <b>Anexo 1.</b> Concentración de DNA por extracción con fenol-cloroformo.....                                         | 63 |
| <b>Anexo 2.</b> Preparación de células electrocompetentes y electroporación (Dower<br><i>et al.</i> 1998) .....       | 64 |
| <b>Anexo 3.</b> Composición del DMEM.....                                                                             | 65 |
| <b>Anexo 4.</b> Composición de medios mínimos. ....                                                                   | 66 |

## RESUMEN

*Escherichia coli* enteropatógena (EPEC) es un patógeno causante de diarrea principalmente en niños. EPEC pertenece a la familia A/E (*Attaching/Effacing*), cuya principal característica es la adherencia íntima a los enterocitos, la eliminación de las microvellosidades y la formación de una estructura tipo pedestal rica en actina. Los genes involucrados en este fenotipo están codificados dentro de la isla de patogenicidad *LEE* (*Locus of enterocyte effacement*, por sus siglas en inglés). El regulador global H-NS reprime la expresión del *LEE* mientras que *Ler*, un regulador codificado en el *LEE*, antagoniza a H-NS. *GrlA* y *GrlR* son reguladores también codificados en esta isla de patogenicidad. *GrlA* es un activador transcripcional que se une a la secuencia espaciadora localizada entre las cajas -35 y -10 del promotor de *ler*, que contiene un motivo ATGT esencial para la activación mediada por este regulador. *GrlR* es un homodímero que interactúa con un monómero de *GrlA*, evitando que éste active a *ler* y, en consecuencia, reprime la expresión del *LEE*. Por su parte, *PerC*, un regulador transcripcional codificado en el plásmido EAF, activa a *ler* de manera independiente a través de un mecanismo aún no definido.

Un análisis preliminar del transcriptoma de *GrlA*, sugirió que el gen *E2348C\_1013* (*E1013*) es regulado por *GrlA*. Este gen está anotado como una proteína hipotética, pero su región promotora comparte una alta identidad con el promotor de *ler*, incluyendo la presencia del motivo ATGT en la región espaciadora. Experimentos de complementación en *E. coli* K12, conteniendo fusiones transcripcionales del promotor de *E1013* al gen *cat* sin promotor, con plásmidos que codifican para diferentes reguladores de EPEC, mostraron que *GrlA*, y la región reguladora comprendida de la posición -43 a +26, respecto al inicio de transcripción, son suficientes para activar la expresión de *E1013*. Una mutación en la primera base (A) de la secuencia ATGT reduce drásticamente la activación de *E1013* por *GrlA*, confirmando la funcionalidad de este motivo para la expresión de este gen. A su vez, la región no traducida (UTR) regula negativamente la expresión de *E1013* de manera independiente a H-NS. Por su parte, *PerC* tiene un efecto al parecer irrelevante en la activación de este gen, en comparación al de *GrlA* y contrario al



que PerC muestra para *ler*. Una búsqueda *in silico* de genes posiblemente regulados por GrlA, reveló una alta correlación entre la presencia de la secuencia NRIR (*n*r*e* regulatory inverted repeat) corriente arriba de la caja -35, y la secuencia ATGT ubicada en la secuencia espaciadora corriente arriba de la caja -10 en el contexto de promotores tipo  $\sigma^{70}$ . Sin embargo, a pesar de que el promotor de *E1013* es activado por GrlA en un fondo heterólogo como *E. coli* K12, la actividad en EPEC silvestre es apenas medible con las fusiones a *cat*, tanto en LB como en DMEM, a diferencia de un promotor del *LEE* que es activado en DMEM. Basados en la observación de que EPEC muestra niveles considerables de secreción mediados por el SSTT en medio mínimo (MM), evaluamos la expresión de los promotores de *E1013* y *LEE2* en MM-N y MM-9, más glicerol o glucosa como fuente de carbono.. Así, se determinó que el medio mínimo con glicerol favorece significativamente la expresión de *E1013* y del operón *LEE2*, sugiriendo que, en estas condiciones, la fuente de carbono es un determinante en la regulación del *LEE* y de otros genes cuya expresión depende de los reguladores codificados en esta isla.

## ABSTRACT

Enteropathogenic *Escherichia coli* (EPEC) is a pathogen that causes diarrhea mainly in children. EPEC belongs to the A/E (Attaching and Effacing) family of bacterial pathogens, which hallmark is the intimate adhesion to enterocytes, the elimination of microvilli, and the formation of an actin-rich pedestal-like structure on the surface of enterocytes. The genes involved in this phenotype are encoded within the LEE pathogenicity island (Locus of enterocyte effacement). The global regulator H-NS represses the expression of the LEE while the LEE-encoded regulator (Ler) antagonizes H-NS-mediated repression. GrIA and GrIR are regulators also encoded within the LEE. GrIA is a transcriptional activator that binds to the spacer sequence located between the -35 and -10 boxes of the *ler* promoter, which contains an ATGT motif essential for activation mediated by this regulator. GrIR is a homodimer that interacts with a GrIA monomer, preventing it from activating the *ler* promoter and, consequently, represses the expression of the LEE. In addition, PerC, a transcriptional regulator encoded in the EAF plasmid, independently activates *ler* through a mechanism not yet defined.

A preliminary analysis of the GrIA transcriptome suggested that the *E2348C\_1013* (*E1013*) gene is regulated by GrIA. This gene is annotated as a hypothetical protein, but its promoter region shares high identity with the *ler* promoter, including the presence of the ATGT motif in the spacer region. Complementation experiments in *E. coli* K12, containing transcriptional fusions of the E1013 promoter to the promoterless *cat* gene, with plasmids encoding different EPEC regulators, showed that GrIA, and the regulatory region spanning positions -43 to +26, are sufficient to activate E1013 expression. The mutation of the ATGT first base dramatically reduces GrIA-mediated E1013 activation, confirming this motif's functionality. In turn, the untranslated region (UTR) negatively regulates the expression of E1013 independently to H-NS. In contrast, PerC has, apparently, an irrelevant effect on the activation of this gene, compared to that of GrIA and contrary to what PerC shows for *ler*. An *in silico* search of genes possibly regulated by GrIA, revealed a high correlation between the presence of the NRIR (nle regulatory inverted repeat)

sequence upstream of the -35 sequence, and the ATGT motif located in the spacer sequence upstream of the -10 box in the context of  $\sigma 70$  type promoters. However, although the E1013 promoter is activated by GrlA in a heterologous background such as *E. coli* K12, the activity in wild type EPEC is barely measurable with cat fusions, both in LB as in DMEM, unlike a LEE promoter that is activated in DMEM. Based on the observation that EPEC shows considerable levels of secretion mediated by the TTSS in minimal medium (MM), we evaluated the expression of the E1013 and LEE2 promoters in MM-N and MM-9, plus glycerol or glucose as carbon source. These experiments revealed that MM with glycerol significantly favors the expression of both the *E1013* and *LEE2* promoters even more than DMEM, suggesting that the carbon source is a regulatory determinant of the LEE and other genes whose expression depends on the regulators encoded by this island. These results are allowing us to look into variations of GrlA action mechanism at different promoters, as well as into the role of the carbon source in LEE regulation.

## INTRODUCCIÓN

*Escherichia coli* enteropatógena (EPEC) y *E. coli* enterohemorrágica (EHEC) son importantes agentes causantes de diarrea infecciosa en humanos. EPEC causa diarrea severa y persistente, particularmente en niños menores de 2 años, sobre todo de países en desarrollo, mientras que EHEC causa colitis hemorrágica y puede desarrollar síndrome urémico hemolítico, el cual puede ser fatal (Gally y Stevens. 2017; Ochoa y Contreras. 2011). Ambos patotipos de *E. coli* junto con el patógeno murino *Citrobacter rodentium*, pertenecen a la familia de patógenos A/E (*Attaching/Effacing*), los cuales se adhieren a las células epiteliales del intestino formando estructuras ricas en actina denominadas “pedestales” y eliminando las microvellosidades. Sin embargo, estos patógenos afectan distintas regiones del intestino: EPEC coloniza preferentemente el intestino delgado (Pearson *et al.* 2016), mientras que EHEC (Gally y Stevens. 2017) y *C. rodentium* (Crepin *et al.* 2016) el intestino grueso de humano y ratón, respectivamente. El modelo de infección canónico de EPEC se ha descrito en tres etapas: (I) la adherencia inicial a los enterocitos mediada por la fimbria BFP, la cual está codificada en el plásmido *EAF*; (II) la translocación de proteínas al enterocito, entre ellas Tir, mediada por el sistema de secreción tipo 3 (SST3); (III) la adherencia íntima entre EPEC y el enterocito gracias a la interacción de Tir, el cual se encuentra en la membrana celular del hospedero, e Intimina, que se encuentra en la membrana externa de la bacteria (figura 1) (Coxen *et al.* 2013).

Los genes necesarios para el ensamblaje del SST3, así como otros factores de virulencia, están codificados en la isla de patogenicidad LEE (*Locus of Enterocyte Effacement*) (Gaytán *et al.* 2016). Los reguladores maestros de los genes de LEE son el regulador negativo H-NS (*Histone-like nucleoid-structuring protein*) y el regulador positivo Ler (*LEE encoded regulator*) como regulador positivo (figura 2). H-NS reprime la expresión uniéndose al DNA donde su oligomerización puede resultar en la formación de filamentos nucleoproteicos rígidos. La interacción entre los oligómeros de H-NS situados en distintas regiones del genoma puede resultar

en la formación de horquillas (Winardhi *et al.* 2014). Ler antagoniza la represión mediada por H-NS interaccionando con el DNA en sitios vecinos al complejo nucleorrepositor (Bustamante *et al.* 2001). Una vez que la concentración de Ler rebasa los niveles necesarios para desreprimir a sus genes blanco, es capaz de unirse a la región promotora del operón *LEE1* y autorregularse negativamente (figura 2) (Berdichevsky *et al.* 2005). Los dos principales activadores transcripcionales de *ler* son PerC y GrlA (Bustamante *et al.* 2011). PerC está codificado en el plásmido EAF y su mecanismo de acción se desconoce hasta el momento. *grlA* y *grlR* conforman el operón *LEE7 (grlRA)* (figura 2). Se sabe que GrlA se une al promotor de *ler* a través de su dominio HTH (*Helix-Turn-Helix*), activando su transcripción y antagonizando la represión mediada por H-NS (Jiménez *et al.* 2010; Islam *et al.* 2011), lo que establece un circuito de regulación positiva en el cual Ler activa el operón *grlRA* y GrlA activa a *ler* (figura 3) (Barba *et al.* 2005). Por su parte, GrlR forma homodímeros en estado nativo e interacciona directamente con el dominio HTH de un monómero de GrlA, formando un heterotrímero (figura 3). Esta interacción evita que GrlA se una al promotor de *ler* y, por tanto, que ejerza su función como regulador positivo, por lo que GrlR es un regulador negativo de los genes de LEE (Padavannil *et al.* 2013).

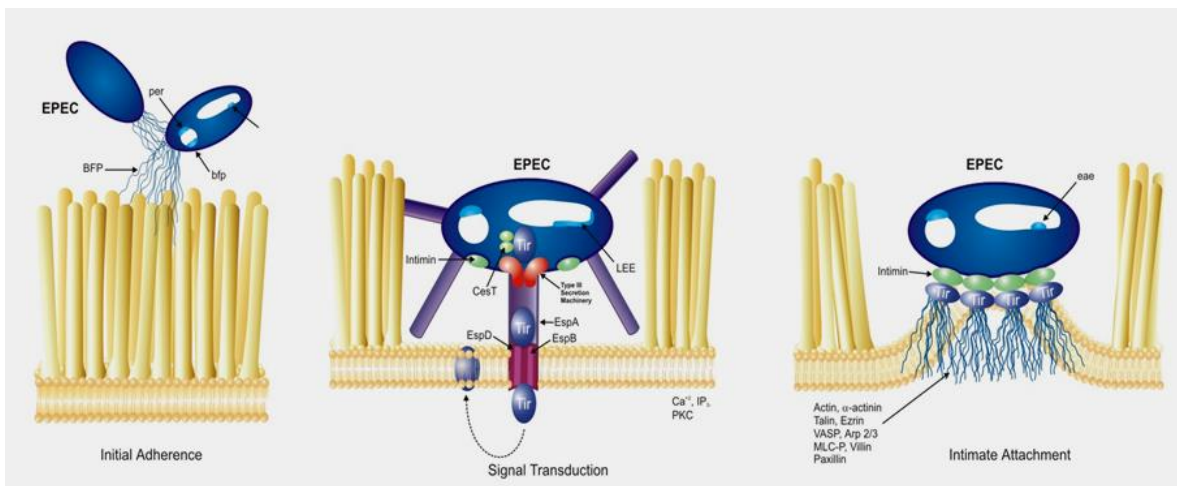


Figura 1. Modelo de infección por EPEC en tres etapas (ver texto).

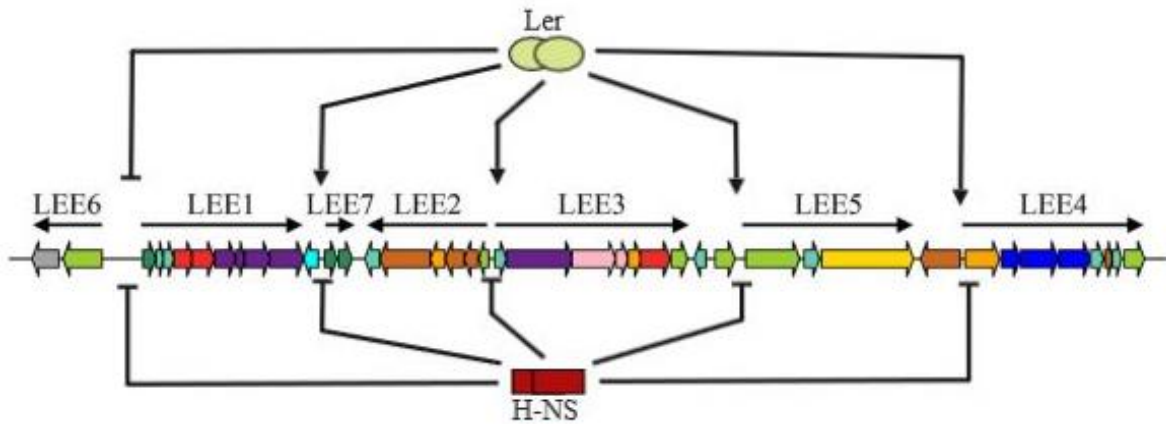


Figura 2. Regulación global del *LEE* mediada por H-NS (negativa) y Ler (positiva). Imagen tomada y modificada de Gaytán *et al* (2016).

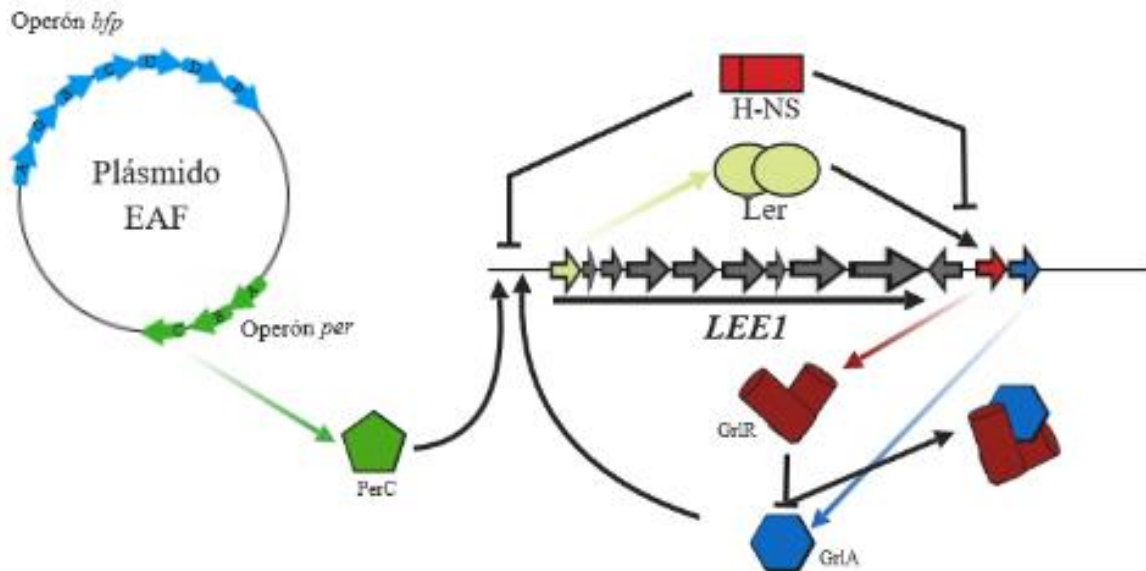


Figura 3. Regulación del operón *LEE1* y *LEE7*. PerC contrarresta la represión de H-NS en el promotor de *LEE1*, activando la expresión de *ler*. Ler antagoniza a H-NS en el resto de los operones del LEE, entre ellos al operón *grlRA* (*LEE7*). GrlA activa al promotor de *LEE1* formando un circuito de regulación positiva Ler-GrlA, mientras GrlR forma homodímeros e interactúa con GrlA; evitando así su unión a DNA.

## ANTECEDENTES

**La unión de GrIA a la secuencia ATGT en la región espaciadora del promotor de *ler* es esencial para su actividad regulatoria.**

GrIA es una proteína con dominio HTH capaz de activar la expresión del *ler*, para evaluar si esta regulación era directa Jiménez *et al* (2010) realizaron ensayos tipo EMSA (*Electrophoretic Mobility Shift Assay*) con las regiones *ler*-260/-50 y *ler*-50/+217 en presencia de la proteína recombinante GrIA fusionada a la proteína MBP (*Maltose binding protein*) (MBP-GrIA) (figura 4). La ausencia de retardo en la movilidad de *ler*-260/-50 en presencia de MBP-GrIA indicó que esta región no contiene un sitio de unión para GrIA. Sin embargo, la presencia de MBP-GrIA provocó un retardo en la movilidad de la región *ler*-50/+217, indicando que esta región contiene al menos un sitio de unión para GrIA.

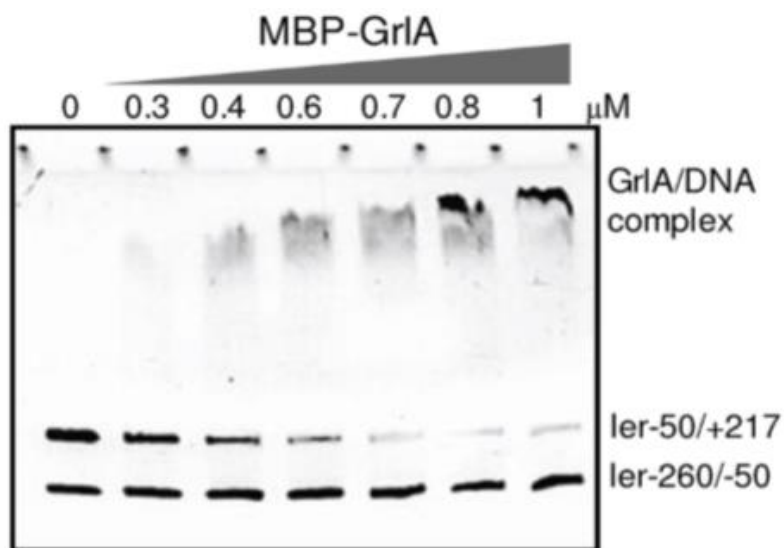


Figura 2. **GrIA interactúa con la región reguladora de *ler*.** EMSA con fragmentos de la región reguladora de *ler* y MBP-GrIA purificada. Imagen tomada de Jiménez *et al* (2010).

Posteriormente, Islam *et al* (2011) demostraron mediante mutagénesis al azar que el motivo ATGT localizado en la región espaciadora entre las cajas -10 y -35 del promotor de *ler* (señalado con asteriscos en la figura 5), es importante para la activación de *ler* mediada por GrlA. Dichos experimentos fueron realizados en la cepa de laboratorio *E. coli* M182 que expresaba de manera ectópica al gen *frlA* y una fusión transcripcional del promotor de *ler* al gen reportero *lacZ*. En este trabajo los autores no demostraron la importancia del motivo ATGT para la unión específica de GrlA al promotor de *ler*.

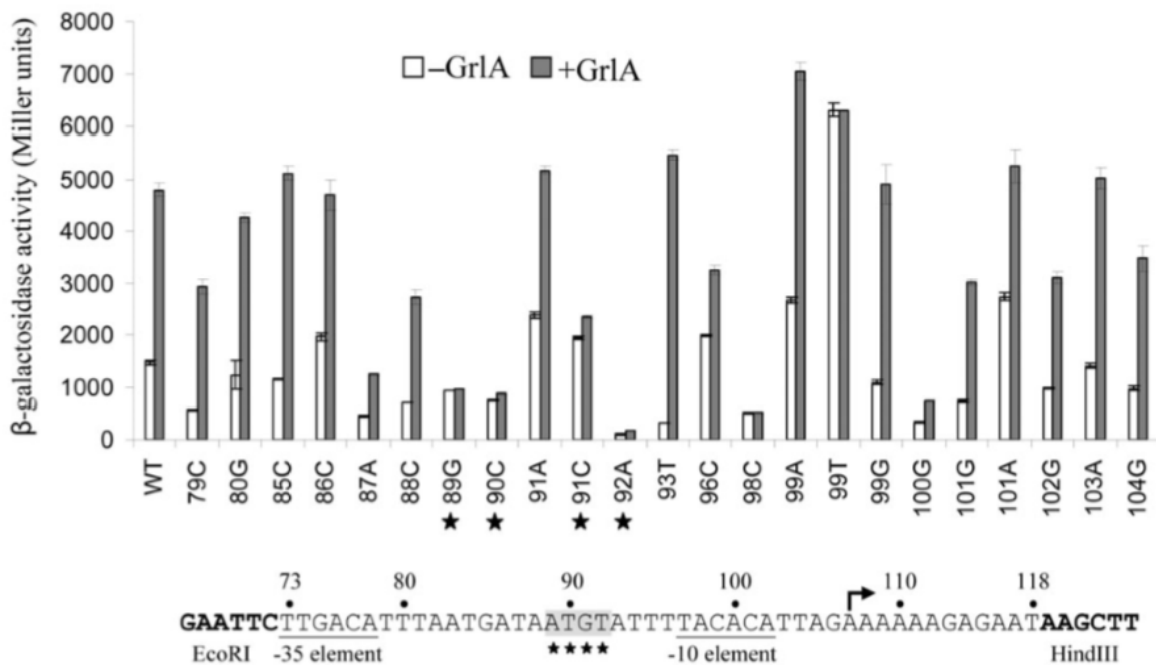


Figura 3. La secuencia ATGT localizada en la región espaciadora del promotor de *ler* es esencial para la activación mediada por GrlA. Ensayo de actividad de β-galactosidasa de fusiones transcripcionales delimitadas por los sitios de restricción EcoRI y HindIII señalados en negritas, mutagenizadas al azar del promotor de *ler* a *lacZ*. Imagen tomada de Islam *et al* (2011).

Con ambas evidencias, se sugiere fuertemente que GrlA se une directamente al promotor de *ler* (figura 4) y que la secuencia ATGT localizada en la región espaciadora juega un papel esencial durante la activación mediada por GrlA y, por tanto, probablemente en la unión de la proteína (figura 6).



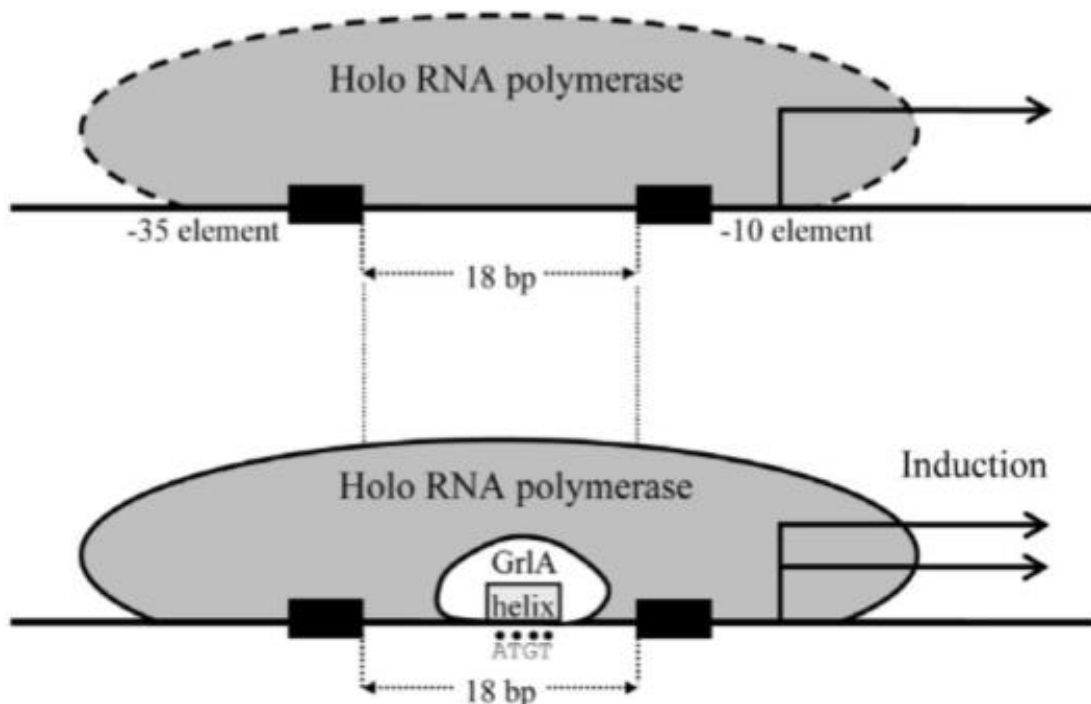


Figura 4. Esquema representando el posible mecanismo de acción de GrlA con el promotor de *ler*. La secuencia espaciadora contiene el motivo ATGT esencial para la activación mediada por GrlA, así como una base de más con respecto a los 17 pb que presenta la mayoría de promotores de *E. coli*. Se propone que la unión de GrlA al promotor compensa la presencia de dicha base adicional permitiendo entonces la correcta interacción de la RNA polimerasa; sin embargo, dicho mecanismo no ha sido caracterizado por completo. Imagen tomada de Islam *et al* (2011).

### ***E1013* es regulado por GrlA**

En los medios LB (*Lysogeny Broth*) y DMEM (*Dulbecco's Modified Eagle's Medium*) los genes de virulencia de EPEC se reprimen e inducen, respectivamente, incluyendo los codificados en la isla LEE (Bustamante *et al.* 2001) y en el plásmido EAF (Puente *et al.* 1996; Martínez-Laguna *et al.* 1999). Con la finalidad de identificar genes regulados por Ler, GrlA y GrlR, se realizó un experimento de *RNA-seq* con RNA de la cepa *E. coli* E2348/69 (EPEC) y de sus mutantes  $\Delta ler$ ,  $\Delta grlA$  y  $\Delta grlR$ , crecidas en LB y DMEM (Cervantes-Rivera, R., datos no publicados). Uno de los

genes que mostró mayor regulación diferencial entre la cepa silvestre y sus mutantes fue el gen de función desconocida *E2348C\_1013* (*E1013*). Para el caso de la mutante  $\Delta grlR$  se observó un aumento cercano a 6 veces la expresión de *E1013* en LB, como se esperaría al ser éste un regulador negativo, mientras que en DMEM su actividad decae en las mutantes  $\Delta grlA$  y  $\Delta ler$ , aunque en esta última la disminución no es tan pronunciada. En conjunto, estos datos sugirieron que *E1013* es regulado directamente por GrlA y no por Ler (figura 7).

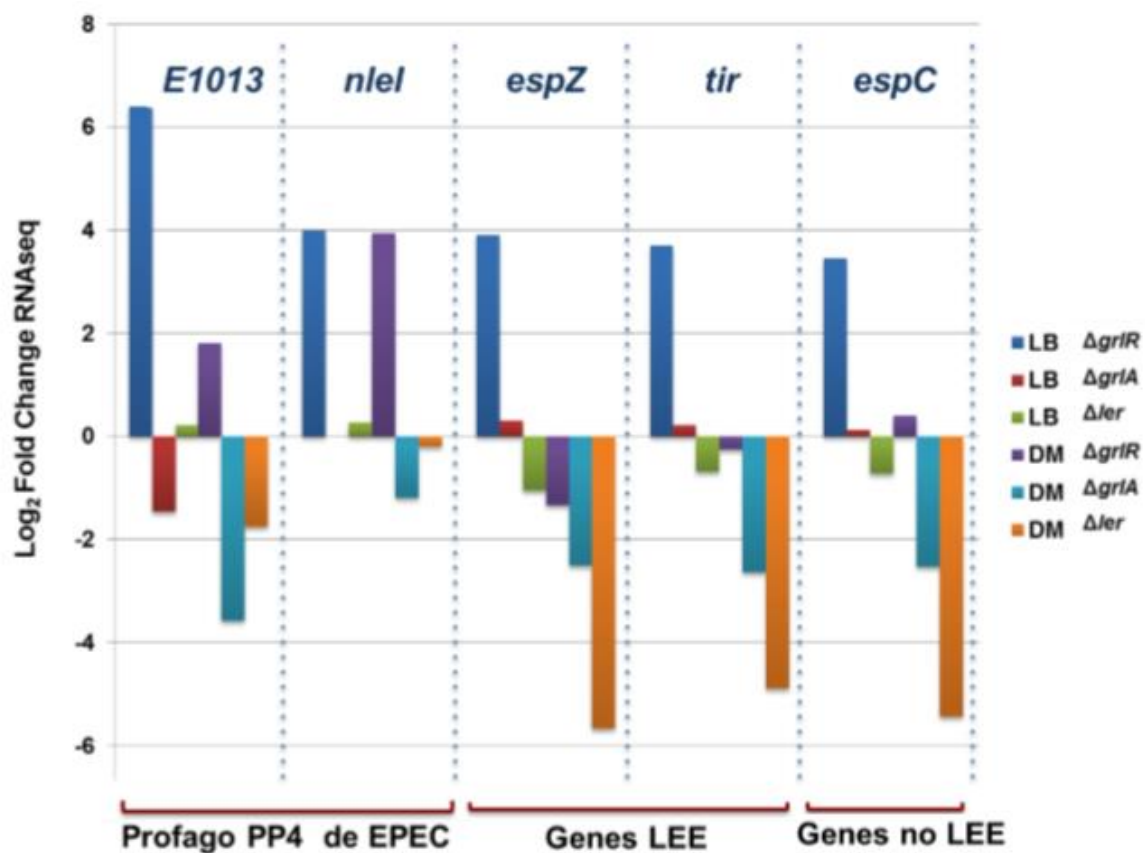


Figura 5. Análisis de RNA-seq proveniente de cultivos crecidos en matraz hasta una  $DO_{600}$  de 0.3. *nleI* es regulado por GrlA, mientras *espZ*, *tir* y *espC* son regulados por Ler, por ello son mostrados como controles. La expresión de las mutantes fue contrastada con la expresión de la cepa silvestre en el mismo medio. Cervantes-Rivera, R. (datos no publicados).

*E1013* es un gen codificado en el profago 4 (PP4), mismo en donde se encuentran los genes *nleG*, *nleB*, *nleC* y *nleD* que codifican para efectores conocidos como Nle (Non-LEE encoded effector), pero que son sustratos del SST3 (figura 8) (Pearson

et al. 2016). Las regiones reguladoras de *E1013* y *ler* están altamente conservadas. Destaca la presencia del motivo ATGT en ambas regiones, este motivo, como se mencionó anteriormente, es crucial durante la activación de *ler* mediada por GrIA (figura 9) (Islam et al. 2011). El papel de los reguladores GrIA y GrIR en la regulación de la expresión de *E1013* se ha corroborado mediante el estudio de una fusión transcripcional de la región reguladora -495/+26 al gen. En similitud a otros genes de virulencia en EPEC, la expresión de el gen *E1013* se vió reprimida en medio LB. Esta represión se pierde en la cepa  $\Delta grIR$ . Este esquema regulatorio es consistente con el modelo de regulación propuesto para circuitos dependientes de Ler y GrIA-GrIR. En medio DMEM, donde se espera activación de los genes de virulencia de EPEC, se observó expresión de *E1013*. Esta expresión es dependiente de *grIA* (figura 10A).

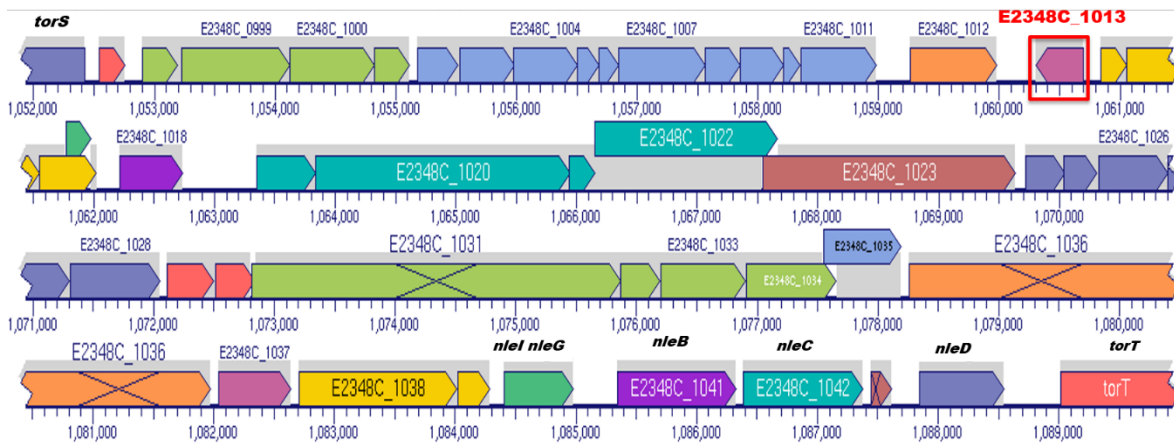


Figura 6. Contexto genómico de *E1013* que forma parte del profrago PP4 de EPEC en la cepa E2348/69. Se indica la posición de los genes *nleG*, *nleB*, *nleC* y *nleD*, los cuales codifican para efectores Nle previamente reportados como sustratos del SST3. Los genes *torS* y *torT* marcan los límites del PP4. Imagen obtenida de biocyc.org.

Para obtener evidencia de regulación directa mediada por GrIA, se llevaron a cabo experimentos de expresión heteróloga de la fusión *E1013-cat* en *E. coli* MC4100 en presencia de los plásmidos pMPM-T3(pT3), pT3/*grIA* y pT3/*grIR* (figura 10B). Únicamente la presencia del plásmido pT3/*grIA* fue capaz de promover la expresión

de la fusión *E1013-cat*. Este resultado sugiere que GrIA es un regulador directo de *E1013* (Cadena, E. 2016).

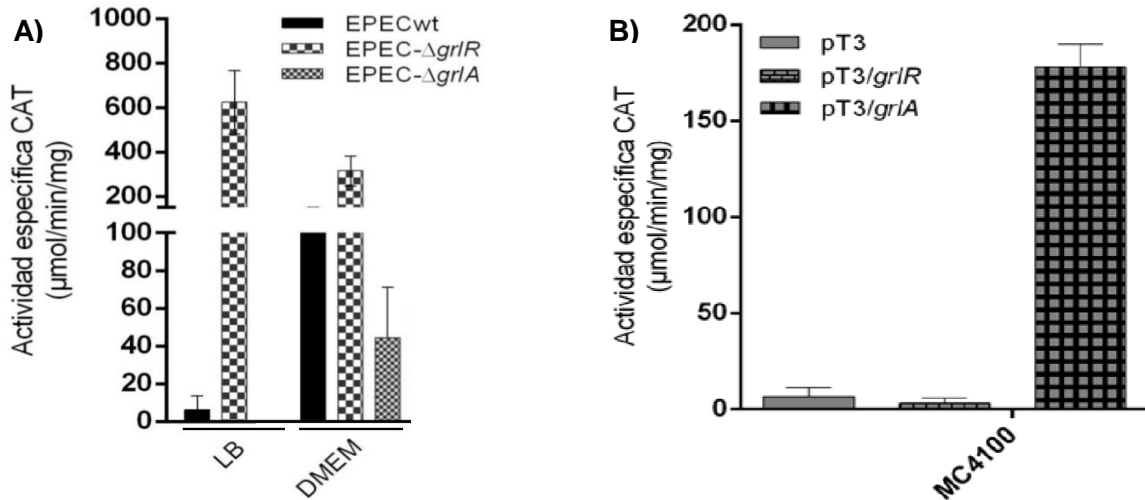


Figura 10. Ensayo de actividad específica de CAT con fusión pKKE1013-495+26 transformada en A) la cepa parental de EPEC,  $\Delta\text{grlA}$  y  $\Delta\text{grlR}$  en medio LB y DMEM, y B) en una cepa de *E. coli* MC4100 complementada con el vector vacío (pT3), pT3/*grlR* y pT3/*grlA* en medio LB. Tomada y modificada de Cadena E. 2016.

Además de *ler*, antes de *E1013*, no se había reportado un gen que fuera regulado directamente por GrIA. Estudiar los mecanismos de acción de GrIA sobre el promotor de *E1013* nos permitirá conocer más sobre su papel de factor transcripcional.

## HIPÓTESIS

La expresión del gen *E1013* de EPEC es regulada en *trans* por los reguladores GrlA y PerC, y en *cis* por el motivo ATGT localizado en la región espaciadora de su promotor.

## OBJETIVO

Caracterizar los mecanismos de regulación del gen *E1013* de EPEC.

### Objetivos específicos

- Identificar y caracterizar elementos de regulación en *cis* del gen *E1013*, incluyendo aquellos que determinan la interacción de GrlA con el promotor.
- Evaluar el papel de PerC en la regulación de *E1013*.
- Determinar condiciones de crecimiento que favorezcan la expresión de *E1013*.

## METODOLOGÍA

### Cepas, plásmidos y medios de crecimiento

Las cepas y plásmidos utilizados en este proyecto se muestran en las tablas 1 y 2, respectivamente. Se utilizó medio LB (10g/L Bacto<sup>®</sup> Triptona de Caseína, 5g/L Bacto<sup>®</sup> Extracto de levadura, 10g/L Golden Bell Reactivos<sup>RM</sup> Cloruro de sodio), Gibco<sup>®</sup> DMEM (0.45 % de glucosa, L-glutamina(+), piruvato de sodio (-)) (anexo 3), medio mínimo N (MM-N) (anexo 4) (Nelson y Kennedy. 1970) y medio mínimo 9 (MM-9) (anexo 4) (Li *et al.* 2000). Sigma-Aldrich<sup>®</sup> Ampicilina (0.2mg/ml), Sigma-Aldrich<sup>®</sup> Tetraciclina (0.012mg/ml) o Sigma-Aldrich<sup>®</sup> Estreptomina (0.1 mg/ml) fueron añadidos a los medios cuando era necesario.

### Generación de fusiones transcripcionales al gen reportero *cat*

Las fusiones al gen reportero *cat* fueron construidas generando fragmentos de DNA por PCR (*Polymerase Chain Reaction*). Las condiciones de PCR fueron de 1 ciclo a 95 °C por 5 minutos en la primera etapa, 30 ciclos a 95 °C por 30 seg, 56 °C por 30 seg y 72 °C por 30 seg en la segunda etapa, finalizando con 1 ciclo a 72 °C por 5 min. Los reactivos para la PCR fueron los siguientes: Oligonucleótidos 0.4 μM (tabla 3) (sintetizados en la Unidad de Síntesis y Secuenciación del IBt de la UNAM), MgCl<sub>2</sub> 2 mM, dNTPs 0.2 mM (Thermo Scientific<sup>TM</sup>), buffer PCR 1X (Buffer PCR 10X: Thermo Scientific<sup>TM</sup>) y DNA Taq Polimerasa 0.3 U (Thermo Scientific<sup>TM</sup>). DNA genómico de EPEC o del plásmido pKKE1013-495+26 (figura 11) fue utilizado como molde para la generación de fusiones con o sin la secuencia líder, respectivamente (figura 11) (Cadena, E. 2016). Los fragmentos generados fueron purificados utilizando el método de extracción por fenol-cloroformo (anexo 1), digeridos con 10

U de BamHI y HindIII (Thermo Scientific™) y nuevamente purificados. Dichos fragmentos fueron ligados al vector pKK232-8, el cual tiene el gen *cat* (cloranfenicol acetiltransferasa) sin promotor (Pharmacia LKB Biotechnology), digerido con las mismas enzimas, utilizando 1 U por reacción de la T4 DNA *Ligase* (Thermo Scientific™). Los productos de ligación se utilizaron para transformar cepas electrocompetentes (anexo 2) de *E. coli* MC4100; aquellas colonias que crecieron en presencia de ampicilina se analizaron por PCR para confirmar la presencia de inserto, utilizando los oligonucleótidos pKKseq1 y pKKseq2 (tabla 3). Clonas positivas fueron secuenciadas en la Unidad de Secuencia y Síntesis del IBt de la UNAM para confirmar la integridad del inserto.

### **Preparación de extractos para ensayo de CAT**

De cultivos crecidos en las diferentes condiciones probadas en este estudio, se obtuvieron muestras de 1.5 ml a los tiempos indicados para cada experimento, las cuales fueron transferidas a un tubo Eppendorf de 1.5 ml. La pastilla celular se concentró por centrifugación a 12,000 rpm durante 2 minutos. Para quitar restos del medio de cultivo, la pastilla se lavó con 800 µl de buffer TDTT (tris-HCl 50 mM pH 7.8 y dithiothreitol 30 µM), se centrifugó y se desechó el sobrenadante. Cada pastilla se resuspendió en 300 µl de buffer TDTT, se sonicó durante 3 minutos con lapsos de 10 seg en encendido y 5 seg en apagado, se centrifugó a 13,500 rpm durante 15 min y se rescató el sobrenadante en un tubo Eppendorf de 1.5 ml nuevo. Las muestras se guardaron a -20°C en caso de no utilizarse inmediatamente.

## Cuantificación de proteína

Se agregaron 10 µl en cada pozo de los lisados en duplicado en una microplaca de 96 pozos utilizando buffer DTT como blanco. Se añadió 200 µl del reactivo del kit Pierce® BCA Protein Assay (Thermo Scientific™; según indicaciones del fabricante) a cada pozo. Se incubó a 37°C durante 30 min y se leyó en el lector de microplacas de 96 pozos ELx808 (Bio-Tek®) a 562 nm (Martínez-Laguna *et al.* 1999).

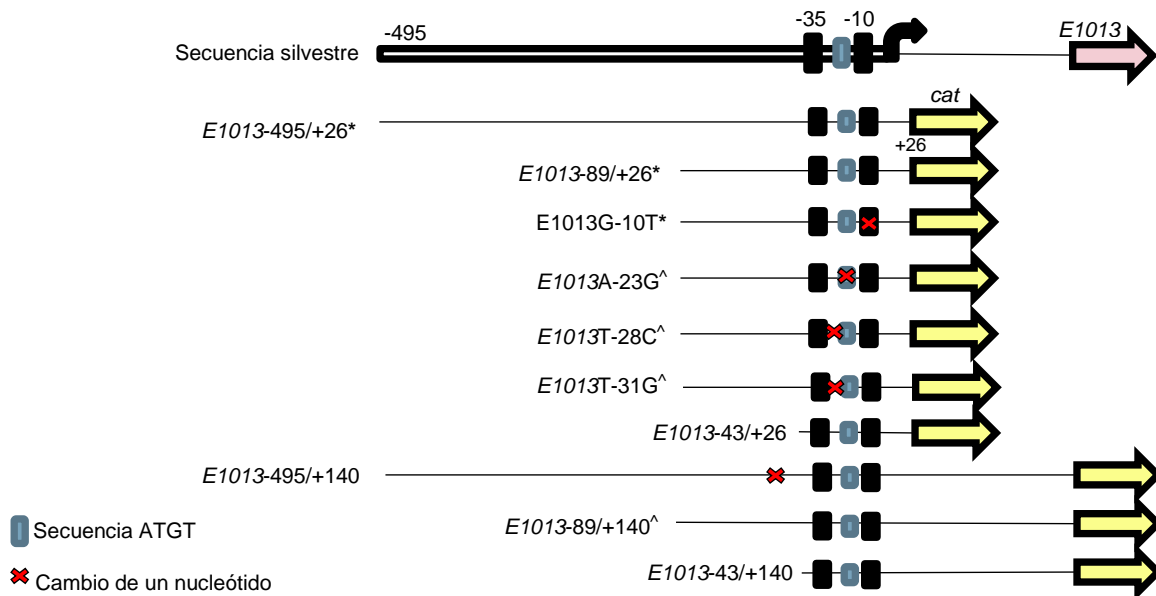


Figura 11. **Fusiones transcripcionales de E1013 al gen reportero *cat*.** Las coordenadas de cada fusión son con respecto al inicio de transcripción reportado previamente (Cadena E. 2016). (\*) indica fusiones previamente construidas y reportadas (Cadena, E. 2016). (^) indica plásmidos construidos por Cadena, E. pero no reportados anteriormente.

## Cuantificación de la actividad CAT

Para determinar la actividad de CAT, se agregaron 5 µl de cada lisado por pozo en una microplaca de 96 pozos por duplicado. Posteriormente, se añadieron 200 µl de



la solución para la reacción (1 mM DNTB, 0.1 mM acetil CoA, 0.1 mM cloranfenicol, 0.1 M tris-HCl pH 7.8) y se leyó durante 5 minutos, cada 15 segundos, a 412 nm en el lector de microplacas de 96 pozos ELx808 en modo de cinética (Bio-Tek®). Para calcular la actividad específica de CAT se divide la actividad sobre la cantidad de proteína. Todos los valores aquí presentados son resultado de al menos tres experimentos independientes, cada uno con su réplica biológica.

### **Búsqueda de promotores posiblemente regulados por GrlA**

Utilizando *perl* como lenguaje de programación y la consola BASH de Ubuntu para su ejecución, se realizó una búsqueda de secuencias que cumplieran las características aparentes de un promotor regulado por GrlA (figura 14): una caja -35 conservada respecto a un promotor  $\sigma^{70}$  TTG[A/T]CA (McClure, W.R. 1985), una caja -10 separada por 18 pb con al menos una G o C que podía tener hasta 3 cambios respecto a su secuencia consenso TATAAT siempre que conservara 4 A/T, y la secuencia AT[A/G]T a una distancia de 4 o 5 pb río arriba de la -10. Esto último debido a que en *C. rodentium* dentro del promotor de *ler* está la secuencia ATAT y no el ATGT que encontramos en *ler* de EPEC y EHEC, o *E1013* (tabla 4); sin embargo, GrlA activa a *ler* en *C. rodentium*. Por otro lado, Islam *et al* (2011) mostró que este cambio en el promotor de *ler* en EPEC no afectaba la actividad de GrlA sobre éste.

Del GenBank se descargaron los siguientes genomas, los cuales fueron utilizados para hacer la búsqueda por ser cepas referencia a patógenos A/E, a excepción de *E. coli* MG1655:

- *Escherichia coli* O127:H6 E2348/69 (EPEC)
- *Escherichia coli* O55:H7 CB9615 (EPEC atípica)
- *Escherichia coli* O157:H7 Sakai (EHEC)

- *Escherichia coli* O157:H7 EDL933 (EHEC)
- *Escherichia coli* O111:H- str. 11128 (EHEC)
- *Escherichia coli* O26:H11 str. 11368 (EHEC)
- *Escherichia albertii* KF1
- *Citrobacter rodentium* ICC168
- *Escherichia coli* MG1655

### **Búsqueda de secuencias NRIR asociadas a un promotor**

Utilizando *perl* como lenguaje de programación y la consola BASH de Ubuntu para su ejecución, se generó un algoritmo que identificara la secuencia NRIR (*nr*/*le regulatory inverted repeat*) (figura 14) en los genomas mencionados anteriormente. Los parámetros se establecieron utilizando la secuencia “CAAAAACCGGAG” como la secuencia base del repetido invertido distal del NRIR, “TCCGG” para el espaciador y “CTCCGGTTTTTG” para el repetido invertido proximal a la caja -35 del promotor. Dichos parámetros incluyeron que puede haber un máximo de dos cambios en cada repetido invertido y 2 en el espaciador. Una vez que el programa encuentra un elemento NRIR, con base en el alineamiento de la figura 15 busca una secuencia -35 “TTGTCA” en su vecindad que permite máximo dos cambios, y una secuencia -10 “TATAAT” que permite 3 bases fuera del consenso pero debe de tener al menos 4 A/T, y la secuencia espaciadora entre la caja -10 y -35 debe ser de 17 ó 18 pb. Si la búsqueda no reconocía un promotor, ya fuera por falta de la caja -35 o -10, pero sí una secuencia NRIR, imprimía ésta en la pantalla más 30 pb rio abajo y la anotación de que no tiene un promotor  $\sigma^{70}$  asociado.

Tabla 1. Cepas utilizadas en este trabajo.

| Cepa                                          | Genotipo      | Referencia                     |
|-----------------------------------------------|---------------|--------------------------------|
| <i>E. coli</i> MC4100                         | Cepa parental | Casadaban, M. 1976.            |
|                                               | $\Delta hns$  | Bustamante <i>et al.</i> 2011. |
| <i>E. coli</i> enteropatógena (EPEC) E2348/69 | Cepa parental | Levine <i>et al.</i> 1978.     |

Tabla 2. Plásmidos utilizados en este estudio.

| Plásmido        | Descripción                                                                                                                                                                 | Referencia                     |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| pKK232-8        | Derivado de pBR322 con el gen <i>cat</i> sin promotor y un gen de resistencia a ampicilina.                                                                                 | Brosius, J. 1984.              |
| pler-260        | Derivado del pKK232-8. Codifica una fusión transcripcional <i>ler-cat</i> desde la región -253 a +216.                                                                      | Bustamante <i>et al.</i> 2011. |
| pSEPZ-11        | Derivado del pKK232-8. Codifica una fusión transcripcional con la región reguladora del operón <i>LEE2</i> .                                                                | Bustamante <i>et al.</i> 2001. |
| pKKE1013-495+26 | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -495 a +26.                                                                     | Cadena, E. 2016.               |
| pKKE1013-89+26  | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +26.                                                                      | Cadena, E. 2016.               |
| pKKE1013G-10T   | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +26. Presenta una mutación en la caja -10 del promotor (TAGAAA → TATAAA). | Cadena, E. 2016.               |

|                 |                                                                                                                                                                                                                  |                 |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| pKKE1013A-23G   | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +26. Presenta una mutación en la caja ATGT en el promotor (ATGT → GTGT).                                       | Este estudio.   |
| pKKE1013T-28C   | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +26. Presenta una mutación dentro del promotor en la posición -28 respecto al inicio de transcripción (T → C)  | Este estudio.   |
| pKKE1013T-31G   | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +26. Presenta una mutación dentro del promotor en la posición -31 respecto al inicio de transcripción (T → G). | Este estudio.   |
| pKK1013-43+26   | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -43 a +26.                                                                                                           | Este estudio.   |
| pKK1013-495+140 | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -495 a +140. Presenta una mutación en la posición -51 respecto al inicio de transcripción (C → T).                   | Este estudio.   |
| pKK1013-89+140  | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -89 a +140.                                                                                                          | Este estudio.   |
| pKK1013-43+140  | Derivado del pKK232-8. Codifica una fusión transcripcional <i>E1013-cat</i> desde la región -43 a +140.                                                                                                          | Este estudio.   |
| pMPM-T3         | Vector de clonación derivado de p15A de bajo número de copias. Codifica un gen de resistencia a kanamicina.                                                                                                      | Mayer, M. 1995. |

|           |                                                                                            |                                |
|-----------|--------------------------------------------------------------------------------------------|--------------------------------|
| pTEPGriA1 | Derivado del pMPM-T3. Codifica al gen <i>griA</i> regulado por el promotor de <i>lac</i> . | Bustamante <i>et al.</i> 2011. |
| pTEPPerC1 | Derivado del pMPM-T3. Codifica al gen <i>perC</i> regulado por el promotor de <i>lac</i> . | Bustamante <i>et al.</i> 2011. |

Tabla 3. Oligonucleótidos utilizados en este proyecto.

| Nombre                | Secuencia                                | Uso                                                                  |
|-----------------------|------------------------------------------|----------------------------------------------------------------------|
| pKKseq1               | CAACGGTGGTATATCCAGTG                     | Secuenciación de insertos en el pKK232-8. <i>Reverse</i> .           |
| pKKseq2               | GAGGCCCTTTCGTCTTCAAG                     | Secuenciación de insertos en el pKK232-8. <i>Forward</i> .           |
| <i>E1013-495-BHI*</i> | ATAG <b>GGATCC</b> ACGCCAACGATATCTTCGTAT | Generación de fusión. Sitio de restricción BamHI. <i>Forward</i> .   |
| <i>E1013-43-BHI</i>   | CGTCATAATACCG <b>GGATCC</b> TTATGATTGAC  | Generación de fusión. Sitio de restricción BamHI. <i>Forward</i> .   |
| <i>E1013+26-H3*</i>   | CCG <b>AAGCTT</b> TGGAGAAATGATATGTAACAC  | Generación de fusión. Sitio de restricción HindIII. <i>Reverse</i> . |
| <i>E1013+139-H3^</i>  | CCGTTGTGT <b>AAGCTT</b> TGAGATAGAAAGCAT  | Generación de fusión. Sitio de restricción HindIII. <i>Reverse</i> . |

En negritas se marcan los sitios de restricción. (\*) oligonucleótidos reportados por Cadena, E (2016). (^) oligonucleótidos diseñados por Cadena, E. pero no reportados anteriormente.

## RESULTADOS Y DISCUSIÓN

### **La región -43+26 del gen *E1013* contiene los elementos necesarios para la activación mediada por GrlA**

Cadena, E (2016) demostró que la fusión de pKKE1013-89+26 se comporta de la misma manera que la fusión pKKE1013-495+26 en una cepa de EPEC, tanto en LB como en DMEM, indicando que los elementos necesarios para la regulación de *E1013* están contenidos entre las posiciones -89 a +26 con respecto al inicio de transcripción.

Con el propósito de conocer si los elementos contenidos sólo en el promotor de *E1013* son suficientes para mediar su activación vía GrlA, como previamente se reportó para el promotor de *ler* (Islam *et al* (2011)), se construyó la fusión pKKE1013-43+26 que abarca desde la posición -43 a la posición +26 con respecto al inicio de transcripción (figura 11). Las fusiones pKKE1013-89+26, pKKE1013-495+26 y pKKE1013-43+26 fueron introducidas en *E. coli* MC4100 en conjunto con el plásmido vacío pMPM-T3 o el plásmido pTEPGrlA1 que expresa GrlA. La actividad de las tres fusiones fue dependiente de la presencia de GrlA y alcanzó niveles similares. Esto sugiere que todos los elementos necesarios para la activación del promotor de *E1013* por GrlA están contenidos dentro de la región -43 a +26 (figura 12).

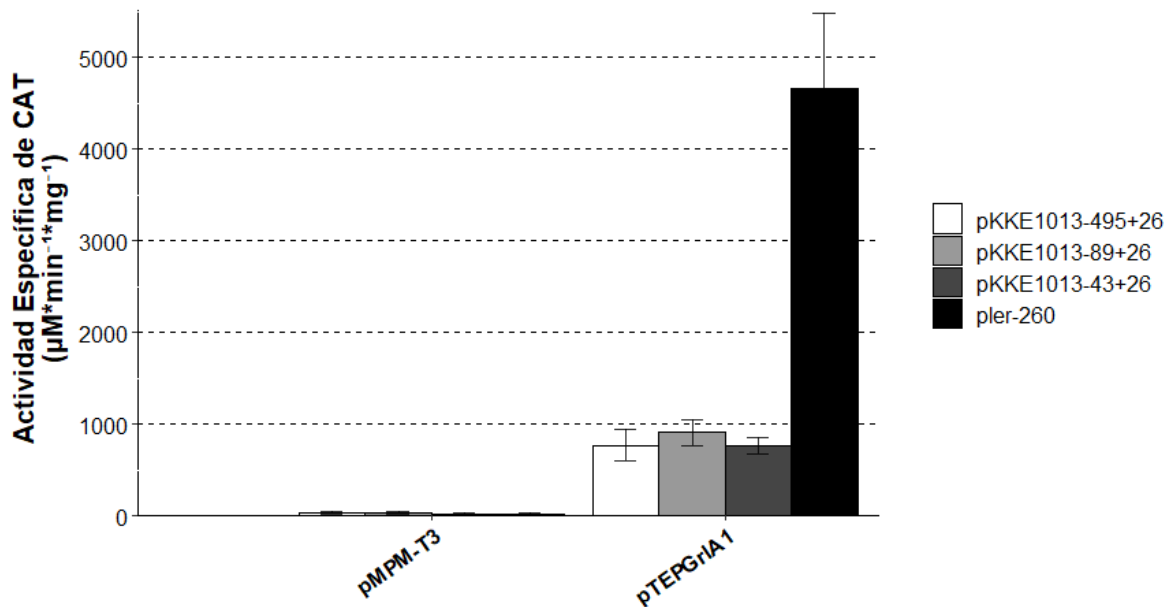


Figura 12. **La región -43+26 es suficiente para mediar la activación del gen *E1013* por GrlA.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de *E. coli* MC4100 conteniendo las fusiones indicadas más el plásmido pMPM-T3 o pTEPGrlA, crecidos en 5 ml de LB en tubo por 9 horas ( $DO_{600} \sim 0.93 \pm 0.073$ ) a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su duplicado biológico. Las barras de error representan la desviación estándar.

### La A del motivo ATGT del promotor de *E1013* es esencial para la activación vía GrlA

La hipótesis de este trabajo plantea que la secuencia ATGT contenida en la región espaciadora del promotor de *E1013*, es crítica para la activación mediada por GrlA. Para evaluar esta hipótesis, se utilizó la fusión pKKE1013-89+26, para introducir mutaciones puntuales en posiciones conservadas en la secuencia espaciadora de los promotores de *E1013* y *ler*. Los nucleótidos que se cambiaron fueron la A de la secuencia ATGT en la posición -22 por una G, la T en la posición -28 por una C y la T en la posición -31 por una G (figura 9).

Las actividades específicas de CAT generadas por las fusiones antes mencionadas en *E. coli* MC4100 con o sin GrlA se compararon con las obtenidas para la fusión

silvestre y la fusión pKKE1013G-10T (figura 9), la cual tiene una mutación de una G por una T en la caja -10 que acerca la secuencia de dicha caja a la del promotor consenso. Esta última mutación mejora la fuerza del promotor y le otorga actividad independiente de GrlA; aunque sigue presentando un incremento significativo de actividad en presencia de GrlA (Cadena, E. 2016).

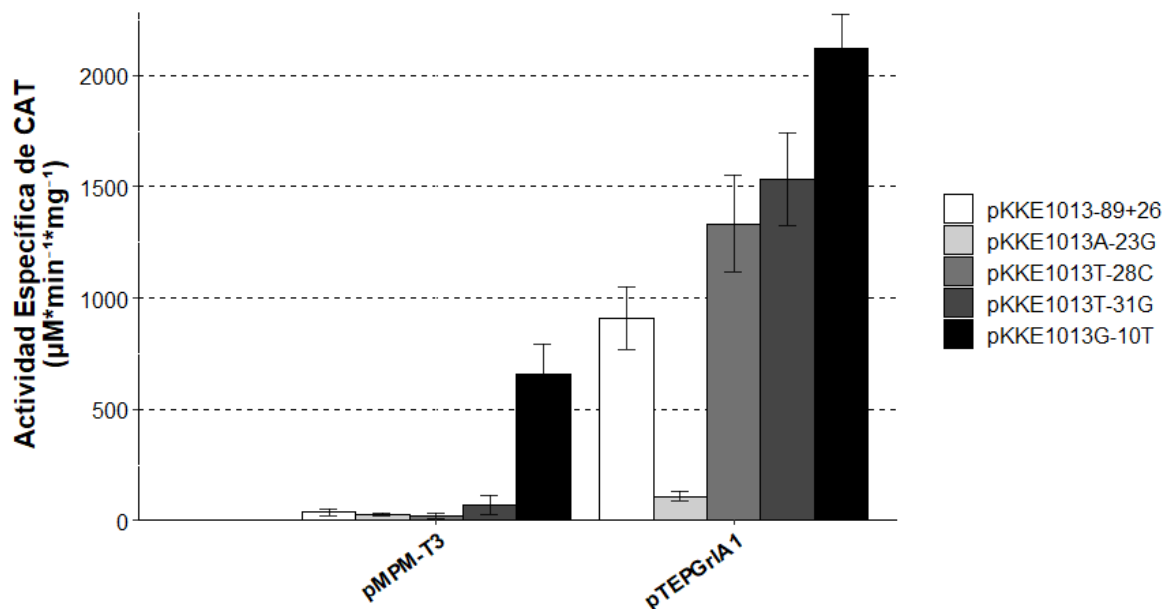


Figura 13. **La A del motivo ATGT del promotor de *E1013* es esencial para la activación mediada por GrlA.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de *E. coli* MC4100 conteniendo las fusiones indicadas más el plásmido pMPM-T3 o pTEPGrlA, crecidos en 5 ml de LB en tubo por 9 horas ( $DO_{600} \sim 0.93 \pm 0.073$ ) a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su duplicado biológico. Las barras de error representan la desviación estándar.

Como se había observado previamente, la expresión de la fusión pKKE1013-89+26 es dependiente de la presencia de GrlA, mientras que la expresión de la fusión pKKE1013G-10T es parcialmente independiente de GrlA aunque puede ser activada por la misma. Por su parte, las fusiones pKKE1013T-28C y pKKE1013T-31G sólo son dependientes de GrlA y muestran un nivel de expresión similar a la fusión pKKE1013-89+26. Esto sugiere que estos nucleótidos no están involucrados en la activación del promotor de *E1013* mediada por GrlA. En contraste, la expresión de la fusión pKKE1013A-23G está muy por debajo de la expresión en la fusión pKKE1013-89+26 aun en presencia de GrlA. Esto indica que este nucleótido es



importante para la expresión de *E1013* y probablemente para la activación mediada por GrlA. Estos datos están en acuerdo con las observaciones de Islam *et al* (2011) en relación a la importancia de este motivo en la expresión de *ler*. Experimentos con fusiones de *E1013* mutando los otros nucleótidos y ensayos de tipo EMSA son necesarios para confirmar el papel de la secuencia ATGT del promotor de *E1013* en la activación mediada por GrlA y, por tanto, en la interacción GrlA-promotor.

### **Promotores posiblemente regulados por GrlA están altamente asociados a la secuencia NRIR en patógenos A/E**

Además de *E1013* y *ler*, hasta el momento solamente se han reportado *flhDC* y *ehxCABD* como blancos de regulación negativa y positiva por GrlA, respectivamente. Sin embargo, en ningún caso se ha demostrado regulación directa (Iyoda *et al.* 2006; Saitoh *et al.* 2008). Por otro lado, García-Ángulo *et al* (2012) describieron que la activación de los genes *nleH1* y *nleB2* de EPEC, localizados en profagos fuera del LEE y cuyos productos codifican para efectores substrato del SST3 (denominados *non-LEE-encoded effectors*), requiere de una región invertida repetida conservada, denominada NRIR (*nle regulatory inverted repeat*), que está ubicada justo corriente arriba de la caja -35 de sus respectivos promotores. La activación de estos genes no se vio afectada significativamente en las cepas de EPEC  $\Delta$ *grlA* y  $\Delta$ *ler*, lo que llevó a la conclusión inicial de que su regulación era independiente de Ler y GrlA y que existía un regulador adicional que interactúa con el NRIR para activar estos genes. Sin embargo, en similitud a lo observado con el promotor de *E1013* y *ler* (Islam *et al.* 2011), la expresión de *nleH1* y *nleB2* puede ser activada en un sistema heterólogo a partir de la expresión ectópica de GrlA (figura 12) (González-Lara, tesis en curso).

La secuencia contenida entre las cajas -35 y -10 de los promotores que contienen la secuencia NRIR en EPEC, conserva características encontradas en los

promotores de *ler* y *E1013*, como son la secuencia ATGT, una caja -35 muy cercana al consenso y una caja -10 con al menos una G o C (figura 14).

El análisis de alineamiento múltiple llevó a proponer que un promotor regulado por GrlA debe presentar como características consenso 1) una caja -35 conservada “TTG[A/T]CA”, 2) una caja -10 conteniendo entre la tercera y sexta posición al menos una G/C, 3) una secuencia espaciadora de 18 pb, 4) y el motivo “AT[G/A]T” a 4 o 5 pb corriente arriba de la caja -10. El cambio de G a A en la tercera posición del motivo ATGT no afecta la activación del promotor de *ler* por GrlA (figura 5) (Islam *et al.* 2011; Barba *et al.* 2005). Utilizando estas características generamos un algoritmo para buscar promotores que pudieran ser blanco de regulación por GrlA en genomas de cepas de referencia pertenecientes a la familia A/E.



Figura 14. Alineamiento múltiple de las regiones reguladoras de los genes *ler*, *E1013* y *nles* que contienen el elemento NRIR. En verde se muestra el elemento palindrómico NRIR, en rojo las cajas -35 y -10 y en azul la secuencia ATGT necesaria para la activación de *ler* mediada por GrlA (Islam *et al.*, 2011, García-Ángulo *et al.*, 2012). A la derecha se indica el número de bases entre el promotor y el codón de inicio. Los asteriscos indican las bases conservadas.

Los resultados de esta búsqueda se presentan en la tabla 4. Identificamos 78 secuencias que cumplieron con las características especificadas y las utilizamos para generar una secuencia *logo* utilizando la plataforma de WebLogo (<https://weblogo.berkeley.edu/logo.cgi>) (figura 15) (Schneider y Stephens. 1999;

Crooks *et al.* 2004). Se puede apreciar el alto grado de conservación de la caja -35, y su cercanía al consenso de los promotores regulados por sigma 70. La base 4, según la numeración de la secuencia *logo*, varía entre T y A, aunque hay una notoria preferencia por T. Contigua a la caja -35, se encuentra una caja “ATGT” (posiciones 6 a la 9), cuya función en la regulación de los promotores que la contienen no ha sido determinada. Es también notoria la predominancia de Gs de la base 12 a la 16, especialmente en la posición 13 donde una G está claramente conservada. Entre las posiciones 17-20 se encuentra el motivo “ATGT”, el cual de acuerdo a evidencias experimentales es esencial para la activación mediada por GrIA de *ler* y *E1013* (figura 13) (Jiménez *et al.* 2010; Islam *et al.* 2011). Seguida de este motivo se observa la conservación de la secuencia “TTGT” de la posición 21 a la 24. Por último, en la caja -10 del promotor, posiciones 25-30, las únicas bases conservadas son las primeras dos bases, TA. La posición 27 no tiene ninguna letra en la secuencia *logo* porque siempre tiene un cambio, entonces no se puede dar un grado de conservación para ningún nucleótido, mientras que de la 28 y 30 la tendencia es por A/G y A/C, respectivamente.

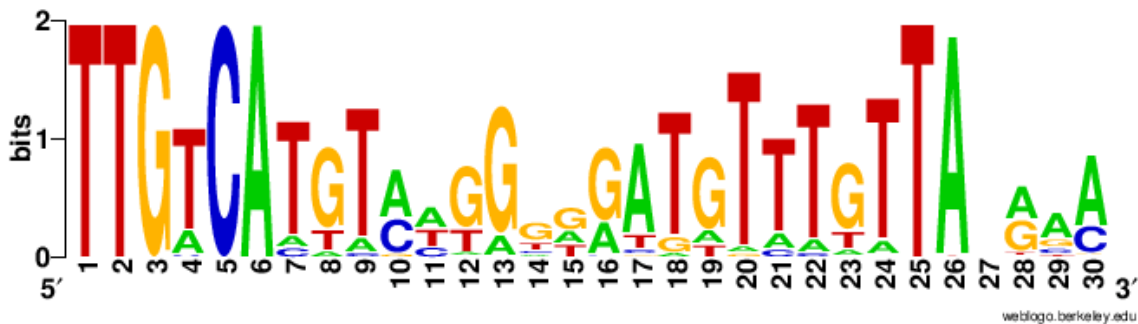


Figura 15. Secuencia *logo* obtenida a partir de las secuencias identificadas en los diferentes genomas de cepas pertenecientes a la familia A/E de genes con posibles promotores regulados por GrIA. Este consenso está presente en la región promotora de más de 15 genes diferentes.

En el consenso de promotores dependientes de  $\sigma^{70}$  no se puede distinguir conservación alguna en la secuencia localizada entre las cajas -10 y -35, ni tampoco en las últimas dos bases de la caja -35 (figura 16) (Shultzaberger *et al.* 2006). Cabe hacer notar que las secuencias representadas en el logo presentado en la figura 15, corresponden a más de 15 genes diferentes encontrados entre los genomas de las

diferentes cepas de bacterias pertenecientes a la familia A/E. Queda por explorar el genoma de otras bacterias en general para determinar si promotores con estas características están presentes en otros organismos que no contienen el LEE.

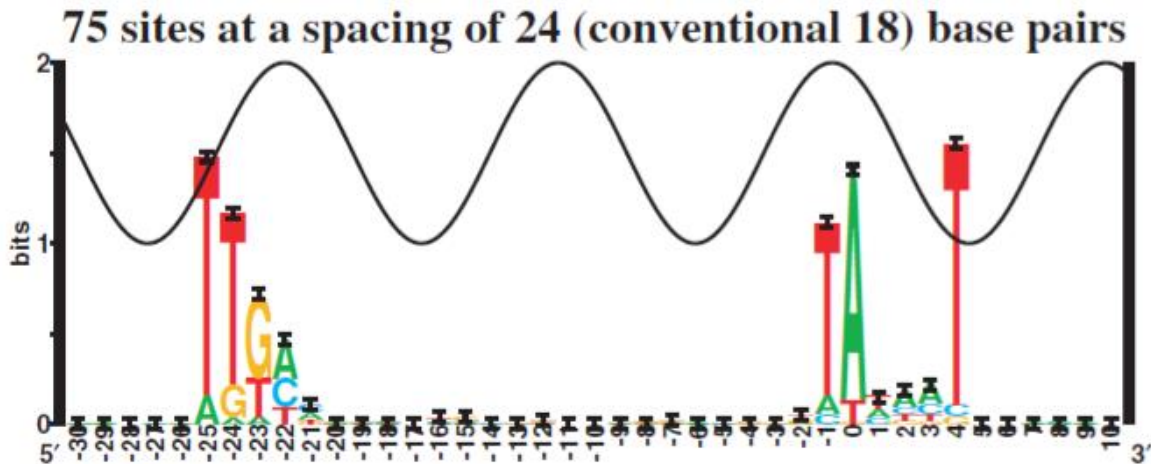


Figura 16. Secuencia logo de los 75 promotores  $\sigma_{70}$  con 18 pb de distancia entre las cajas -10 y -35 en *Escherichia coli* K-12. Tomado y modificado de Shultzaberger et al (2006).

Varios de los promotores identificados por nuestro algoritmo contienen el elemento regulador denominado NRIR, una secuencia repetida invertida o palíndroma que García-Ángulo *et al* (2012) demostraron es necesaria para la activación de los genes *nleH1* y *nleB2* de EPEC. En dicho trabajo se reportó también que las regiones reguladoras de los genes mostrados en la figura 14, así como 9 regiones reguladoras en EHEC EDL933, 8 en EHEC Sakai y 3 en *C. rodentium*, contienen el elemento NRIR. Este elemento también fue localizado, en algunos casos, corriente abajo de algunos genes pero sin estar asociado a un promotor. La posible función del NRIR para estos últimos genes no ha sido establecida.

La tabla 4 muestra las características de los 26 genes que arrojó una nueva búsqueda para el elemento NRIR, que en total representan 58 secuencias dentro de cepas prototipo de patógenos A/E, excepto en *E. coli* MG1655 que fue utilizada

como control. Cabe mencionar que 11 de 26 de estos genes corresponden a parálogos/ortólogos de *nleG*, cuyo número de copias varía dependiendo de la cepa, yendo de un solo gen en EPEC E2348/69 a tres en *C. rodentium* y 14 en EHEC EDL933; no todos los parálogos de *nleG* poseen un secuencia NRIR. Ninguna diferencia se observó en el número de genes recuperados cuando se aumentaron a tres o cuatro los cambios permitidos en cada invertido repetido. Las secuencias identificadas fueron también utilizadas para generar una secuencia *logo* del NRIR (figura 17).

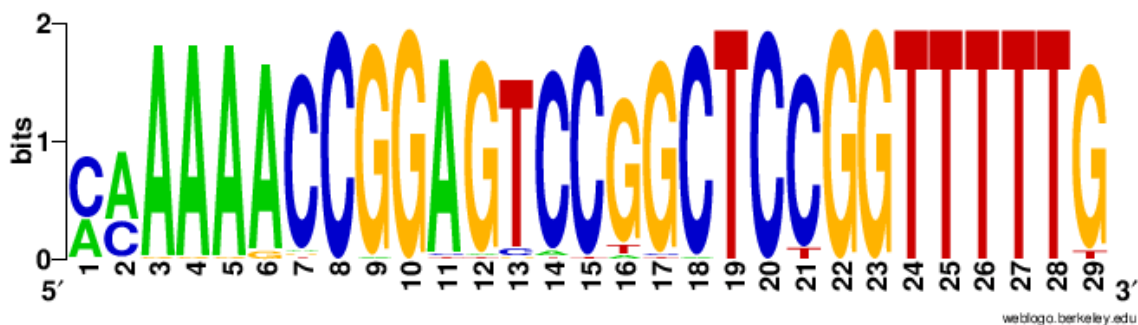


Figura 17. **Secuencia *logo* obtenida a partir de todas las secuencias NRIR con un promotor  $\sigma^{70}$  asociado.** El *logo* se generó a partir de las 58 secuencias recuperadas que representan 26 genes diferentes, algunos de ellos presentes en todas las cepas incluidas en el estudio o en genes sólo presentes en una de las cepas.

Dada la naturaleza de la búsqueda, no resulta sorprendente la conservación del NRIR. Para propósitos del proyecto, sólo se seleccionaron aquellas secuencias que tuvieran anexa una región promotora. Algo a destacar, es que a pesar de los parámetros algo permisibles en la búsqueda del promotor asociado al NRIR, la secuencia *logo* de los promotores identificados (tabla 4) muestra una mayor conservación respecto a la secuencia de la figura 15, resaltando aún más la conservación de la caja -35 la cual conserva una secuencia “TTGTCA” en 56/57 casos, la secuencia “ATGT” adicional anexada a la caja -35 del promotor, el motivo “ATGT” principal en las posiciones 17 a la 20, la conservación de la secuencia “TGTT” entre el motivo “ATGT” y la caja -10, así como la poca conservación en la caja -10. Por otra parte, es notoria la poca conservación en la secuencia de la figura 18 de las posiciones 12 a la 16, aunque no deja de haber una preferencia muy marcada a la presencia de Gs en la misma.

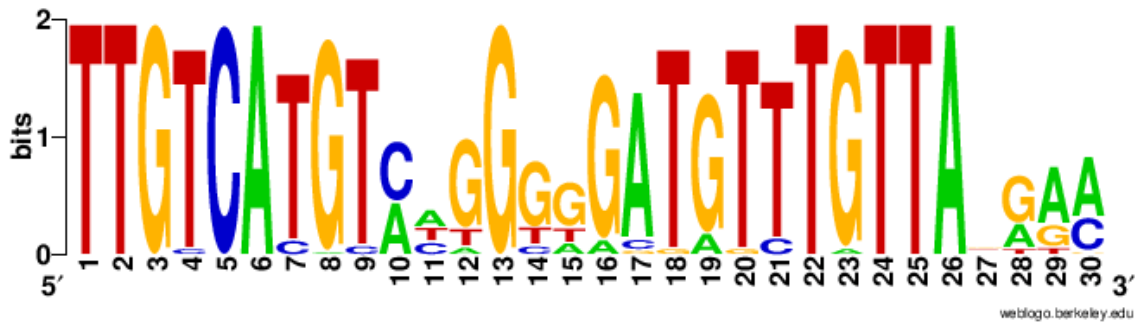


Figura 18. Secuencia *logo* de los promotores asociado al NRIR productos de la búsqueda.

Notablemente, en la tabla 4 hay casos particularmente interesantes. Principalmente, resalta que la mayoría de los genes identificados que poseen un promotor con todas las características que lo hacen candidato a ser regulado por GrlA, coinciden en tener el elemento NRIR corriente arriba. Entre los genes que carecen del NRIR, encontramos a *yodA* de *E. coli* O55:H7, el cual es un ortólogo idéntico de *zinT* de *E. coli* O157:H7 EDL933 y Sakai, aunque en el primer caso el producto del gen está anotado en el GenBank como un precursor de la Epimerasa Ribulosa 3-fosfato y en los otros dos como una proteína con dominio de unión a zinc y cadmio para transporte del periplasma al citosol. Sin embargo, en el KEGG (*Kyoto Encyclopedia of Genes and Genomes*), el motivo conservado anotado para *yodA* es de la familia *zinT*, por ello consideramos que puede ser un error de anotación.

Otra particularidad son los genes *irhA* y *pstS* en *C. rodentium*. Para *irhA* no hay evidencia de algún papel en virulencia, mientras que se ha descrito un papel del regulón de PhoBR para la virulencia de EPEC (Crépin *et al.* 2011). El sistema PST es capaz de regular la actividad del sistema de dos componentes PhoBR al suministrar el transporte específico de  $PO_4^-$  del periplasma al citosol. Una mutante *C. rodentium*  $\Delta pstS$  coloniza deficientemente a su hospedero murino y la mutante de EPEC str. E128012 (atípica) se adhiere menos a células en cultivo. En estas mutantes el regulador de respuesta actúa constitutivamente sobre sus blancos. En consecuencia, para una doble mutante  $\Delta pstS \Delta phoB$  se restaura la adherencia de

EPEC a células en cultivo y la colonización en ratón por *C. rodentium* (Cheng *et al.* 2009).

El cuarto gen con NRIR, identificado en EPEC E2348/69 como *E2348C\_1441* y anotado como un gen que codifica para una proteína hipotética, se encuentra conservado en la cepa *E. coli* O55:H7; sin embargo, en EHEC O157:H7, aunque conserva la secuencia NRIR corriente arriba del promotor, el “ATGT” cercano a la secuencia -10 está sustituido por un trecho de Gs (tabla 4). Por otra parte, su marco de lectura presenta un codón de paro en el codón 23, contrastando con los 180 codones predichos para el marco de lectura del mismo gen en EPEC, lo que en su conjunto sugiere que para el caso de EHEC su función fue descontinuada dando lugar a un pseudogen. Aunque cabe aclarar que ni la regulación, ni la función de este gen ha sido estudiada.

Por último, algo interesante a resaltar es que el promotor del homólogo de *ospB* en *E. coli* O111:H- y *E. coli* O26:H11, y el promotor de *nleB2* de EPEC E2348/69 y O55:H7, incluyendo el NRIR, son idénticos, a pesar de que codifican para proteínas que son estructural y funcionalmente diferentes en su totalidad. En el caso de EPEC, NleB es capaz de interactuar con FADD, TRADD y RIPK1, proteínas receptoras de señales de apoptosis, y de transferirles una molécula de N-acetylglucosamina, inactivándolas, inhibiendo así la apoptosis (Pearson *et al.* 2016). Sin embargo, NleB2 no demostró ser eficiente en la interacción con las proteínas mencionadas anteriormente y tampoco con su actividad enzimática, por lo que su función como efector en EPEC se desconoce (Pearson *et al.* 2013). Por su parte, *ospB* también tiene un ortólogo en *C. rodentium* y, aunque no se encontró una secuencia NRIR en la región reguladora del gen o un promotor  $\sigma^{70}$  con las características de ser regulado por GrlA, se demostró que su producto OspB es sustrato del SST3 (Deng *et al.* 2010). La función de esta proteína no se ha demostrado para ningún patógeno A/E, pero en *Shigella flexneri* OspB promueve la proliferación celular al inducir a un regulador maestro de crecimiento celular, mTORC1, de manera dependiente de

IQGAP1 (Lu *et al.* 2015). Falta demostrar que estos genes son efectivamente regulados por GrlA, así como elucidar el papel regulador del NRIR.

La mayoría de los genes que se listan en la Tabla 4 codifican para efectores o son genes relacionados con virulencia, a excepción de aquellos descritos anteriormente en *C. rodentium*, aunque no se descarta un posible papel de *insA* en la infección.

Hasta el momento, sólo *E1013* y *ler* han sido caracterizados como blancos directos de GrlA (figuras 12 y 13, y datos no publicados) (Jiménez *et al.*, 2010; Islam *et al.*, 2011). Este análisis nos permitió ampliar el repertorio de genes potencialmente regulados por GrlA, así como identificar modelos adicionales que esperamos nos ayuden a entender mejor, además del mecanismo de acción de GrlA, cuáles fueron las presiones selectivas que dieron como resultado la generación de un casete de regulación que integra diversos elementos como el NRIR, el motivo ATGT y un promotor de características particulares que controla la expresión de diversos genes con secuencias codificantes que dan lugar a proteínas con características completamente diferentes.



Tabla 1. Genes encontrados con un promotor posiblemente regulado por GrIA.

| Gen          | Función anotada                                                 | Cepas <sup>a</sup>                                                                              | NRIR <sup>b</sup> | Promotor (-35 /18pb/ -10) <sup>c</sup>            |
|--------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------|---------------------------------------------------|
| <i>E1013</i> | Proteína hipotética                                             | <i>E. coli</i><br>O127:H6,<br>O55:H7,<br>O157:H7,<br>O111:H-,<br>O26:H11,<br><i>E. albertii</i> | Ausente           | TTGACA ATAATGATA <b>ATGT</b> TTATA TAGAAA         |
| <i>ler</i>   | Regulador transcripcional.                                      | <i>E. coli</i><br>O127:H6,<br>O55:H7,<br>O157:H7,<br>O111:H-,<br>O26:H11,<br><i>E. albertii</i> | Ausente           | TTGACA <u>TTT</u> AATGATA <b>ATGT</b> ATTT TACACA |
|              |                                                                 | <i>C. rodentium</i>                                                                             |                   | TTGACA <u>TTT</u> AAGGATA <b>ATAT</b> ATTT TACACA |
| <i>yodA</i>  | Precursor de la Epimerasa Ribulosa-3-Fosfato.                   | <i>E. coli</i> O55:H7                                                                           | Ausente           | TTGTCA <u>TAT</u> GTTACA <b>ATAT</b> AACAT TACACA |
| <i>zinT</i>  | Proteína de unión a Zinc.                                       | <i>E. coli</i><br>O157:H7                                                                       |                   |                                                   |
| <i>irhA</i>  | Regulador transcripcional del operón de la NADH deshidrogenasa. | <i>C. rodentium</i>                                                                             | Ausente           | TTGTCA <u>CGT</u> GATATAC <b>ATAT</b> TTAC TTTCAC |
| <i>pstS</i>  | Transportador ABC de fosfato, proteína de unión al sustrato.    | <i>C. rodentium</i>                                                                             | Ausente           | TTGTCA <u>TAA</u> AACTGTC <b>ATAT</b> TTCG TACATT |

| Gen          | Función anotada           | Cepa <sup>a</sup>                                            | NRIR <sup>b</sup>                | Promotor (-35 /18pb/ -10) <sup>c</sup>             |
|--------------|---------------------------|--------------------------------------------------------------|----------------------------------|----------------------------------------------------|
| <i>nleI</i>  | Efactor sustrato del SST3 | <i>E. coli</i><br>O127:H6                                    | AAAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <u>TGCC</u> CATGGTG <b>ATGT</b> TTGT TATGAC |
| <i>nleH1</i> | Efactor sustrato del SST3 | <i>E. coli</i><br>O127:H6,<br>O55:H7                         |                                  | TTGTCA <b>TGT</b> ACGGGGG <b>ATGT</b> CTGT TAGGAA  |
|              |                           | <i>E. coli</i><br>O157:H7,<br>O111:H-,<br>O26:H11            | CAAAAACCGGAG TCCGG CTCCGGTTTTTG  | TTGTCA <b>TGT</b> ATGGGGG <b>ATGT</b> CTGT TAGGAA  |
| <i>nleH2</i> | Efactor sustrato del SST3 | <i>E. coli</i><br>O127:H6,<br>O55:H7,<br>O111:H-,<br>O26:H11 | ACAAAACCGGAG TCCGG CTCCGGTTTTTG  | TTGTCA <b>TGT</b> CATGGTG <b>ATGT</b> TTGT TAAGAA  |
|              |                           | <i>E. albertii</i>                                           | CAAAAACCGGAG TCCGG CTCCGGTTTTTG  |                                                    |
|              |                           | <i>C. rodentium</i>                                          | CAAAAATCGGAG TCCGG CTCCGGTTTTTG  | TTGTCA <b>TGT</b> CAGGGTG <b>ATAT</b> TTGT TAATAA  |
| <i>E1441</i> | Proteína hipotética       | <i>E. coli</i><br>O127:H6,<br>O55:H7                         | ACAAAACCGGAG TCCGG CTCCGGTTTTTG  | TTGTCA <b>TGT</b> ATAGGGG <b>ATAT</b> TTGT TATTAG  |
|              |                           | <i>E. coli</i><br>O157:H7                                    | CAAAAACCGGAG CCCGG CTCCGGTTTTTG  | TTGTCA <b>TGT</b> ATAGGGGGGGTTAT TAGAGA            |
| <i>ospB</i>  | Efactor sustrato del SST3 | <i>E. coli</i><br>O111:H-,<br>O26:H11                        | CAAAAACCGGAG TCCGG CTCCGGTTTTTG  |                                                    |
| <i>nleB2</i> | Efactor sustrato del SST3 | <i>E. coli</i><br>O127:H6,<br>O55:H7                         |                                  | TTGTCA <b>TGT</b> CCGGTGG <b>ATGT</b> TTGT TAGGAA  |
|              |                           | <i>E. coli</i><br>O157:H7                                    | ACAAAACCGGAG TCCGG CTCCGGTTTTTG  |                                                    |
| <i>nleB</i>  | Pseudogen                 | <i>C. rodentium</i>                                          | CAAAAGCCGGAA TACAC CTCCGGTTTTTG  | TTGTCA <b>TGT</b> CAGGCTG <b>ATGT</b> TTGT TAGGAA  |

| Gen                | Función anotada                                   | Cepa <sup>a</sup>                                            | NRIR <sup>b</sup>               | Promotor (-35 /18pb/ -10) <sup>c</sup>            |
|--------------------|---------------------------------------------------|--------------------------------------------------------------|---------------------------------|---------------------------------------------------|
| <i>espX</i>        | Proteína hipotética                               | <i>E. coli</i><br>O157:H7                                    |                                 |                                                   |
|                    | Efecto sustrato del SST3                          | <i>E. coli</i><br>O111:H-,<br>O26:H11                        | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> CCGGGGA <b>ATAT</b> TTGT TAGATA |
|                    |                                                   | <i>C. rodentium</i>                                          | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGCCA <b>TGT</b> CCGGGGA <b>ATAT</b> TTGT TAGATA |
| <i>insA</i>        | Represor IS1 putativo                             | <i>E. coli</i><br>O111:H-                                    | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> CCGGTGG <b>ATGT</b> TTGT TAGGAA |
| <i>nleG</i>        | Efecto sustrato del SST3                          | <i>E. coli</i> O55:H7                                        | CAAAAACCGGAG GCCGG CTCCGGTTTTTG | TTGTCA <b>TGC</b> CATGGTG <b>ATGT</b> TTGT TATGAA |
| <i>nleG</i>        | Efecto sustrato del SST3                          | <i>E. coli</i> O55:H7                                        | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGC</b> CATGGTG <b>ATGT</b> TTGT TATGAA |
| <i>nleG</i>        | Proteína hipotética,<br>probablemente <i>nleG</i> | <i>E. coli</i><br>O55:H7,<br>O157:H7,<br>O111:H-,<br>O26:H11 | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> CAGGGAG <b>ATGT</b> TTGT TATGAA |
| <i>nleG</i>        | Efecto sustrato del SST3                          | <i>E. coli</i><br>O55:H7,<br>O157:H7,<br>O26:H11             | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> ATGGGGG <b>CTGT</b> TTGT TATGAC |
| <i>nleG</i>        | Proteína hipotética<br>homóloga a <i>nleG</i>     | <i>E. coli</i><br>O55:H7,<br>O157:H7,<br>O111:H-             | CAAAAACCGGAG TCCTG CTCTGGTTTTTG | TTGTCA <b>CGT</b> ATGGGGG <b>ATGT</b> TTGT TACAGC |
|                    |                                                   | <i>E. coli</i><br>O157:H7<br>EDL933                          | CAAAAACCGGAG TCCGG CTCCGGTTTTTG |                                                   |
|                    |                                                   | <i>E. coli</i><br>O26:H11                                    | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> ATGGGGG <b>ATGT</b> TTGT TACAGC |
|                    |                                                   | <i>C. rodentium</i>                                          | CAAAAGCCGGCG TTCGG ATCCGGTTTTTG | TTGTCA <b>TGT</b> CAGGGGG <b>ATGT</b> TTGT TATTGC |
| <i>nleG6</i><br>-3 | Pseudogen                                         | <i>E. coli</i><br>O157:H7                                    | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> ACGGTGG <b>ATGT</b> TTGT TATGAC |
| <i>nleG</i>        | Pseudogen                                         | <i>E. coli</i><br>O111:H-                                    | CAAAAACCGGAG TCCGG CTCCGGTTTTTG | TTGTCA <b>TGT</b> CCGGTGG <b>ATGT</b> TTGT TATGAC |

| Gen                | Función anotada          | Cepa <sup>a</sup>                                 | NRIR <sup>b</sup>                  | Promotor (-35 /18pb/ -10) <sup>c</sup> |
|--------------------|--------------------------|---------------------------------------------------|------------------------------------|----------------------------------------|
| <i>nleG2</i><br>-1 | Pseudogen                | <i>E. coli</i><br>O157:H7,<br>O111:H-,<br>O26:H11 | ACAAAACCGGAG TCCGG CTCCGGTTTTTG    | TTGTCA TGTAAAGGCAGATGTTTGT TAAAGC      |
| <i>nleG</i>        | Pseudogen                | <i>E. coli</i><br>O111:H-                         | CAAAAAACCGGAG TCCGG CTCCGGTTTTTGTG | TTGCCA TATAAGGCAGATGTTTGT TACAGC       |
| <i>nleG</i>        | Efecto sustrato del SST3 | <i>E. coli</i><br>O26:H11                         | AAAAAACCGGAG TCTGG CTCCGGTTTTTTG   | TTGTCA TGT CATGGT GATGTTTGT TAATGA     |
| <i>nleG1</i>       | Efecto sustrato del SST3 | <i>C. rodentium</i>                               | TGGGGGGCAGTT CACAT CTCCGGTTTTTG    | TTGTCA TGTCAGGTTGATGTTTGT TAAGAA       |

<sup>a</sup> Cepas en las que se encontró la secuencia indicada. <sup>b</sup> Secuencia NRIR (nle regulatory inverted repeat) en el caso de los genes que lo contienen corriente arriba del promotor. <sup>c</sup> En rojo se resalta la secuencia "AT[G/A]T" que se ha confirmado experimentalmente para la activación mediada por GrlA en el caso de *ler* y *E1013* y en azul un ATGT adicional contiguo a la caja -35 del promotor, encontrado en los genes que contienen el NRIR cuya función no ha sido analizada. Las secuencias subrayadas en ambos casos corresponden a motivos degenerados. Se incluyen las variantes a la secuencia encontrada para un mismo gen en diferentes cepas.

**La región correspondiente a la región líder no codificante de *E1013* ejerce un efecto represor sobre su expresión.**

Similar a la región reguladora de *ler*, el transcrito de *E1013* incluye una región líder no codificante de 119 pb (figura 11). En el caso de *ler*, esta región está involucrada de manera importante en la regulación negativa del gen (Bustamante et al. 2011). Esto nos llevó a preguntarnos si esta región estaba implicada en la regulación de *E1013*. Para abordar la pregunta se comparó la actividad de fusiones que contienen diferentes trechos de la región 5' reguladora corriente arriba del promotor y hasta la posición +140 del gen, incluyendo los primeros 7 codones del marco de lectura, o hasta la posición +26 lo cual elimina la mayor parte de la región líder no codificante (figura 11).

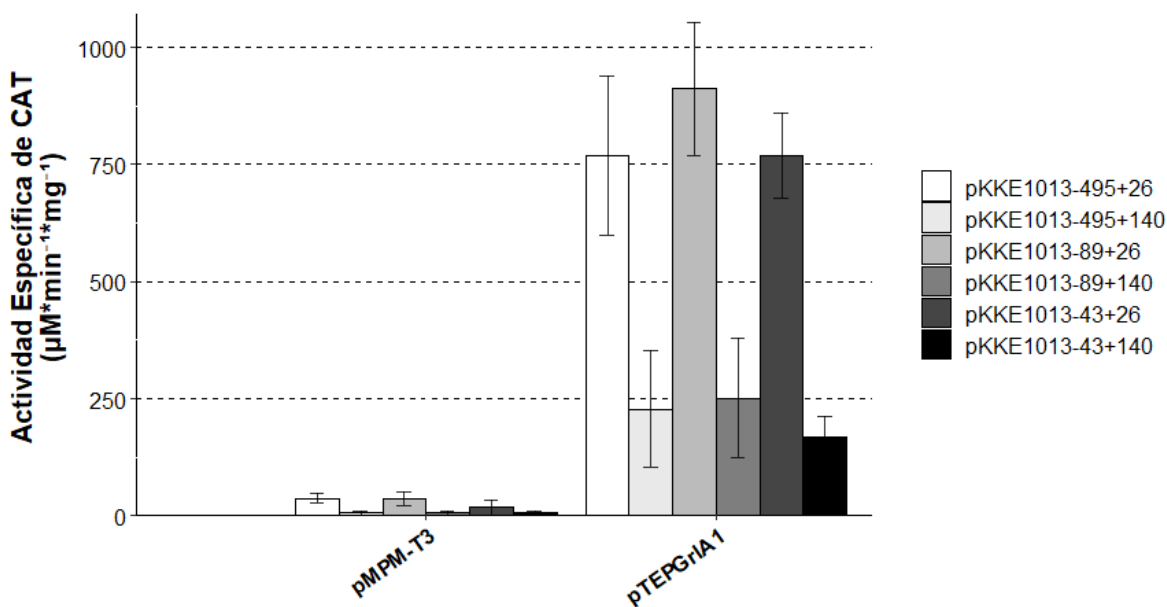


Figura 19. **La región no codificante localizada entre el inicio de transcripción y el codón de inicio de *E1013* regula negativamente su expresión.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de *E. coli* MC4100 transformada con los plásmidos pMPM-T3 o pTEPGriA1 más las fusiones indicadas. Los cultivos fueron crecidos en 5 ml de LB en tubo por 9 horas ( $DO_{600} \sim 0.93 \pm 0.073$ ) a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar.

En la figura 19 se muestra que la actividad en presencia de GrlA, de todas las fusiones en las que se elimina la mayor parte de la región líder no codificante (pKKE1013-495+26, pKKE1013-89+26 y pKKE1013-43+26) es mayor comparada con la de aquellas en las que esta región se mantuvo intacta (pKKE1013-495+140, pKKE1013-89+140 y pKKE1013-43+139). Esto sugiere que la región líder de *E1013* tiene un efecto negativo en la expresión de este gen. Este resultado sugiere que la regulación de *E1013* es similar a la de *ler*, ya que se ha demostrado que este gen es regulado negativamente a nivel transcripcional por la unión de H-NS a esta región no codificante (Bustamante et al. 2011).

### **La expresión de *E1013* es independiente de la presencia de H-NS**

Se sabe que H-NS se une al DNA preferentemente a secuencias ricas en A-T. En su mayoría los genes adquiridos por transferencia horizontal, como es el caso del *LEE* o de los profagos, tienen un contenido de A+T mayor al promedio del genoma (Winardhi et al. 2014), lo cual favorece que su regulación sea controlada por H-NS (Bustamante et al. 2001).

Jimenez et al (2010) reportaron que GrlA, además de actuar como activador, también antagoniza la función represora de H-NS sobre el promotor de *ler*. Por su parte, el promotor de *E1013*, además de ser regulado positivamente por GrlA como el promotor de *ler* (figuras 12 y 13), cuenta también con un sitio de regulación negativa en su región líder no codificante (figura 19), lo que llevó a hipotetizar que al igual que *ler*, H-NS podría estar regulando negativamente a *E1013*. Para evaluar esta hipótesis se analizó la actividad de las fusiones pKKE1013-89+26, pKKE1013-89+140 y pler-260 en *E. coli* MC4100 y una mutante  $\Delta hns$  (figura 20).

En la cepa silvestre se observa, que las fusiones pler-260 y pKKE1013-89+140 se expresan sólo en presencia del plásmido pTEPGrlA1, mientras que la fusión que contiene la región líder *E1013*-89+140 no se expresa. En la mutante  $\Delta hns$  vemos

que la fusión pler-260 es activa independiente de GrIA, y que la presencia de GrIA estimula adicionalmente su expresión como se reportó anteriormente (Jimenez *et al.*, 2010). Por su parte, la expresión de las fusiones pKKE1013-89+26 y pKKE1013-89+140 no se afecta por la ausencia de *hns*.

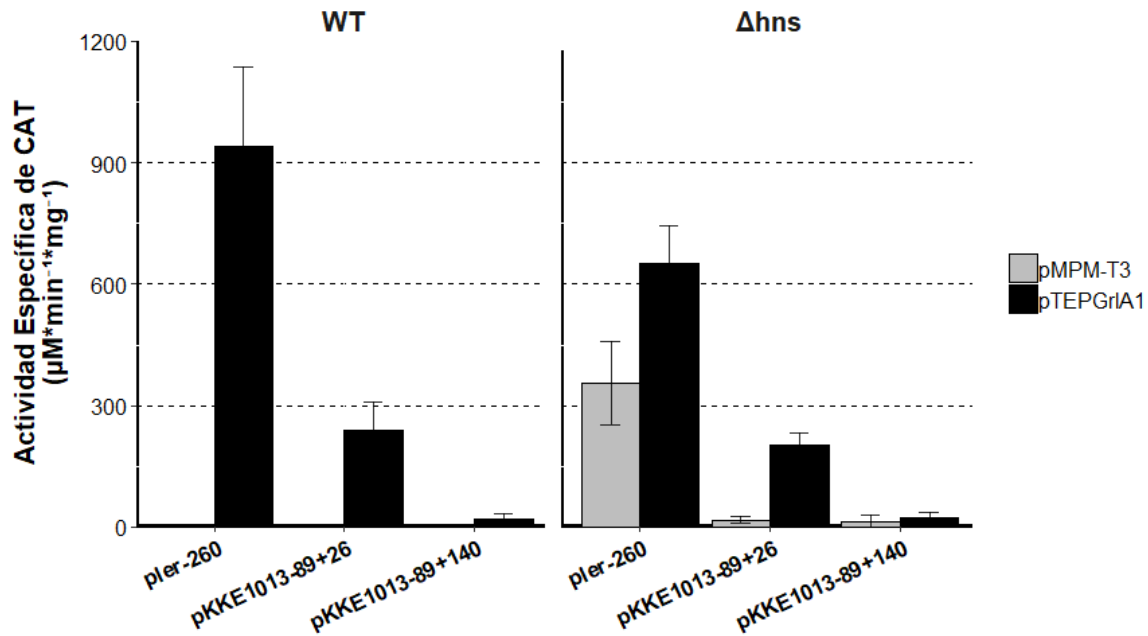


Figura 20. **H-NS no reprime a H-NS.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de *E. coli* MC4100 y la mutante en *hns* transformada con los plásmidos pMPM-T3 o pTEPGriA1 más las fusiones indicadas. Los cultivos fueron crecidos en matraz a una DO<sub>600</sub> de 0.8 a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar.

En su conjunto estos resultados indican que, a pesar de su alta homología con el promotor de *ler*, el promotor de *E1013* no es regulado negativamente por H-NS y sugieren que la activación mediada por GrIA podría estar asociada con el reclutamiento de la RNA polimerasa y/o la formación de un complejo abierto de transcripción activo, o antagonizando un represor diferente a H-NS. Queda por determinar el mecanismo que silencia al promotor aun en presencia de GrIA cuando la región líder de *E1013* está presente y cuáles son las implicaciones en el contexto cromosomal donde el gen parece ser activo de acuerdo a los resultados de RNAseq

(figura 7).

### **PerC activa modestamente al promotor de *E1013***

Además de GrlA, en EPEC la expresión de *ler* es también activada por PerC, una proteína codificada en el operón *per* del plásmido EAF que no tiene dominios de unión a DNA conocidos (Bustamante *et al.*, 2011). Sin embargo, su mecanismo de acción o los elementos de la región reguladora requeridos para su función no han sido definidos a la fecha. Dada la similitud entre los promotores de *ler* y *E1013*, nos preguntamos si aquellos requeridos por PerC estarían también conservados. Analizamos la actividad de las fusiones pKKE1013-495+26, pKKE1013-495+140, pKKE1013-43+26 y pKKE1013-43+140 en *E. coli* MC4100 en presencia del el vector vacío pMPM-T3 y su derivado pTEPPerC1 a partir del cual se expresa PerC. La fusión *pler*-260 fue utilizada como control positivo de la activación mediada por PerC (figura 21).

A diferencia del promotor de *ler*, cuya expresión se incrementa hasta 400 veces en presencia de PerC, la expresión del promotor de *E1013* muestra un incremento modesto de cerca de cuatro veces en presencia de este regulador y dando actividades cerca del límite de detección confiable del ensayo. Aunque preliminarmente se podría concluir que PerC no es un regulador relevante de *E1013* en las condiciones ensayadas, su papel debe ser analizado en más detalle, ya que las actividades alcanzadas para este promotor son relativamente bajas en comparación con la de otros promotores que son regulados por Ler o GrlA. Cabe contrastar que los niveles de actividad que alcanzan las fusiones de *ler* y *E1013* en presencia de GrlA difieren en no más de tres veces (figura 20), mientras que con PerC la diferencia es muy evidente (figura 21).



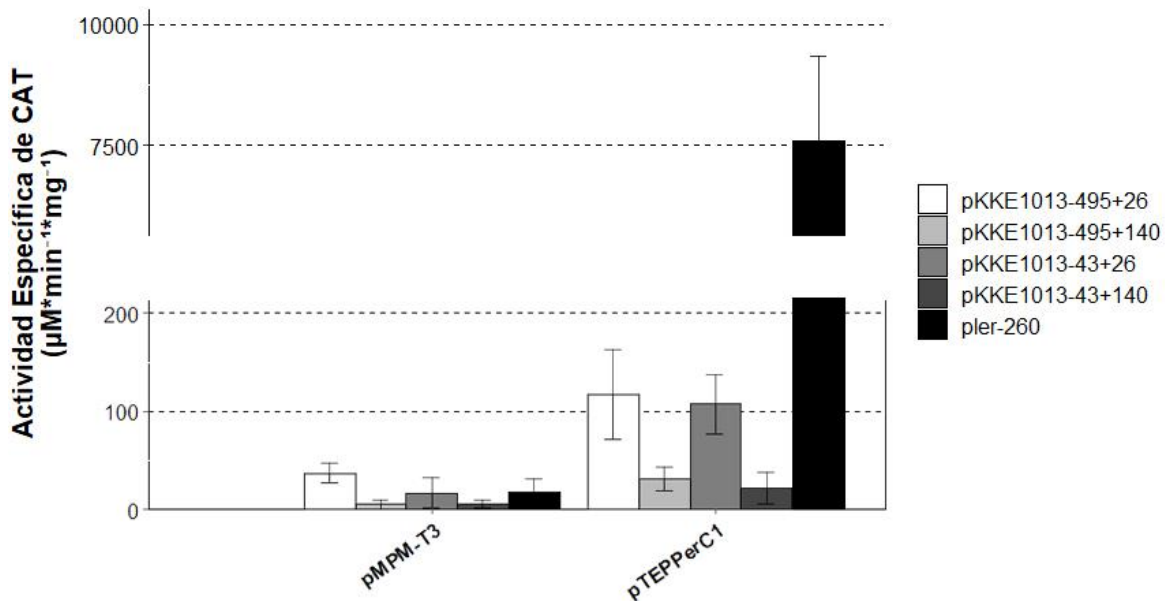


Figura 21. **PerC no tiene un efecto sobre el promotor de *E1013***. Ensayo de actividad específica de CAT a partir de muestras de cultivos de *E. coli* MC4100 transformada con los plásmidos pMPM-T3 o pTEPPerC1 más las fusiones indicadas. Los cultivos fueron crecidos en 5 ml de LB en tubo por 9 horas ( $DO_{600} \sim 0.93 \pm 0.073$ ) a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar.

*E. coli* enterohemorrágica (EHEC) posee en su genoma homólogos de *perC*, denominados *pch* (*perC*-homologue). Abe *et al* (2008), realizaron un *ChIP-on-chip* con PchA que generó la única evidencia experimental a la fecha de la posible capacidad de esta proteína de unirse al DNA. Posteriormente, Fukui *et al* (2016) observó que los sitios de unión de PchA coincidían, en su mayoría, con sitios de unión de H-NS y su parálogo StpA, proponiendo que Pch actúa como un antirepresor que antagoniza la represión mediada por proteínas de estructuración del nucleóide como H-NS, StpA, Hha y YdgT sobre el *LEE*.

Por su parte, se ha reportado que PerC no muestra afinidad por el promotor de *ler* *in vitro* (Porter *et al.*, 2005), sugiriendo que su papel como activador de *ler* en EPEC podría involucrar componentes adicionales. Más aún, tanto PerC como los parálogos de Pch son capaces de activar los promotores de *ler* de EPEC y EHEC en una cepa de *E. coli* K12 (Porter *et al.*, 2005). Esto no quiere decir que

necesariamente se conserve el mismo mecanismo en ambos patotipos, ya que las condiciones en las que actúan son diferentes. En EPEC, PerC es el regulador principal del *LEE* cuando la bacteria es cultivada en DMEM a 37°C sin agitación en una atmósfera de 5% de CO<sub>2</sub>, pero no así en cultivos agitados sin CO<sub>2</sub> donde no hay ninguna afectación en la expresión del *LEE* en la  $\Delta$ *perC* (Bustamante *et al.* 2011). Por su parte, en EHEC los genes *pchA* y *pchB* son esenciales para la expresión del *LEE* (Iyoda y Watanabe. 2004).

A pesar de haber diferencias entre la regulación del *LEE* de EHEC y EPEC, se ha propuesto que PerC antagoniza la represión mediada por H-NS en EPEC (Bustamante *et al.* 2011), encajando en el modelo propuesto por Fukui *et al.* (2016). En contraste, no observamos efecto de H-NS sobre la expresión de *E1013* (figura 14), en línea con la ausencia de regulación por parte de PerC, características que distinguen la regulación entre *ler* y *E1013*.

### ***E1013* se expresa en bajos niveles en DMEM**

Después de demostrar que solamente se necesita de la región -43 a +26 de *E1013* para la activación de su promotor por GrlA, nos preguntamos cuál es su región mínima reguladora. Para responder esta pregunta, evaluamos la actividad de los distintos recortes de la región reguladora de *E1013* en EPEC E2348/69 en cultivos de LB y DMEM (figura 22).

Ninguno de los recortes generó actividad significativa en ambos medios, indicando que al menos en estas condiciones *E1013* es prácticamente inactivo. Asimismo, al no observar diferencia de actividad en los distintos recortes, se decidió usar las fusiones pKKE1013-43+26 y pKKE1013-43+140 como referentes para futuros ensayos de actividad.

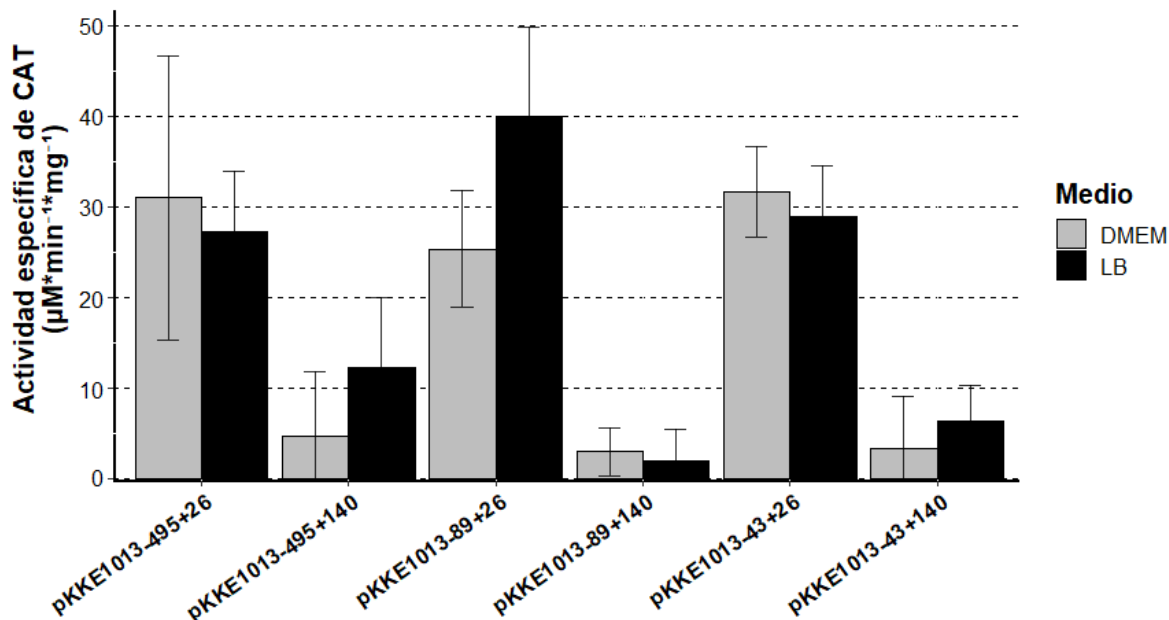


Figura 227. ***E1013* se expresa en niveles muy bajos en DMEM.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de EPEC E2348/69 silvestre transformada con las fusiones indicadas. Los cultivos fueron crecidos en 5 ml de DMEM o LB en tubo por 9 horas ( $DO_{600} 0.85 \pm 0.37$ ) a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar.

### ***E1013* es activo en medio mínimo N con glicerol**

En el transcriptoma realizado por Hazen *et al* (2015), se aprecia muy poca transcripción de *E1013* tanto en LB como en DMEM, sin superar las 10 y 100 lecturas, respectivamente, aunque respalda el diferencial de transcripción que observamos entre los dos medios (figura 7). El DMEM y el LB han sido ampliamente utilizados para comparar la expresión del *LEE* en condiciones de crecimiento que la favorecen o no, respectivamente. Se ha demostrado que el regulador positivo de *ler* en cultivos de DMEM agitados es GrIA y que éste también es más abundante en esta condición en comparación con el LB (Bustamante *et al.* 2011). Sin embargo, a pesar de que GrIA expresado a partir de un plásmido es capaz de inducir la expresión del promotor *E1013* en un fondo heterólogo como *E. coli* K12 (figura 12), éste no parece ser suficiente para activarlo en EPEC bajo estas condiciones de

crecimiento (figura 22). Esto sugiere, entre otras posibilidades, que el promotor de *E1013* es regulado negativamente por otro u otros reguladores que impiden la función de GrlA, y que su expresión se favorece en condiciones distintas a las probadas a la fecha.

Previamente se reportó que EPEC secreta proteínas a través del SST3 en cultivos crecidos en medio mínimo 9 (MM-9) (Deng *et al.* 2005), el cual se clasifica como un medio mínimo químicamente definido que contiene cloruro de amonio como fuente de nitrógeno y glucosa como fuente de carbono (anexo 4). Considerando que este medio también favorece la secreción y por tanto la expresión del LEE, decidimos analizar la expresión de *E1013* en medios mínimos como una alternativa para encontrar una condición que favoreciera la activación de su promotor en fondo silvestre, ya que éstos pueden ser modificados fácilmente en su composición.

La expresión de *E1013* se determinó en la cepa E2348/69 silvestre conteniendo las fusiones pKKE1013-43+26 y pKKE1013-43+140 (tabla 2) a partir de cultivos en LB y DMEM que se usaron como medios control, y en medio mínimo N (MM-N) con glicerol al 0.5 % (MM-N·Glicerol) o con glucosa al 0.48 % (MM-N·Glucosa) como fuente de carbono. La fusión pSEPZ-11, que contiene la región reguladora del operón *LEE2* y es muy sensible a cambios en la expresión del *LEE*, se usó como control (Bustamante *et al.* 2001).

La fusión pKKE1013-43+26 tuvo una actividad de CAT en DMEM de  $150 \text{ U} \cdot \text{mg}^{-1} \pm 60$ , mientras que la fusión pKKE1013-43+140 fue de  $43 \text{ U} \cdot \text{mg}^{-1} \pm 24$ . Esto indica que, aunque los niveles de expresión de la fusión corta (pKKE1013-43+26) sean muy bajos, se alcanza a expresar en DMEM en niveles previamente reportados por Cadena, E. (2016) y se puede apreciar un decremento de actividad de aproximadamente de tres veces cuando la fusión tiene la región líder (figura 23), confirmando el efecto negativo que ejerce sobre la expresión de *E1013* (figura 19). Estos datos sugieren que existe un regulador negativo conservado tanto en EPEC como en *E. coli* K-12 que requiere la secuencia líder para efectuar su función, o que

dicha región adopta una conformación que desfavorece la expresión a nivel transcripcional o postranscripcional que aún no se ha explorado.

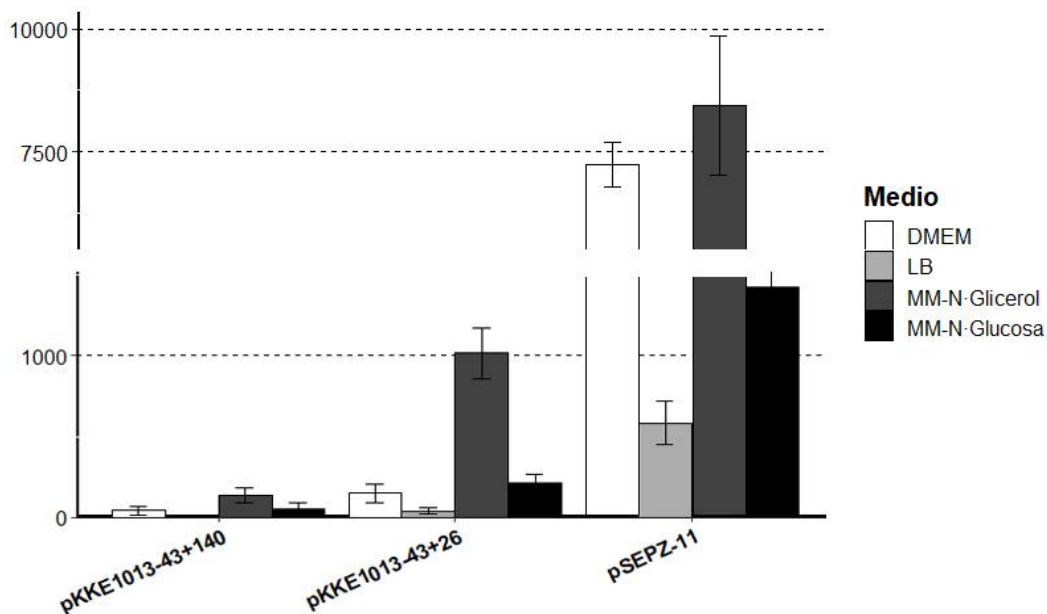


Figura 238. **La fusión pKKE1013-43+26 es activa en MM-N-Glicerol.** Ensayo de actividad específica de CAT a partir de muestras de cultivos de EPEC transformados con las fusiones indicadas. Los cultivos fueron crecidos en 5 ml del medio indicado en tubo por 9 horas a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar.

De manera interesante, en MM-N la fusión pKKE1013-43+26 mostró niveles significativamente más altos de actividad cuando el medio contiene glicerol (MM-N-Glicerol) como fuente de carbono; mientras que con glucosa tiene una actividad muy similar a la reportada en DMEM (figura 23). Por su parte, la expresión dirigida por la fusión pKKE1013-43+140 sigue siendo regulada negativamente debido a la presencia de la región líder, con una actividad 7 veces menor respecto a la de pKKE1013-43+26.

Por otra parte, de los resultados mostrados en la figura 23, resalta que la actividad de la fusión pSEPZ-11 en MM-N-Glicerol sea similar a la del DMEM pues uno de los componentes básicos del MM-N es sulfato de amonio (7.416 mM), al que

previamente se le asoció un efecto represor sobre la expresión de los operones *bfp*, *per* y *LEE2* al adicionarlo al DMEM, el cual no contiene sales de amonio en su composición sino aminoácidos como fuente de nitrógeno (anexo 3) (Puente *et al.*, 1996; Martínez-Laguna *et al.*, 1999; Bustamante *et al.*, 2001). Más aún, Deng *et al* (2005) observaron secreción de proteínas dependiente del SST3 a partir de cultivos en MM-9, otro medio mínimo que también contiene amonio (18.7 mM) y glucosa como fuente de carbono. Estas observaciones plantean la posibilidad de que el efecto represor del sulfato de amonio en DMEM, sea debido a la combinación con la composición particular de este medio y no a la presencia de sales de amonio en sí.

También es de notar la reducción en actividad de la fusión pSEPZ-11 en MM-N·Glucosa con respecto al mismo medio con glicerol, lo cual sugiere que, bajo estas condiciones de crecimiento, la fuente de carbono juega un papel importante para la regulación del *LEE*. Hasta donde sabemos no hay ningún reporte que describa actividad del *LEE* en medios con glicerol como única fuente de carbono.

### **La expresión del *LEE* y *E1013* se estimula en medios mínimos con glicerol como fuente de carbono**

El que la activación de *E1013* y del *LEE* se favorece en medio mínimo con glicerol como fuente de carbono se volvió de interés en el estudio, ya que hasta el momento es la única condición en donde se observa una activación significativa de la fusión pKKE1013-43+26 con respecto al LB y el DMEM. Para confirmar que la fuente de carbono es el factor determinante y no la fuente de nitrógeno o sales presentes en el MM-N, se diseñó un experimento en donde se evaluó la expresión de las mismas fusiones pSEPZ-11 (figura 24), pKKE1013-43+26 (figura 25) y pKKE1013-43+140 (figura 26) en EPEC silvestre, a partir de cultivos en LB, DMEM, MM-N y MM-9 con las variantes descritas a continuación. LB, DMEM, DMEM más 30mM de amonio (DMEM·NH<sub>4</sub>(+)), MM-N con glucosa y sin amonio (MM-N·Glucosa·NH<sub>4</sub>(-)), MM-N

con glucosa y con amonio 14.832 mM (concentración original de la composición del medio; MM-N·Glucosa), MM-N con glucosa y con amonio 30mM (MM-N·Glucosa·NH<sub>4</sub>(+)), MM-N con glicerol y sin amonio (MM-N·Glicerol·NH<sub>4</sub>(-)), MM-N con glicerol y amonio 14.832 mM (MM-N·Glicerol), MM-N con glicerol y amonio 30 mM (MM-N·Glicerol·NH<sub>4</sub>(+)), MM-9, MM-9 con ácidos casamino sustituyendo al amonio como fuente de nitrógeno (MM-9·Aa·NH<sub>4</sub>(-)) y MM-9 con glicerol sustituyendo a la glucosa como fuente de carbono (MM-9·Glicerol·Glucosa(-)).

En la figura 24 observamos que la expresión de la fusión pSEPZ-11 se reprime y se induce en medio LB y DMEM, respectivamente. A su vez, la expresión de la fusión disminuye en DMEM·NH<sub>4</sub>(+), consistente con lo reportado por Bustamante *et al* (2001). De manera interesante, el promotor del *LEE2* contenido en esta fusión muestra niveles significativos de actividad en ambos medios mínimos con glicerol. Esto sugiere que el glicerol actúa como inductor o que la glucosa ejerce represión catabólica o un fenómeno similar que abate la expresión de este promotor.

Por su parte, como se mostró anteriormente, la fusión pKKE1013-43+26 se expresa a niveles basales en LB y sólo muestra un ligero incremento en actividad en DMEM (figura 25). Sin embargo, al igual que la fusión pSEPZ-11 (figura 24), el promotor de *E1013* se expresa al menos 5 veces más cuando la fuente de carbono es glicerol en ambos medios mínimos.

De acuerdo a su participación como regulador negativo de la expresión de *E1013*, la presencia de la región líder en la fusión pKKE1013-43+140 no permite que se exprese a niveles por arriba del basal, aún en presencia de glicerol (figura 26), confirmando el fuerte efecto represor que ésta región ejerce sobre la expresión de *E1013*.

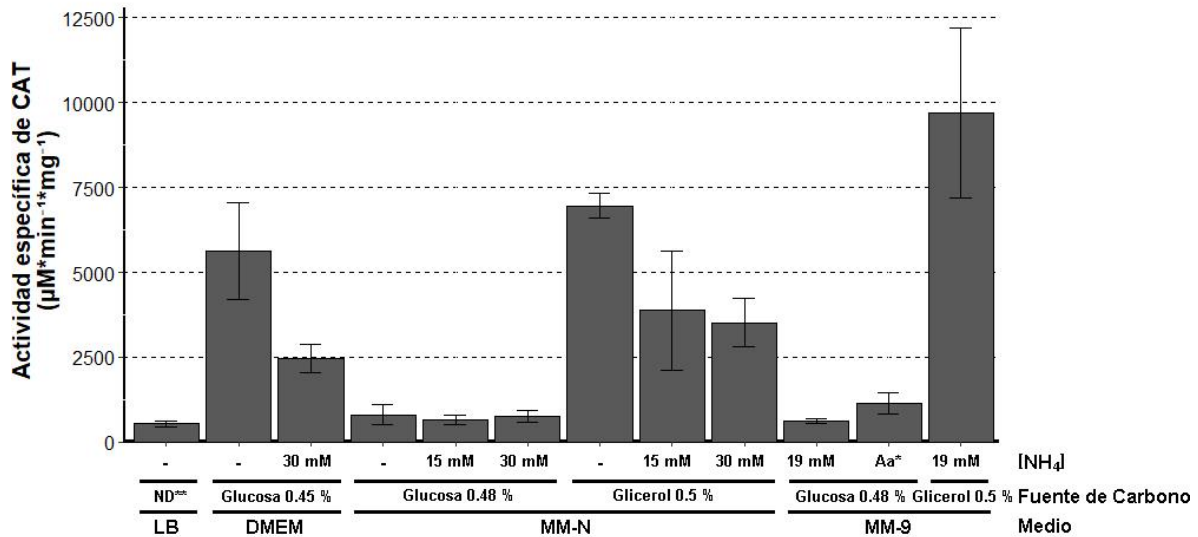


Figura 24. **La fuente de carbono y no de nitrógeno regulan la expresión de la fusión pSEPZ-11 en medios mínimos.** Ensayo de actividad específica de CAT en la cepa EPEC E2348/69 silvestre transformada con la fusión pSEPZ-11 en cultivos crecidos en 5 ml del medio indicado en tubo por 8 horas a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar. \* Cambió de sales de amonio por ácidos casamino como fuente de nitrógeno. \*\* No determinado.

En su conjunto, estos resultados plantean la posibilidad de que los genes del LEE y *E1013* estén sujetos a represión catabólica, y que reguladores transcripcionales como CRP (Escalante *et al.* 2012) y Cra (Shimada *et al.* 2010; Chabarría *et al.* 2014, Bley-Folly *et al.* 2018) podrían estar involucrados en dicha regulación a través de los mecanismos descritos en la figura 27. Lo anterior, nos abre la posibilidad de estudiar un aspecto de la regulación del LEE de EPEC poco explorado, esto es, la relación que podría haber entre la expresión de los factores de virulencia de esta bacteria y su estado metabólico, en particular en torno a la disponibilidad de fuentes de carbono.



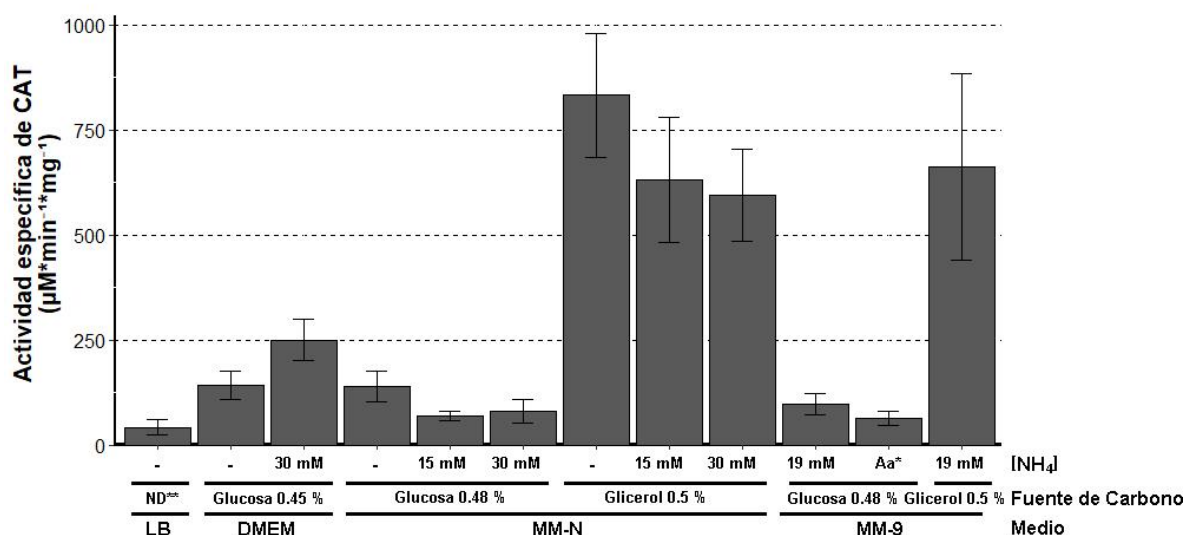


Figura 25. **La fuente de carbono y no de nitrógeno regulan la expresión de la fusión pKKE1013-43+26 en medios mínimos.** Ensayo de actividad específica de CAT con la cepa EPEC E2348/69 silvestre transformada con la fusión pKKE1013-43+26 en cultivos crecidos en el medio indicado en tubo por 8 horas a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar. \* Cambió de sales de amonio por ácidos casamino como fuente de nitrógeno. \*\* No determinado.

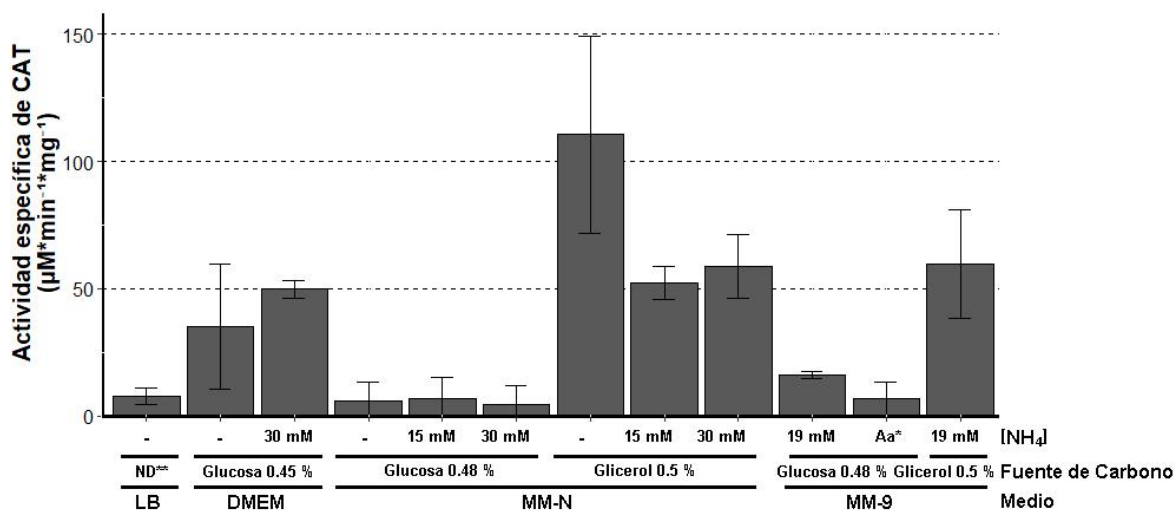


Figura 26. **La región líder no codificante regula negativamente la expresión de E1013 aún en condiciones donde la fusión pKKE1013-43+140 es activa.** Ensayo de actividad específica de CAT con la cepa EPEC transformada con la fusión pKKE1013-43+26 en cultivos crecidos en el medio indicado en tubo por 8 horas a 37°C en agitación. La gráfica representa el promedio de al menos tres experimentos independientes con su réplica biológica. Las barras de error representan la desviación estándar. \* Cambió de sales de amonio por ácidos casamino como fuente de nitrógeno. \*\* No determinado.



## CONCLUSIONES

La región entre las posiciones -43 a +26 de *E1013*, con respecto a su sitio de inicio de transcripción, contiene todos los elementos necesarios para la activación de su promotor por GrlA, incluyendo la secuencia ATGT ubicada en la secuencia espaciadora entre las cajas -10 y -35 del promotor (figura 12). La funcionalidad de esta secuencia se demostró mediante la sustitución de la T en la cuarta posición por una G, la cual disminuyó dramáticamente la expresión del promotor de *E1013* (figura 13). Este resultado respalda el modelo propuesto por Islam *et al* (2011).

*E1013* y *ler* comparten un promotor altamente conservado en patógenos A/E (figura 15). De manera interesante, la mayoría de los genes que presentan este promotor también conservan la secuencia NRIR corriente arriba del mismo (tabla 4).

Los promotores de *E1013* y de *ler* comparten las características de un promotor semi-conservado que es activado por GrlA (figura 14). Así mismo, ambos poseen una región líder no conservada que regula negativamente la expresión de sus respectivos promotores (figura 19). Para el caso de *ler*, la regulación negativa mediada por la región líder involucra a H-NS, no siendo así para el promotor de *E1013* para el cual la ausencia de H-NS no hace ninguna diferencia con respecto a la cepa parental (figura 20). La región líder que regula negativamente la expresión de *E1013*, está contenida entre las posiciones +26 a +140 con respecto al inicio de transcripción (figura 22). A su vez, la regulación del promotor de *E1013* también se distingue de la de *ler* en que PerC no participa en su activación (figura 21).

*E1013* y el *LEE* se expresan en medios mínimos (MM-N o MM-9) con glicerol como fuente de carbono, compartiendo una represión provocada por el amonio añadido

al medio (figura 25 y 26); efecto compartido en DMEM con la fusión pSEPZ-11, no para pKKE1013-43+26 o pKKE1013-43+140. El mecanismo responsable de esta represión se desconoce; sin embargo, este efecto diferencial del amonio dependiente del medio en la expresión del *LEE* o *E1013*, resalta la importancia de considerar el impacto que la composición del medio tiene en la fisiología de EPEC.

## PERSPECTIVAS

Generar fusiones transcripcionales con mutaciones individuales en todas las bases de la secuencia ATGT del promotor de *E1013* para determinar su importancia en la activación mediada por GrlA, así como en la interacción de GrlA con el DNA por ensayos tipo EMSA.

Generar una fusión transcripcional cuyo extremo 3' quede entre la región +26 y +140 de *E1013* para evaluar su actividad y empezar a vislumbrar el mecanismo de regulación negativa ejercida sobre esta región. Asimismo realizar ensayos de RT-PCR de punto final para evaluar si la regulación es a nivel transcripcional o a nivel post-transcripcional, y hasta qué punto se transcribe *E1013*.

Evaluar la actividad de *E1013* y del *LEE* en MM-N·Glicerol en cepas mutantes en los genes que codifican para los reguladores CRP y Cra para evaluar su posible participación en la regulación del *LEE* y *E1013*. Alternativamente, sobre expresar dichos reguladores a partir de plásmidos que contengan los respectivos genes clonados y determinar el efecto sobre la expresión de las fusiones transcripcionales.

Evaluar si los genes contenidos en la tabla 4 son efectivamente regulados por GrlA, así como determinar qué diferencias le otorga la secuencia NRIR a estos promotores. A su vez, determinar si éstos se expresan en condiciones de virulencia y si ayudan a la misma.

## BIBLIOGRAFÍA

Abe, H., Miyahara, A., Oshima, T., Tashiro, K., Ogura, Y., Kuhara, S., Ogasawara, N., Hayashi, T., Tobe, T. 2008. Global regulation by horizontally transferred regulators establishes the pathogenicity of *Escherichia coli*. *DNA Research*. 15: 25-38.

Barba, J., Bustamante, V.H., Flores-Valdez, M.A., Deng, W., Finlay, B.B. y Puente, J.L. 2005. A positive regulatory loop controls expression of the locus of enterocyte effacement-encoded regulators Ler and GrlA. *Journal of Bacteriology*. 187(23): 7918-7930.

Berdichevsky, T., Friedberg, D., Nadler, C., Rokney, A., Oppenheim, A. y Rosenshine, I. 2005. Ler is a negative autoregulator of the *LEE1* operon in Enteropathogenic *Escherichia coli*. *Journal of Bacteriology*. 187(1): 349-357.

Bley-Folly, B., Ortega, A.D., Hubmann, G., Bonsing-Vedelaar, S., Wijma, H.J., Van der Meulen, P., Millas-Argeitis, A. y Helnemann, M. 2018. Assessment of the interaction between the flux-signaling metabolite fructose-1, 6-biphosphate and the bacterial transcription factors CggR and Cra. *Molecular Microbiology*. 109(3): 278-290.

Brosius, J. 1984. Plasmid vectors for the selection of promoters. *Gene*. 27: 151-160.

Bustamante, V.H., Santana, F.J., Calva, E. y Puente, J.L. 2001. Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Molecular Microbiology*. 39(3): 664-678.

Bustamante, V.H., Villalba, M.I., García-Angulo, V.A., Vázquez, A., Martínez, L.C., Jiménez, R. y Puente, J.L. 2011. PerC y GrlA independently regulate Ler expression in enteropathogenic *Escherichia coli*. *Molecular Microbiology*. 82(2): 398-415.

Cadena, E., 2016. Regulación de los genes *E2348C\_1012* y *E2348C\_1013* por los reguladores GrlA y GrlR en *Escherichia coli* enteropatógena. Tesis (Maestría). Universidad Nacional Autónoma de México.

Casadaban, M.J. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu- *Journal of Molecular Biology*. 284: 875-883.

Chavarría, M., Durante-Rodríguez, G., Krell, T., Santiago, C., Brezovsky, J., Damborsky, J. y de Lorenzo, V. 2014. Fructose 1-phosphate is the one and only physiological effector of the Cra (FruR) regulator of *Pseudomonas putida*. *FEBS openbio*. 4(2014): 377-386.

Cheng, C., Tennant, S.M., Azzopardi, K.I., Bennett-Wood, V., Hartland, E.L., Robins-Browne, R.M. y Tauschek, M. 2009. Contribution of the *pst-pshoU* operon to cell adherence by atypical enteropathogenic *Escherichia coli* and virulence of *Citrobacter rodentium*. *Infection and Immunity*. 77(5): 1936-1944.

Crooks, G.E., Hon, G., Chandonia, J, Brenner, S.E. 2004. WebLogo: a sequence logo generator. *Genome Research*. 14(1): 1188-1190.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M. y Finlay, B.B. 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clinical Microbiology Reviews*. 26(4): 822-880.

Crepin, V.F., Collins, J.W., Habibzay, M. y Frankel, G. 2016. *Citrobacter rodentium* mouse of bacterial infection. *Nature Protocols*. 11(10): 1851-1876.

Datsenko, K.A. y Wanner, B.L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS*. 97(12): 6640-6645.

Deng, W., Li, Y., Hardwidge, P.R., Frey, E.A., Pfuetzner, R.A., Lee, S., Gruenheid, S., Strynadka, N.C.J., Puente, J.L., Finlay, B.B. 2005. Regulation of type III secretion hierarchy of translocators and effectors in attaching and effacing bacterial pathogens. *Infection and Immunity*. 73(4): 2135-2146.

Deng, W., de Hoog, C.L., Yu, H.B., Li, Y., Croxen, M.A., Thomas, N.A., Puente, J.L., Foster, L.J. y Finlay, B.B. 2010. A comprehensive proteomic análisis of the type III secretome of *Citrobacter rodentium*. *The Journal of Biological Chemistry*. 285(9): 6790-6800.

Dower, W.J., Miller, J.F. y Ragsdale, C.W. 1998. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acid Research*. 16(13): 6127-6145.



Escalante, A., Cervantes, A.S., Gosset, G. y Bolivar, F. 2012. Current knowledge of the *Escherichia coli* phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. *Applied Microbiology Biotechnology*. 94: 1483-1494.

Gally, D.L. y Stevens, M.P. 2017. Microbe Profile: *Escherichia coli* O157:H7 - notorious relative of the microbiologist's workhorse. *Microbiology*. 163(1): 1-3.

García-Ángulo, V.A., Martínez-Santos, V.I., Villaseñor, T., Santana, F.J., Huerta-Saquero, A., Martínez, L.C., Jiménez, R., Lara-Ochoa, C., Téllez-Sosa, J., Bustamante, V.H. y Puente, J.L. 2012. A distinct regulatory sequence is essential for the expression of a subset of *nle* genes in Attaching and Effacing *Escherichia coli*. *Journal of Bacteriology*. 194(20): 5589-5603.

Gaytán, M.O., Martínez-Santos, V.I., Soto, E. y González-Pedrajo, B. 2016. Type three secretion system in attaching and effacing pathogens. *Frontiers in Cellular and Infection Microbiology*. 6(129).

Görke, B. y Stülke, J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology*. 6: 613-624.

Hazen, T.H., Daugherty, S.C., Shetty, A., Mahurkar, A.A., White, O., Kaper, J.B. y Rasko, D.A. 2015. RNA-Seq analysis of isolate- and growth phase-specific differences in the global transcriptomes of enteropathogenic *Escherichia coli* prototype isolates. *Frontiers in Microbiology*. 6: 569.

Islam, Md.S., Bingle, L.E.H., Pallen, M.J. y Busby, S.J.W. 2011. Organization of the *LEE1* operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Molecular Microbiology*. 79(2): 468-483.

Iyoda, S. y Watanabe, H. 2004. Positive effects of multiple *pch* genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escherichia coli* O157:H7 to Hep-2 cells. *Microbiology*. 150: 2357-2371.

Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T., Ohnishi, M. y Watanabe, H. 2006. The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. *Journal of Bacteriology*. 188(6): 5682-5692.

Jiménez, R., Cruz-Migoni, S.B., Huerta-Saquero, A., Bustamante, V.H. y Puente J.L. 2010. Molecular characterization of GrlA, a specific positive regulator of *ler* expression in enteropathogenic *Escherichia coli*. *Journal of Bacteriology*. 192(18): 4627-4642.

Levine, M.M., Bergquist, E.J., Nalin, D.R., Waterman, D.H., Homick, R.B., Young, C.R. y Sotman, S. 1978. *Escherichia coli* strains that cause diarrhea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. *Lancet*. 1: 1119-1122.

Li, Y., Frey, E., Mackenzie, A.M.R. y Finlay, B.B. 2000. Human response to *Escherichia coli* O157:H7 infection: antibodies to secreted virulence factors. *Infection and Immunity*. 68(9): 5090-5095.

Lu, R., Herrera, B.B., Eshleman, H.D., Fu, Y., Bloom, A., Li, Z., Sacks, D.B. y Goldberg, M.B. 2015. *Shigella* effector OspB activates mTORC1 in a manner that depends on IQGAP1 and promotes cell proliferation. PLOS Pathogens. 11(10).

Martínez-Laguna, Ygnacio., Calva, E. y Puente, J.L. 1999. Autoactivation and environmental regulation of *bfpT* expression, the gene coding for the transcriptional activator of *bfpA* in enteropathogenic *Escherichia coli*. Molecular Microbiology. 33(1): 153-166.

Mayer, M.P. 1995. A new set of useful cloning and expression vectors derived from pBlueScript. Gene. 163: 41-46.

McClure, W.R. 1985. Mechanism and control of transcription initiation in prokaryotes. Annual Review of Biochemistry. 54:171-204.

Mellies, J.L. y Lorenzen, E. 2014. Enterohemorrhagic *Escherichia coli* virulence gene regulation. Microbiology. 2(4): 1-15.

Nelson, D.L. y Kennedy, E.P. 1970. Magnesium transport in *Escherichia coli*: Inhibition by cobaltous ion<sup>+</sup>. The journal of Biological Chemistry. 246(9): 3042-3049.

Njoroge, J.W., Nguyen, Y., Curtis, M.M., Moreira, C.G. y Sperandio, V. 2012. Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli*. mBio. 3(5): 1-12.

Ochoa, T.J. y Contreras, C.A. 2011. Enteropathogenic *E. coli* (EPEC) infection in children. *Current Opinion in infectious Disease*. 24(5): 478-483.

Padavannil, A., Jobichen, C., Mills, E., Velazquez-Campoy, A., Li, K., Leung, K.Y., Mok, Y.K., Rosenchine, I. y Sivaraman, J. 2013. Structure of GrIR-GrIA complex that prevents GrIA activation of virulence genes. *Nature Communications*. 2546.

Pearson, J.S., Giogha, C., Ong, S.Y., Kennedy, C.L., Kelly, M., Robinson, K.S., Lung, T.W.F., Mansell, A., Riedmaier, P., Oates, C.V.L., Zaid, A., Mühlen S., Crepin, V.F., Marches, O., Ang, C., Williamson, N.A., O'Reilly, L.A., Bankovacki, A., Nachbur, U., Infusini, G., Webb, A.I., Silke, J., Strasser, A., Frankel, G. y Hartland, E.L. 2013. A type III effector antagonizes death receptor signaling during bacterial gut infection. *Nature*. 501: 247-251.

Pearson, J.S., Giogha, C., Wong-Fok-Lung, T. y Hartland, E.L. 2016. The genetics of enteropathogenic *Escherichia coli* virulence. *Annual Review of Genetics*. 50: 493-513.

Porter, M.E., Mitchell, P., Free, A., Smith, D.G.E. y Gally, D.L. 2005. The *LEE1* promoters from both Enteropathogenic and Enterohemorrhagic *Escherichia coli* can be activated by PerC-Like proteins from either organism. *Journal of Bacteriology*. 187(2): 458-472.

Puente, J.L., Bieber, D., Ramer, S.W., Murray, W. y Schoolnik. G.K. 1996. The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals. *Molecular Microbiology*. 20(1): 87-100.

Saitoh, T., Iyoda, S., Yamamoto, S., Lu, Y, Shimuta, K., Ohnishi, M., Terajima, J. y Watanabe, H. 2008. Transcription of the *ehx* enterohemolysin gene is positively regulated by GrlA, a global regulator encoded within the Locus of Enterocyte Effacement in enterohemorrhagic *Escherichia coli*. *Journal of Bacteriology*. 190(4): 4822-4830.

Schneider, T.D. y Stephens, R.M. 1999. Sequence logos: a new way to display consensus sequences. *Nucleic Acid Research*. 18(20): 6097-6100.

Shimada, T., Yamamoto, K. y Ishihama, A. 2011. Novel members of the Cra regulon involved in carbon metabolism in *Escherichia coli*. *Journal of Bacteriology*. 193(3): 649-659.

Shultzaberger, R.K., Chen, Z., Lewis, K.A. y Schneider, T.D. 2006. Anatomy of *Escherichia coli*  $\sigma^{70}$  promoters. *Nucleic Acids Research*. 35(3): 771-788.

Winardhi, R.S., Gulvady, R., Mellies, J.L. y Yan, J. 2014. Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic *Escherichia coli* competes off histone-like nucleoid structuring protein (H-NS) through noncooperative DNA binding. *The Journal of Biological Chemistry*. 289(20): 13739-13750.

## ANEXOS

### **Anexo 1.** Concentración de DNA por extracción con fenol-cloroformo.

1. Agregar medio volumen de acetato de amonio 7.5 M pH 7.8 y mezclar por pipeteo.
2. Agregar 250  $\mu$ l de fenol y 250  $\mu$ l de cloroformo:alcohol-isoamílico (24:1). Mezclar por inmersión hasta combinar ambas fases.
3. Centrifugar a 12,000 rpm por 5 min a T°amb.
4. Retirar cuidadosamente la fase acuosa (superior) y pasar a un tubo limpio.
5. Agregar 250  $\mu$ l de cloroformo:alcohol-isoamílico (24:1), mezclando de la misma manera que en el paso 2.
6. Repetir paso 3 y 4.
7. Añadir 2 volúmenes de etanol absoluto e incubar por 1 hr a -80°C.
8. Centrifugar a 13,500 rpm por 30 min. Descartar el sobrenadante.
9. Agregar 1 ml de etanol al 70 % y mezclar.
10. Centrifugar a 12,000 rpm por 5 min. Descartar todo el sobrenadante.
11. Dejar secando la pastilla a T°amb y dejar resuspendiendo con agua MQ durante toda la noche a 4°C.

**Anexo 2.** Preparación de células electrocompetentes y electroporación (Dower *et al.* 1998)

1. Cultivar la cepa deseada toda la noche en agitación a 37°C adicionando los antibióticos necesarios (en caso de tener resistencia en la cepa).
2. Inocular en dilución 1:100 en un matraz con medio LB con los antibióticos necesarios.
3. Incubar a 37°C en agitación hasta que el cultivo llegue a una DO<sub>600</sub> de 0.6-0.8.
4. Transferir el cultivo a tubos de centrifugación de 50 ml.
5. Centrifugar a 8,000 rpm durante 8 min a 4°C.
6. Descartar el sobrenadante (en caso de tener más cultivo, repetir el paso 4, 5 y 6 las veces necesarias).
7. Lavar la pastilla con 40 ml de agua MQ a 4°C.
8. Centrifugar a 8,000 rpm durante 8 min a 4°C.
9. Descartar el sobrenadante y repetir una vez desde el paso 7.
10. Resuspender en 25 ml de glicerol al 10 % a 4°C.
11. Centrifugar a 8,000 rpm durante 8 min a 4°C.
12. Descartar el sobrenadante.
13. Resuspender con glicerol al 10 % de acuerdo al tamaño de la pastilla y según la concentración celular deseada (generalmente entre 200-400 µl para un medio inicial de 100 ml). En caso de no utilizar para electroporar inmediatamente, guardar a -80°C hasta 6 meses.
14. Agregar DNA al tubo con electrocompetentes (si es plásmido circular, utilizar 100 ng por transformación) y mezclar por pipeteo.
15. Electroporar a 2.5 kV en una cámara de electroporación BioRad®.
16. Recuperar con el cultivo completando a 1 ml con medio SOC (950 µl de medio SOC si se utilizó 50 µl de células electrocompetentes) y pasarlo a un tubo.
17. Incubar 1 hr en agitación a 37°C.
18. Sembrar en cajas con el marcador de selección y dejar crecer durante toda la noche a 37°C.

**Anexo 3.** Composición del DMEM.

| <b>DMEM</b>              |                                        |           |             |           |
|--------------------------|----------------------------------------|-----------|-------------|-----------|
| <b>Compuesto</b>         |                                        | <b>pM</b> | <b>mg/L</b> | <b>mM</b> |
| <b>Aminoácidos</b>       | <b>Glicina</b>                         | 75.0      | 30.0        | 0.4000    |
|                          | <b>Hidrocloruro de L-arginina</b>      | 211.0     | 84.0        | 0.3981    |
|                          | <b>L-cisteína · 2HCl</b>               | 313.0     | 63.0        | 0.2013    |
|                          | <b>L-glutamina</b>                     | 146.0     | 580.0       | 3.9726    |
|                          | <b>Hidrocloruro de L-Histidina·H2O</b> | 210.0     | 42.0        | 0.2000    |
|                          | <b>L-iso-leucina</b>                   | 131.0     | 105.0       | 0.8015    |
|                          | <b>L-leucina</b>                       | 131.0     | 105.0       | 0.8015    |
|                          | <b>Hidrocloruro de L-lisina</b>        | 183.0     | 146.0       | 0.7978    |
|                          | <b>L-metionina</b>                     | 149.0     | 30.0        | 0.2013    |
|                          | <b>L-fenilalanina</b>                  | 165.0     | 66.0        | 0.4000    |
|                          | <b>L-serina</b>                        | 105.0     | 42.0        | 0.4000    |
|                          | <b>L-treonina</b>                      | 119.0     | 95.0        | 0.7983    |
|                          | <b>L-triptófano</b>                    | 204.0     | 16.0        | 0.0784    |
|                          | <b>L-tirosina</b>                      | 181.0     | 72.0        | 0.3978    |
| <b>L-valina</b>          | 117.0                                  | 94.0      | 0.8034      |           |
| <b>Vitaminas</b>         | <b>Cloruro de colina</b>               | 140.0     | 4.0         | 0.0286    |
|                          | <b>Pantotenato de D-calcio</b>         | 477.0     | 4.0         | 0.0084    |
|                          | <b>Ácido fólico</b>                    | 441.0     | 4.0         | 0.0091    |
|                          | <b>Niacinamida</b>                     | 122.0     | 4.0         | 0.0328    |
|                          | <b>Hidrocloruro de piridoxal</b>       | 206.0     | 4.0         | 0.0194    |
|                          | <b>Riboflavina</b>                     | 376.0     | 0.4         | 0.0011    |
|                          | <b>Hidrocloruro de tiamina</b>         | 337.0     | 4.0         | 0.0119    |
|                          | <b>I-inositol</b>                      | 180.0     | 7.2         | 0.0400    |
| <b>Sales inorgánicas</b> | <b>Cloruro de calcio</b>               | 147.0     | 264.0       | 1.7959    |
|                          | <b>Nitrato férrico</b>                 | 404.0     | 0.1         | 0.0002    |
|                          | <b>Sulfato de magnesio</b>             | 246.0     | 200.0       | 0.8130    |
|                          | <b>Cloruro de potasio</b>              | 75.0      | 400.0       | 5.3333    |
|                          | <b>Bicarbonato de sodio</b>            | 84.0      | 3700.0      | 44.0476   |
|                          | <b>Cloruro de sodio</b>                | 58.0      | 6400.0      | 110.3448  |
|                          | <b>Fosfato de sodio monobásico</b>     | 154.0     | 141.0       | 0.9156    |
| <b>Otros</b>             | <b>D-glucosa</b>                       | 180.0     | 4500.0      | 25.0000   |
|                          | <b>Rojo de fenol</b>                   | 376.4     | 15.0        | 0.0399    |
|                          | <b>HEPES</b>                           | 238.3     | 5958.0      | 25.0021   |



**Anexo 4.** Composición de medios mínimos.

| <b>Medio mínimo N (MM-N)</b>         |           |             |           |
|--------------------------------------|-----------|-------------|-----------|
| <b>Compuesto</b>                     | <b>pM</b> | <b>mg/L</b> | <b>mM</b> |
| <b>Cloruro de potasio</b>            | 75.00     | 360         | 4.800     |
| <b>Sulfato de amonio</b>             | 132.14    | 980         | 7.416     |
| <b>Sulfato de potasio</b>            | 174.26    | 86          | 0.494     |
| <b>Fosfato de potasio monobásico</b> | 154.00    | 136         | 0.883     |
| <b>Ácidos casamino</b>               |           | 1000        |           |
| <b>Cloruro de magnesio</b>           | 95.20     | 19          | 0.200     |
| <b>Glicerol</b>                      | 92.10     | 6300        | 68.404    |
| <b>Glucosa</b>                       | 180.16    | 4800        | 26.643    |
| <b>Tris-HCl 0.1M pH 7.5</b>          |           |             |           |

| <b>Medio mínimo 9 (MM-9)</b>         |           |             |           |
|--------------------------------------|-----------|-------------|-----------|
| <b>Compuesto</b>                     | <b>pM</b> | <b>mg/L</b> | <b>mM</b> |
| <b>Fosfato de sodio dibásico</b>     | 142.00    | 12800       | 90.14     |
| <b>Fosfato de potasio monobásico</b> | 154.00    | 3000        | 19.48     |
| <b>Cloruro de sodio</b>              | 58.44     | 500         | 8.56      |
| <b>Cloruro de amonio</b>             | 53.50     | 1000        | 18.69     |
| <b>Glucosa</b>                       | 180.16    | 4000        | 22.20     |
| <b>Cloruro de calcio</b>             | 111.00    | 0.0111      | 0.0001    |
| <b>Sulfato de magnesio</b>           | 120.36    | 0.1204      | 0.0010    |