



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**TERMORIBOSWITCHES:
MOLÉCULAS DE RNA REGULATORIAS
RESPONSIVAS A TEMPERATURA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

B I Ó L O G A

P R E S E N T A:

VIRIDIANA ÁVILA MAGAÑA



**DIRECTOR DE TESIS:
DR. ENRIQUE MERINO PÉREZ
2009**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de datos del jurado

1. Avila
Magaña
Viridiana
56 55 64 85
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
301501389
2. Datos del tutor
Dr.
Enrique
Merino
Pérez
3. Datos del sinodal 1
Dr.
Gabriel
del Río
Guerra
4. Datos del sinodal 2
Dra.
Rosa María
Gutiérrez
Ríos
5. Datos del sinodal 3
Dr.
Arturo Carlos II
Becerra
Bracho
6. Datos del sinodal 4
M en C
José Ricardo
Ciria
Merce
7. Datos del trabajo escrito
Termoriboswitches: Moléculas de RNA regulatorias responsivas a
temperatura
94 p.
2009

Esta tesis fue elaborada en el Instituto de Biotecnología (IBt), UNAM, Cuernavaca, Morelos, en el laboratorio del grupo de Genómica Computacional del Departamento de Microbiología Molecular, bajo la dirección del Dr. Enrique Merino Pérez, con el financiamiento y apoyo económico del IBt, DGAPA y CONACYT.

Se contó con el siguiente comité revisor de tesis:

Dr. Enrique Merino Pérez
Dr. Gabriel del Río Guerra
Dra. Rosa María Gutiérrez Ríos
Dr. Arturo Carlos Il Becerra Bracho
M.C. José Ricardo Ciria Mercé

Agradecimientos

Son aquellas pequeñas cosas,
que nos dejó un tiempo de rosas
en un rincón,
en un papel
o en un cajón.
JM SERRAT.

Quisiera agradecer a todos los que han estado de manera importante durante mi formación académica y mi crecimiento personal.

Particularmente en esta odisea a Enrique Merino, quien se ha convertido en una de esas personas favoritas, a las que les agradezco el estar aquí, para crear, discutir, formar, crecer, recodar, aprender, trabajar, amoldar, viajar, volar y sobretodo disfrutar el quehacer cotidiano, con esas pequeñas cosas que al final se transforman en un intercambio fructífero que va de maravilla y consiguen alimentar el espíritu y las grandes expectativas. Te quiero Lucas.

Al comité que con gran dedicación ha revisado mi tesis, gracias por sus comentarios, sugerencias, y tiempo para hacer de esta una mejor tesis: Rosa María Gutiérrez, Gabriel del Río, Ricardo Ciria, Arturo Becerra y Enrique Merino.

En el lado A de la cinta (o séase en la parte bioinformática) agradezco a Patricia Oliver por su paciencia, dedicación y su gran forma de enseñar a programar, gracias.

En el lado B, experimental, a Zuemy Rodríguez y José Luis Rodríguez, por instruirme y estar al pendiente de mis resultados experimentales siempre.

Por supuesto en ambos lados siempre hubieron personas apoyándome, explicándome e intercambiando su valiosa experiencia, cuyo resultado se traduce en una amistad: Ana Gutiérrez, Christian Martínez, Mario Alberto Martínez, Nancy Mena, Ricardo Ciria, Ma. Luisa Tabche, Alberto Hernández, Miguel Yáñez, Daniel Segura y Raúl Noguez.

En fin a todos mis compañeros y amigos del laboratorio, a Don Lalo de manera muy especial por prepararme todas mis ocurrencias.

A todos los grandes amigos que hice en el Instituto y que han estado a lo largo de todo este tiempo ayudándome de cierta forma a concretar este proyecto, complaciendo mis locuras, mi emoción y mis frustraciones: Erika Guerra, Francisco Gasteazoro, Lorena Sánchez mi comadre, Adriana Espinosa, Georgina Peñalosa, Alberto Echeverría, Omar Piña, Karla Martínez, Diego Cuervo, y a todas las amistades ustedes saben quienes son, porque han estado ahí. Gracias.

A mis ñoras del corazón, por estar al pendiente, darme los mejores ánimos y postres, por la gran complicidad, empatía y amor que nos tenemos. Nada es lo mismo sin ustedes queridas. A Tania Valdivia por las risas, la hermandad y el tiempo, te adoro.

A Fabiola Téllez y a sus pequeñas grandes, por dejarme ser parte de su cotidianidad y ser cómplices en el rancho, te quiero.

A toda mi familia por ser parte de un todo, mi vida pues. En verdad gracias a cada uno de ustedes, que ya saben: los amo.

A mis padres, sin ellos esto no tendría este sabor, dulce-picoso-ácido, que te envuelve y enchina la piel, les estoy infinitamente agradecida y los amo.

A Mariana por realmente cumplir con las expectativas que envuelven a esa palabra, siempre ahí, mi amorosa incondicional protectora, gracias hermana.

Dedicatoria



De manera muy especial a Rosa Luz Herrera Estañol, gran mujer y abuela.
Te amo siempre.

ÍNDICE

INTRODUCCIÓN 1

1.	Revisión de los mecanismos de regulación durante el choque térmico	3
1.1.	Implicaciones celulares de la termorregulación	3
1.2.	Termorregulación basada en proteínas	4
1.3.	Regulación de los genes asociados al choque térmico dependientes de factores sigma.....	4
1.4.	Termorregulación basada en DNA.....	6
1.5.	El papel del RNA en la regulación genética. Sensores de RNA.....	6
1.6.	Termorregulación vía RNA.....	10
1.6.1.	Riboswitches del gen <i>rpoH</i> regulan la traducción de σ_{32} en respuesta a cambios de temperatura	12
1.6.2.	Un termómetro que se activa con el choque térmico por frío	12
1.6.3.	Termómetros de RNA que juegan un papel importante en la virulencia....	13
1.6.3.1.	¿Lisis o lisogenia?: El desarrollo del fago lambda	13
1.6.3.2.	Regulación de VirF por elementos de RNA en <i>Yersinia pestis</i>	14
1.6.3.3.	Regulación de PrfA por elementos de RNA en <i>Listeria monocytogenes</i> ...	15
1.6.3.4.	FourU un termómetro más en <i>Salmonella enterica</i>	15
2.	Represión de la expresión de los genes de choque térmico por ROSE: el termómetro de RNA modelo	16

JUSTIFICACIÓN 21

HIPÓTESIS 23

OBJETIVOS 23

1.1. METODOLOGÍA Y DESARROLLO BIOINFORMÁTICO 25

1.1.1	Grupos de genes ortólogos (COGs) asociados al choque térmico por calor y factores sigma	25
1.1.2.	Regiones de regulación intergénicas	27
1.1.3.	Genomas no-redundantes	28
1.1.4.	Agrupamiento de secuencias iMUR por COG de acuerdo a su filogenia ...	30
1.1.5.	Buscando firmas conservadas: MEME y MAST	30

1.1.6.	Predicción de la estructura secundaria	32
--------	--	----

1.2. RESULTADOS Y DISCUSIÓN BIOINFORMÁTICA 33

1.2.1.	Análisis de los COGs relacionados con el choque térmico por calor (“shock”).....	33
1.2.1.1.	Elementos reguladores conocidos basados en RNA.....	33
1.2.1.2.	Sitios de interacción DNA-proteína.....	38
1.2.1.3.	Motivos encontrados en arqueobacterias que aún no han sido descritos .	38
1.2.2.	Análisis de los COGs relacionados con factores sigma.....	41
1.2.2.1.	COG1191 (DNA-directed RNA polymerase specialized sigma subunit): Un sensor basado en RNA, implicado en la esporulación en <i>Bacillus subtilis</i>	42
1.2.2.2.	El caso de los factores sigma del COG0568 (RLE0166) en Proteobacteria (DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)	46
1.2.2.3.	COG1595: Factores sigma involucrados durante el estrés extracitoplasmático. Dos casos diferentes: Firmicutes y Proteobacteria ...	49
1.2.2.3.1.	Factores sigma involucrados durante el estrés extracitoplasmático en Firmicutes	50
1.2.2.3.2.	Un factor sigma involucrado en la expresión de genes relacionados con el transporte de citrato férrico en Proteobacteria.....	51

2.1. ESTRATEGIA EXPERIMENTAL 56

2.2. METODOLOGÍA EXPERIMENTAL 58

2.2.1.	Amplificación de los insertos de PCR: <i>ROSE</i> , <i>cat</i> y <i>sacB</i>	58
2.2.2.	Reacción de fosforilación-ligación entre <i>ROSE</i> y <i>cat</i>	58
2.2.3.	Clonación del producto <i>ROSE-cat</i> en el vector pMOS-Blue	59
2.2.4.	Apertura del vector pMOS-Blue <i>ROSE-cat</i> por PCR, Clonación de <i>sacB</i> a través de la fosforilación-ligación	60
2.2.5.	Ensayos de crecimiento para probar y seleccionar la construcción de interés.....	61
2.2.6.	Otras consideraciones metodológicas	63

2.3. RESULTADOS Y DISCUSIÓN EXPERIMENTAL 64

2.3.1	Amplificación de los insertos por PCR: <i>ROSE</i> , <i>cat</i> y <i>sacB</i>	64
-------	---	----

2.3.2 La expresión de la construcción pMOS-Blue *ROSE-cat* está
termorregulada por *ROSE* 65

2.3.3 Construcción genética pMOS-Blue *ROSE-cat -sacB*..... 66

DISCUSIÓN GENERAL Y CONCLUSIONES 72

PERSPECTIVAS 77

REFERENCIAS 78

APÉNDICES 83

1. Apéndice I: Scripts de los programas bioinformáticos elaborados para esta tesis
(Perl)..... 83

Introducción

Los seres vivos interactúan con su entorno de manera dinámica, formando parte de éste e interpretando el mundo que les circunda a través de un arreglo sensorial. En particular, las bacterias y arqueobacterias, son capaces de habitar y explotar múltiples ambientes y nichos ecológicos: las hay de vida libre en el suelo, agua dulce, agua salada y aire; todas ellas adaptadas a las variaciones que dichos ambientes presentan. De igual forma hay las formas de vida simbiotes, que viven en conjunto con otro organismo, en donde ambos se benefician, por ejemplo aquellas bacterias presentes en nódulos de una planta que se encargan de fijar nitrógeno. Otro amplio grupo de interés médico son las bacterias patógenas, causantes de distintas enfermedades.

Es sorprendente que existan microorganismos adaptados a condiciones de temperatura extremas: termófilos, mesófilos y psicrófilos. La cuestión subyacente son los mecanismos moleculares protectivos que son establecidas a nivel celular y molecular. Así, las bacterias y otros seres vivos tienen la capacidad de responder a estímulos externos, a través de múltiples mecanismos que les permiten percibir fluctuaciones en su medio y contender con dichos cambios, adaptándose de forma versátil a éste; algunos de los parámetros que deben ser regulados finamente, son por ejemplo, la disposición de nutrientes, osmolaridad, pH y temperatura, entre otros que son necesarios para la sobrevivencia. Dichas respuestas dependen de sensores, los cuales finalmente al ser vistos a nivel molecular, se encuentran intrínsecamente ligados a la regulación y expresión concertada del genoma; de manera muy general genes que se encontraban activados necesitan ser desactivados y genes latentes serán activados. En torno a esta última idea gira la visión de los seres vivos comprendidos como una red genética y metabólica compleja capaz de evolucionar, en donde cada elemento regulador, ya sea el propio DNA codificante, secuencias aledañas a éste, las proteínas o RNAs pequeños que regulan a su expresión, todos ellos funcionan como sensores on/off (encendido/apagado).

Los mecanismos moleculares que han regido el principio de protección y adaptación al medio incluyen a los elementos que pueden regular la expresión específica de los genes. Estos dispositivos sensoriales toman formas innumerables y explotan la diversidad funcional de las proteínas. Generalmente en bacterias, la regulación genética implicada en respuestas al medio se realiza a nivel transcripcional por la acción de proteínas reguladoras.

Las proteínas reguladoras generalmente son factores transcripcionales, que se unen al DNA en una secuencia particular en las regiones que flanquean a los genes, denominadas intergénicas. Estas regiones son de gran importancia debido a que en ellas están contenidos los

elementos de regulación en donde está la información necesaria para que sus genes correspondientes se expresen bajo determinadas condiciones ambientales.

Algunas de estas condiciones ambientales como pueden ser nutrientes o factores de crecimiento son sensadas por la célula: literalmente ella puede interactuar químicamente con éstos; pero ¿qué ocurre con los parámetros físicos tales como la luz, ambiente redox, fluidez en la membrana y la temperatura? Si bien es cierto que la mayoría de los sensores explotan la diversidad funcional ilimitada de las proteínas, existen unos cuantos que toman la ventaja del uso de otro biopolímero fundamental: el RNA.

El dogma central de la Biología Molecular describe el flujo y la regulación de la información de la siguiente manera: fluye del DNA a las proteínas, usando una molécula adaptadora, el RNA. De tal modo que se puede pensar que un gen equivale a una proteína y esta a su vez lleva a cabo una función (Figura1). Sin embargo hasta hace poco tiempo a través de estudios se ha demostrado que el RNA es una molécula versátil que esta implicada en diversos procesos moleculares y celulares (Mattick, 2003).

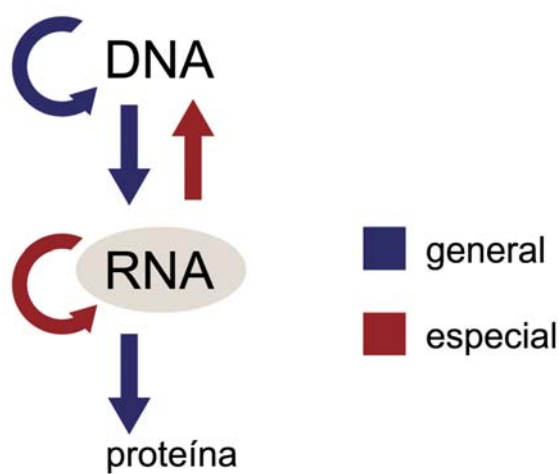


Figura 1. El dogma central de la Biología Molecular. De acuerdo al esquema general la información va desde el DNA hasta las proteínas, usando al RNA como molécula intermediaria. Aunque recientemente se ha dado la explosión del descubrimiento del RNA como una molécula versátil, con muchas más funciones de las que se pensaban.

Recientemente han sido descubiertos termosensores basados en regiones regulatorias de mRNAs cuyo producto proteico se encuentra relacionado con la respuesta al estrés por calor, funcionando a través de cambios estructurales en la secuencia del mRNA, inducidos por cambios en la temperatura, modulando así su propia expresión. A estos mRNA se les conoce como termómetros de RNA (Narberhaus et al., 2006).

Así el descifrar la información incluida en las regiones regulatorias de mRNA que posiblemente estén involucrados en el estrés por calor en microorganismos, llevará a una mejor y más completa comprensión del modo en que estos actúan dinámicamente con el medio, respecto

a la temperatura, así como sus capacidades y limitaciones que supone el uso de nuevos dispositivos sensoriales moleculares basados en RNA.

En el presente trabajo se integra una revisión de los mecanismos moleculares de regulación durante el estrés por choque que han sido descritos hasta la fecha. En una segunda instancia, se realiza una búsqueda bioinformática exhaustiva de nuevos termómetros de RNA compartidos entre bacterias y arqueobacterias de genomas secuenciados y que pueden ser potencialmente caracterizados mediante ensayos biológicos funcionales. En adición a lo anterior, en la presente tesis se presenta como un caso particular de estudio, se realiza una propuesta teórico-experimental para la caracterización de un termómetro de RNA ampliamente conservado en organismos bacterianos: ROSE.

1. Revisión de los mecanismos de regulación durante el choque térmico

1.1. Implicaciones celulares de la termorregulación.

La temperatura se encuentra entre los parámetros físicos que deben ser monitoreado constantemente y regulado finamente debido a que a temperaturas elevadas las proteínas celulares se desnaturalizan. Una serie de procesos son controlados por la temperatura en organismos procariontes: la expresión de los genes que codifican para proteínas de choque térmico por calor (Heat Shock Proteins), de choque frío (Cold Shock Proteins). La inducción de las proteínas de choque térmico por calor (HSPs) para proteger a la célula del daño por la inducción de calor es un proceso biológico universal (Yura et al., 1993). Por lo común, estos procesos están asociados a estrés, por lo que es necesaria la presencia de mecanismos sensibles capaces de regular y responder a los cambios térmicos en el medio. La señal de calor es interpretada por la bacteria como protección de ambientes severos, limitación o abundancia de nutrientes y la oportunidad de crecer y reproducirse. Así la exposición a temperaturas extremas en las bacterias tiene un gran impacto a nivel de las biomoléculas que conforman a estos. Las temperaturas altas desnaturalizan a las proteínas celulares e incrementan la permeabilidad de la membrana celular (Vigh et al., 1998). Por otro lado, las bajas temperaturas ocasionan que la membrana se encuentre rígida y estabilizan las estructuras del mRNA, lo que comúnmente disminuye la eficiencia de traducción (Phadtare, 2004).

Las células son capaces de monitorear temperaturas dañinas que se desplacen hacia arriba o hacia abajo del rango normal celular, induciendo la expresión de las proteínas de choque térmico por calor o por frío, respectivamente. Sin embargo, es destacable que las bacterias no monitorean únicamente temperaturas extremas, además monitorean los cambios en la temperatura dentro del rango de crecimiento fisiológico (Narberhaus, 2002).

Aunado a lo anterior, la temperatura es una señal fundamental en el proceso de virulencia en organismos patógenos, ya que indica el momento en que éstos se encuentran en su estado de vida libre o en contacto con su huésped. Dado lo anterior, dichos organismos requieren de termosensores para inducir los genes de virulencia que les permiten invadir y colonizar a los hospederos.

1.2. Termorregulación basada en proteínas.

Los mecanismos moleculares mediante los cuales las células llevan a cabo el sensado de la temperatura es mediado tanto por proteínas como por ácidos nucleicos. En ambos casos, dichas moléculas sufren cambios conformacionales que inducen formas activas o inactivas del regulador en cuestión.

En relación a las proteínas reguladoras cabe señalar que en general existen pocas proteínas con una capacidad intrínseca para monitorear la temperatura. Ejemplo de lo anterior es la proteína de unión a DNA TlpA en *Salmonella typhimurium*, la cual es directamente sensible a la temperatura (Storz, 1999). TlpA, está codificada en un plásmido de virulencia, y es un represor transcripcional de su propia síntesis. La conformación del dominio de enrollamiento de TlpA es sensible a la temperatura ambiental y regula la multimerización y la actividad de unión al DNA. A la temperatura del hospedero, TlpA se presenta de manera monomérica no-funcional (Hurme et al., 1997). Aunque el único blanco transcripcional de TlpA que se conoce es el mismo, la conservación de *tlpA* que se presenta en plásmidos de virulencia de varias cepas de *Salmonella* hace pensar que esta proteína está posiblemente involucrada en el proceso de infección. En muchos otros casos el monitoreo de la temperatura es indirecto y requiere de sistemas regulatorios de retroalimentación negativa que actúan ajustando la cantidad y actividad de proteínas reguladoras.

La regulación negativa de los genes de choque térmico en bacterias Gram-positivas como *B. subtilis*, depende de la unión de una proteína represora con la secuencia blanco en la región del promotor. El ejemplo de este mecanismo de regulación es CIRCE (Controlling Inverted Repeat of Chaperone Expression). CIRCE funciona como sitio de unión para la proteína represora HrcA, y también promoviendo la formación de una estructura secundaria no estabilizada en el mRNA bajo condiciones normales (Nocker et al., 2001).

1.3. Regulación de los genes asociados al choque térmico dependientes de factores sigma.

Una serie de genes se induce si se somete a las células bacterianas a incrementos repentinos de temperatura, los cuales codifican para las proteínas de choque térmico por calor

también conocidas como “Heat Shock Proteins” (HSPs). La estructura primaria de la mayoría de estas proteínas se encuentra altamente conservada durante la evolución. Existen dos clases de HSPs: las chaperonas moleculares (DnaK y GroEL) y las proteasas dependientes de ATP (Lon, Clp y FtsH). Las primeras se encargan de asegurar que los polipéptidos se plieguen de manera correcta en la célula, mientras que las proteasas degradan aquellas con plegamiento erróneo y que no son capaces de replegarse en su conformación tridimensional nativa (Schumann, 2003).

Todos los genes asociados al choque térmico se expresan a temperatura ambiente, la mayoría a bajos niveles y son inducidos momentáneamente después del choque térmico y su regulación en el organismo modelo *bona fide* en bacterias Gram negativas, *E. coli*, depende de los factores sigma σ_{32} (RpoH) y σ_E . El regulón de σ_{32} está constituido por el conjunto de chaperonas clásicas (GroEL/S y DnaK/J), proteasas (Lon y FtsH) y otras HSPs mientras que el regulón de σ_E codifica para una proteasa periplásmica (HtrA), así como para diferentes proteínas involucradas en el plegamiento proteico, inducidas durante estrés extracitoplasmático (Yura y Nakahigashi, 1999). En dicha bacteria, las proteínas de choque térmico por calor representan menos del 2% del extracto de proteínas totales a una temperatura de 30°C y notablemente del 20 al 25% de las proteínas totales a 46°C. Dado lo anterior, resulta evidente que la célula requiere mantener un balance entre el efecto protector de las HSPs y la carga metabólica que implica la sobre-expresión proteica, lo cual se logra a través de todo un circuito de regulación entorno a σ_{32} ya que su actividad está regulada precisamente por la interacción con DnaK y la co-chaperona DnaJ. Estas chaperonas, además de interactuar con las proteínas desnaturalizadas por el incremento de temperatura, también se unen a σ_{32} , limitando su unión con la RNA polimerasa. El circuito de regulación consiste en la competencia de las proteínas blanco de estas chaperonas. En altas temperaturas la cantidad de proteínas desnaturalizadas es muy alta, por lo que estas chaperonas estarán básicamente interactuando con las proteínas desnaturalizadas y de manera limitada con σ_{32} . En esta condición los genes asociados al choque térmico cuya transcripción es dependiente de σ_{32} , son altamente expresados. A bajas temperaturas, la cantidad de proteínas desnaturalizadas es mucho menor y por tanto estas chaperonas estarán disponibles para interactuar con σ_{32} , inhibiendo su actividad transcripcional. Este mecanismo de regulación es un circuito de retroalimentación negativa conocido como circuito de retroalimentación secuestrador (El Samad et al., 2005).

Cabe señalar que el circuito de retroalimentación negativa de σ_{32} anteriormente mencionado, responde principalmente a la acumulación celular de proteínas desnaturalizadas y no a un cambio en la temperatura intracelular de las bacterias (Waldminghaus et al., 2007b). Para

ello, la expresión de *rpoH* presenta un segundo nivel de regulación basado en un termómetro de RNA en respuesta al choque térmico, tal y como será descrito en la sección 1.6.1.

1.4. Termorregulación basada en DNA.

Otra biomolécula termo-sensible es el DNA, en donde se ha observado que su topología es importante para modular la expresión genética en respuesta a la temperatura (Hurme y Rhen, 1998). Los cambios en el super-enrollamiento del DNA afectan la actividad de ciertos promotores. Aún no es claro si se trata de un efecto inducido directamente por la temperatura, o si es mediado a través de proteínas presentes capaces de sensor cambios, que hasta ahora no han sido descritas. Un ejemplo del papel del superenrollamiento del DNA se presenta en bacterias patógenas, las cuales tienen una proteína parecida a las histonas, H-NS, que se une al DNA como oligómero y protege al *loci* para que no ocurra la transcripción a temperaturas por debajo de la temperatura fisiológica. Así, las temperaturas altas separan los sitios de unión de H-NS, favoreciendo la disociación de los oligómeros. Los monómeros de H-NS se liberan rápidamente del DNA, permitiendo la transcripción a través de la exposición de sitios de metilación. Tal es el caso del gen *pap* que codifica para el pili en *E. coli*; o el caso de *Shigella flexneri*, en donde el promotor *virF*, el cual codifica para un activador de VirB (este último se encarga de activar algunos operones de genes de virulencia) presenta un cambio estructural a 37°C que lo hace menos accesible a la proteína H-NS y consecuentemente competente durante la transcripción (Newman y Weiner, 2002).

1.5. El papel del RNA en la regulación genética. Sensores de RNA.

De acuerdo a la visión tradicional del Dogma central de la Biología Molecular, los ácidos nucleicos almacenan las instrucciones de los procesos celulares, mientras que las proteínas realizan los procesos celulares a través de funciones enzimáticas y de regulación (Fig. 1). Este simple esquema ha cambiado sustancialmente en las dos últimas décadas en las que se ha venido trabajando con RNA. Aunque prevalece la visión de que esta molécula únicamente tiene un papel de intermediario en el dogma DNA→RNA→ proteínas, se han descubierto nuevas funciones y actividades en el RNA.

A lo largo de la escala filogenética, en todos los organismos el RNA desempeña funciones en la mediación de procesos enzimáticos tales como el corte de RNA, el splicing y la traducción (Lai, 2003). Los RNAs no-codificantes (ncRNA) también están relacionados con una amplia variedad de mecanismos de regulación genética que operan a nivel de DNA y mRNA (Dernburg y Karpen, 2002). La relación estructura-función de las ribozimas y los tRNAs ha sido estudiada como paradigma de la complejidad de las moléculas de RNA (Waldminghaus et al., 2007a).

Los RNAs celulares no son moléculas lineales, forman estructuras tridimensionales muy complejas a través del apareamiento intramolecular de regiones complementarias en la secuencia. Estas características estructurales son usadas para llevar a cabo muchas funciones biológicas. Entre dichas funciones se encuentran los sensores de RNA, que al llevar a cabo cambios conformacionales en respuesta a condiciones ambientales, pueden regular y/o modular la expresión genética. Se han distinguido dos clases principales: elementos que actúan en *cis*, es decir que se encuentran embebidos en el mRNA, y los elementos que actúan en *trans*, como es el caso de los ncRNAs pequeños que funcionan apareándose con sus correspondientes RNAs blanco mediante interacciones tipo *Watson-Crick*. En el caso de bacterias, esta interacción es auxiliada por la chaperona de RNA Hfq y puede tener un efecto sobre la expresión génica positivo (aumentando la estabilidad del mRNA o impidiendo la formación de estructuras secundarias que secuestran al sitio de unión a ribosoma o Shine-Dalgarno SD) o negativo (al unirse al SD, impidiendo el acceso del ribosoma). En organismos superiores, la unión de los ncRNAs a sus correspondientes mRNA blancos tienen siempre un efecto negativo en la expresión génica, ya que sirven para señalar su degradación a través del complejo proteico llamado RISC (Gottesman, 2005) (Figura 2).

Los sensores de RNA en *cis* que sufren cambios estructurales en respuesta a señales físicas o químicas también son conocidos como riboswitches, ya que monitorean el estado metabólico de la célula al unirse, con gran especificidad y gran afinidad, a ciertos metabolitos esenciales (Narberhaus *et al.*, 2006) (Figura 3). Los riboswitches están localizados en el extremo 5' de la región no-traducida (UnTranslated RNA o UTR) del mRNA de genes que codifican para la biosíntesis o el transporte del ligando y su efecto de represión o activación se presenta sobre el mismo gen. Su arquitectura se compone de una región receptora mejor conocida como aptámero, que se caracteriza por presentar una secuencia consenso específica para la molécula sustrato o ligando (ésta puede ser un precursor vitamínico, un nucleótido o un aminoácido), la cual transduce la señal vía una plataforma de expresión, comúnmente un atenuador transcripcional o traduccional (Mandal y Breaker, 2004).

Los mecanismos de regulación que son llevados a cabo por riboswitches incluyen la terminación prematura de la transcripción, el auto-corte del mRNA, así como el bloqueo de la traducción por el enmascaramiento de la secuencia SD (Nudler y Mironov, 2004) (Figura 4A y 4B). Se estima que en *Bacillus subtilis* el 2% de todos sus genes podría estar controlado por riboswitches, lo cual nos habla de la importancia de los elementos de RNA regulatorios en el control genético en procariontes (Mandal *et al.*, 2003).

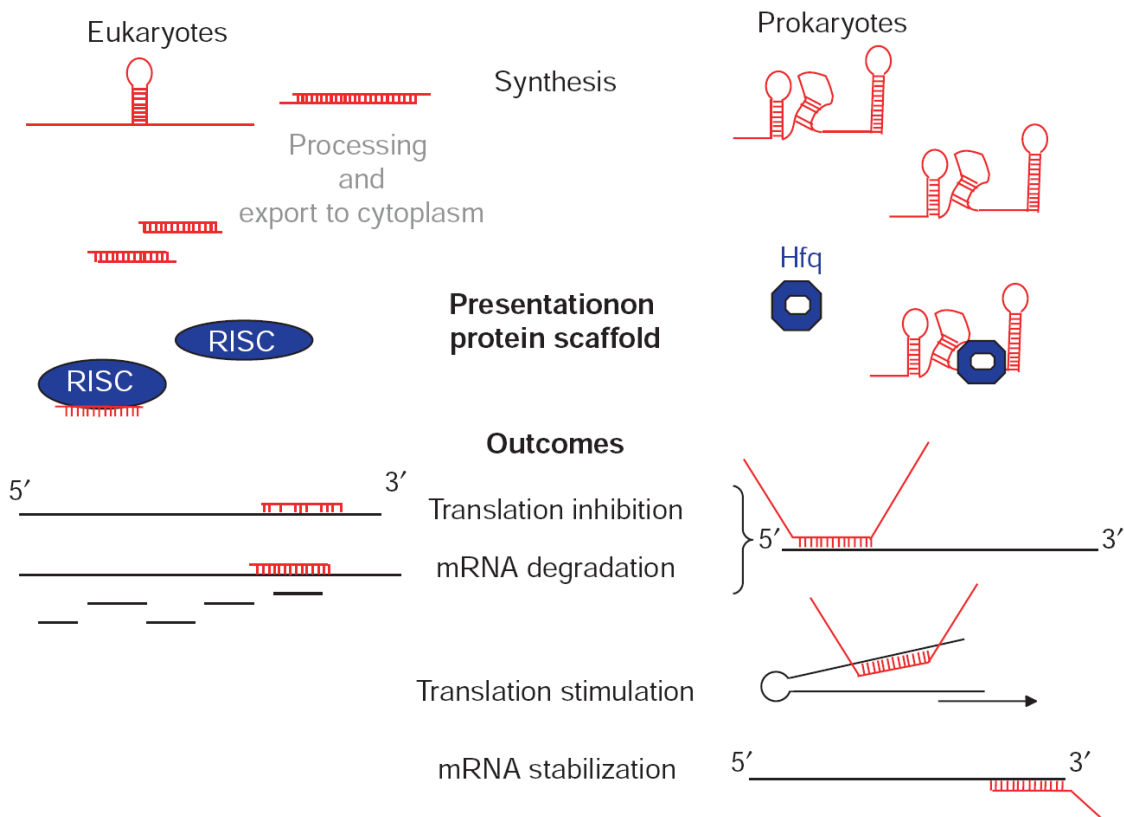


Figura 2. Función de RNAs pequeños no codificantes en eucariontes y procariontes. Los RNAs regulatorios no codificantes se muestran en rojo, los mRNAs blanco se muestran en negro. Las proteínas involucradas en la función de estos RNAs no codificantes se muestran en azul. La síntesis de los microRNAs eucariontes (izquierda) es inicialmente un pasador largo, después de un evento de procesamiento inicial, el pasador es exportado desde el núcleo y posteriormente procesado por la ribonucleasa Dicer a una molécula de doble cadena de 22 nt. En algunos casos múltiples microRNAs derivan de un solo transcrito. Se muestra como una de las cadenas de este doble RNA se une a la proteína RISC, este complejo reconoce al mRNA blanco y se aparean en el extremo 3'UTR. Un apareamiento impreciso propicia la inhibición de la traducción. Si existe el apareamiento preciso entonces el mRNA es degradado. Los RNAs pequeños en procariontes son producidos como un transcrito único y generalmente se encuentran altamente estructurados (derecha). No requieren de un procesamiento posterior, se unen directamente a la proteína hexamérica Hfq, para interactuar con el mRNA blanco. Subsecuentemente, el RNA pequeño y el mRNA blanco se aparean, generalmente en el extremo 5' cerca del inicio de la traducción. El apareamiento puede resultar en la inhibición de la traducción (microRNAs y algunos RNAs pequeños procariontes), la degradación del mRNA (RNAs de interferencia pequeños y muchos RNAs pequeños procariontes), la estimulación de la traducción, así como la estabilización del mRNA (RNAs pequeños procariontes). Tomado de (20).

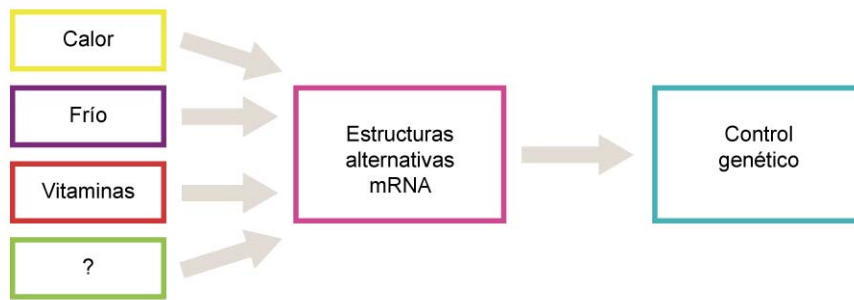


Figura 3.
Principio de la regulación por sensores de RNA.
Modificado de (38).

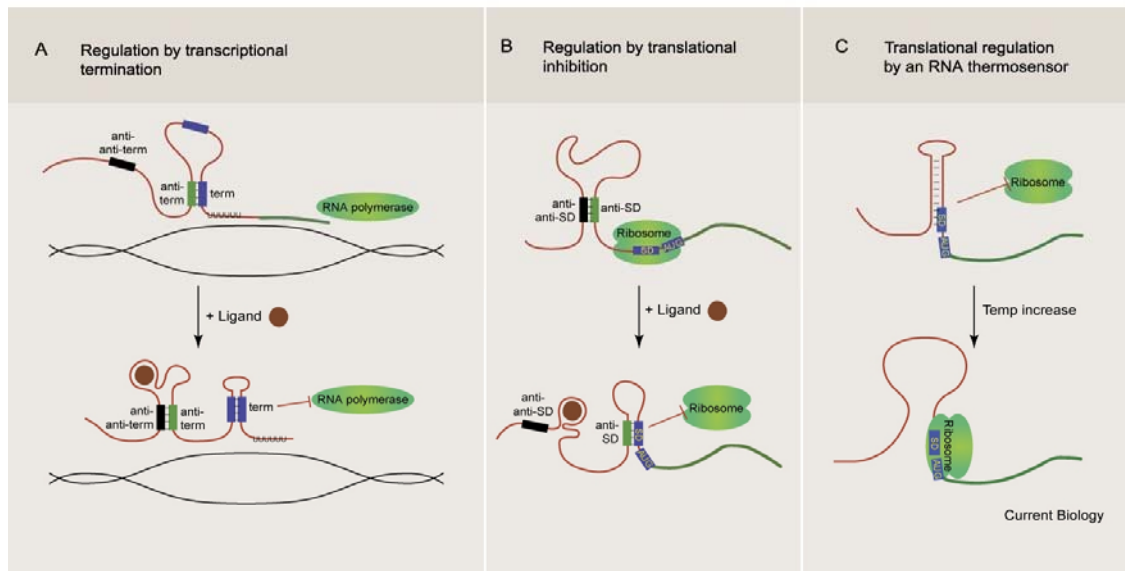


Figura 4. Diferentes estrategias en la regulación genética relacionadas con el RNA. Moléculas pequeñas se unen a los riboswitches, estos actúan por atenuación transcripcional (A) e inhibición traduccional (B). (A) Mecanismo de término de la transcripción. (Arriba) En la ausencia del metabolito relevante, parte del terminador se une al anti-terminador y no es funcional. Ocurre la transcripción pasando la secuencia de poli-uridinas seguidas de la secuencia terminadora y se extiende hasta la secuencia codificante. (Abajo) En presencia del metabolito el ligando se une al RNA el cual adopta una conformación alternativa, en la cual el anti-terminador se une al anti-anti-terminador. Lo anterior permite la formación de un pasador estructuralmente distintivo que induce la liberación de la RNA polimerasa seguida de la secuencia poli-uridina. (B) Mecanismo de inhibición traduccional. (Arriba) En ausencia del cofactor de unión, la secuencia anti-SD es secuestrada por la secuencia anti-anti-SD; esta conformación permite que el ribosoma se una a la secuencia SD y ocurra así la traducción. (Abajo) El ligando reestructura el RNA lo cual permite que la secuencia anti-SD se una a la secuencia SD, ocurriendo la inhibición de la traducción. (C) Regulación traduccional por un riboswitch térmico. (Arriba) Un estructura tallo-asa en donde la secuencia SD se encuentra apareada previene la traducción a temperaturas bajas. (Abajo) Un incremento en la temperatura hace que esta estructura se abra (fusión), permitiendo el acceso del ribosoma a la secuencia SD, permitiendo la traducción de los genes río abajo. Tomada de (29).

1.6. Termoregulación vía RNA.

Otros sensores de RNA que actúan en *cis* son los termómetros de RNA o termoriboswtiches, que en contraste con los riboswtiches de unión a metabolito, responden a una señal física global: la temperatura intracelular, que es un parámetro que debe ser sentido constantemente. A diferencia de los riboswtiches clásicos, los termoriboswtiches no se caracterizan por un dominio de unión a ligando altamente conservado, ya que no existe ninguna molécula *per se* que reconocer (Narberhaus *et al.*, 2006). En su lugar, la funcionalidad de estos termo-riboswtiches está basada en cambios conformacionales que transducen la señal térmica al aparato de expresión genética. (Chowdhury *et al.*, 2006). Dichos cambios controlan el inicio de la traducción enmascarando la secuencia Shine-Dalgarno y algunas veces al codón de inicio AUG, de manera que estos sitios realizan interacciones del tipo Watson-Crick a bajas temperaturas. Al incrementar la temperatura, algunos elementos de la estructura secundaria se funden localmente, permitiendo la exposición del RBS, de tal modo que el ribosoma puede unirse a este sitio y traducir el mRNA (Serganov y Patel, 2007). Dicho mecanismo presenta variantes dependiendo de los procesos de regulación particulares de cada termómetro de RNA que ha sido descrito, teniendo en común que todos ellos se encuentran en *cis* en la región 5' no-traducida del mRNA, regulando la expresión genética río abajo, actúan a nivel traduccional modulando la eficiencia de dicho proceso (Figura 5).

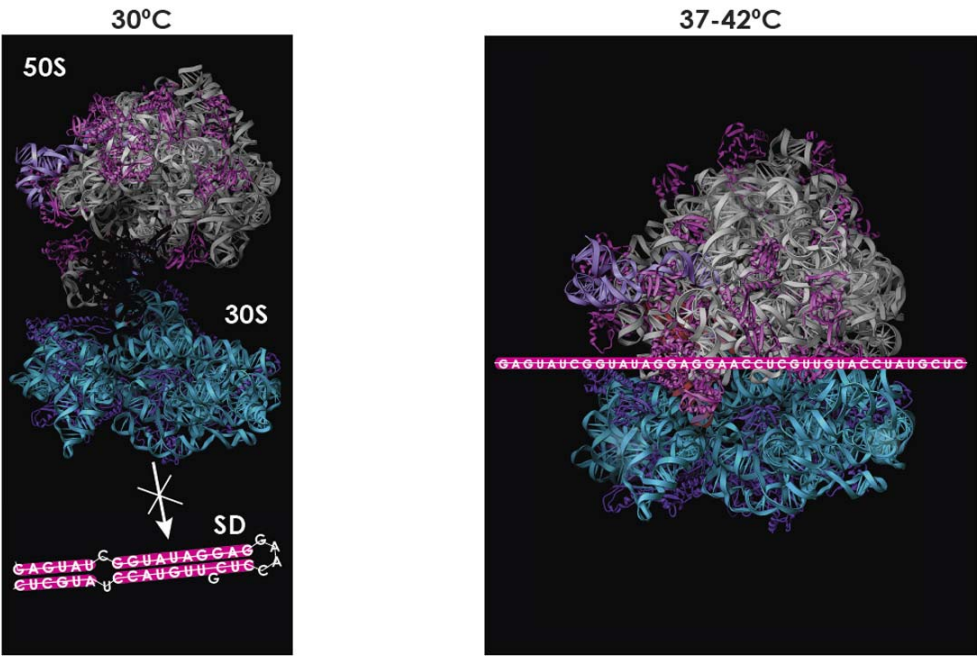


Figura 5. Mecanismo general de la regulación genética por termoriboswtiches. La secuencia SD y en algunas ocasiones el codón de inicio AUG se encuentran apareados a temperaturas bajas. Durante el choque térmico esta estructura se funde y se abre, permitiendo la unión del ribosoma a la secuencia SD y la subsiguiente traducción.

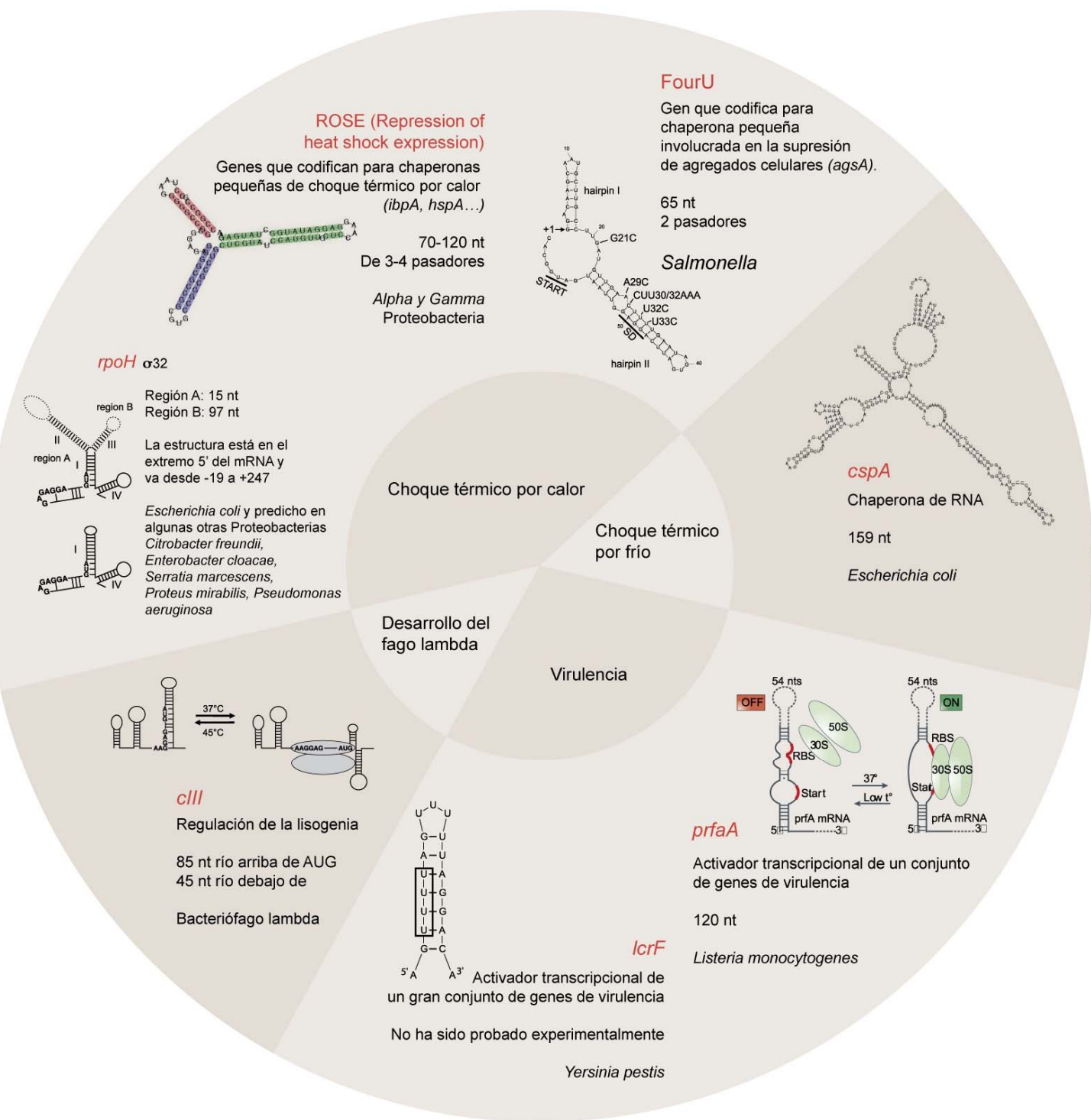


Figura 6. Termoriboswitches que han sido descritos hasta la fecha. Los termoriboswitches se encuentran en genes relacionados con el choque térmico por calor, por frío, virulencia y en el desarrollo del fago lambda. En bacterias patógenas, de vida libre, simbiotes, etc. La información del mecanismo particular así como sus características detalladas se describen a lo largo del texto.

Un grupo de genes que parecen ser candidatos a ser controlados por la temperatura son aquellos relacionados con los choques térmicos por calor o frío. Otro grupo puede ser los genes de virulencia ya que en bacterias patógenas que viven en el agua, el suelo o alimentos, experimentan una variación en la temperatura ambiental al entrar al organismo hospedero. La variación en la temperatura puede no ser lo suficientemente severa como para inducir la respuesta de choque térmico, pero si es importante como señal para la invasión, colonización y la inducción de genes de virulencia. Dichos procesos son regulados en parte por termómetros de RNA. Un caso menos intuitivo sería involucrar la regulación del fago lambda durante la infección de *E. coli* que también está modulada por un termómetro de RNA, y sin embargo fue el primer termómetro descrito hace aproximadamente 15 años (Altuvia, 1989).

1.6.1. Riboswitches del gen *rpoH* regulan la traducción de σ_{32} en respuesta a cambios de temperatura.

Como se mencionó en la sección 1.3, la transcripción del gen *rpoH* que codifica a σ_{32} es regulada mediante un circuito de retroalimentación en respuesta a la acumulación celular de proteínas desnaturalizadas. En adición a esta regulación, la expresión de *rpoH* está modulada por dos segmentos intragénicos en la secuencia codificante denominados región A y región B (Morita et al., 1999). A través de fusiones entre *rpoH* y *lacZ* se determinó que los niveles celulares elevados de σ_{32} en la célula a altas temperaturas dependen de la presencia de un termómetro de RNA, que consiste en una estructura secundaria que se forma entre las regiones A y B, bloqueando el sitio de inicio de la traducción y parte del RBS a temperaturas bajas (Storz, 1999). Es destacable el hecho de que este termómetro, a diferencia de la mayoría de los casos que han sido descritos, forme la estructura secundaria más allá de la región río arriba e incluya un segmento de la región codificante (Figura 6).

Parte relevante del estudio de este termómetro es que a través de mutagénesis se demostró que la estructura secundaria del mRNA, más que la secuencia de nucleótidos, es determinante para la termorregulación: Mutaciones puntuales de ciertos nucleótidos comprometían a *rpoH* aumentando la expresión de la fusión reportera, mientras que las mutaciones compensatorias que mantenían el apareamiento de las bases mediante interacciones Watson-Crick, reestablecían una expresión normal (El Samad et al., 2005; Narberhaus et al., 2006).

1.6.2. Un termómetro que se activa con el choque térmico por frío.

Otro de los primeros estudios que revelaban a la estructura secundaria del mRNA como un termosensor se relacionaba con la traducción de un gen involucrado en la producción de

proteínas de choque térmico por frío (Cold Shock Proteins) como mecanismo de protección celular. En *E. coli*, la principal proteína de choque térmico es CspA y su inducción aumenta de manera importante a bajas temperaturas. Esta proteína se une al RNA, sin presentar aparentemente especificidad por alguna secuencia y se piensa que es una chaperona de RNA que facilita la traducción a temperaturas bajas, no permitiendo la formación de estructuras secundarias (Phadtare, 2004). La expresión de *cspA* es regulada a distintos niveles, entre ellos la transcripción, la antiterminación, la estabilidad del mRNA y la traducción.

Una característica importante del mRNA de *cspA* es la región intergénica no-traducida 5' que consiste en 159 pares de bases y que ha sido postulada como un termómetro de RNA en frío. Yamanaka *et al* (1999) llevaron a cabo un experimento en el cual sustraían la región intergénica de este gen, y observaron que la expresión no correlacionaba con la cantidad de transcrito, sugiriendo que la secuencia líder controla la eficiencia de traducción. Se ha postulado que estructuras secundarias alternativas en el RNA en un rango de 37° a 15°C modulan la accesibilidad del sitio de unión al ribosoma (SD). La estabilidad del *cspA* mRNA puede estar controlada también por estructuras de RNA alternativas, dada la rápida degradación de este transcrito a través de la RNasa E la cual ocurre a temperaturas altas cuando el SD se encuentra desenmascarado y accesible (Yamanaka, 1999) (Figura 6).

El mecanismo del termómetro en frío de *cspA* es similar al descrito para el mRNA *cIII*, ya que se forma una estructura secundaria a bajas temperaturas que permite la unión del ribosoma y la traducción. Existe una región río debajo de la secuencia SD, que es llamada la caja “upstream”, que es importante para reclutar al ribosoma (Xia *et al.*, 2002). No se ha determinado si esta región contribuye directamente a la formación de una estructura secundaria que intervenga con el apareamiento del SD.

1.6.3. Termómetros de RNA que juegan un papel importante en la virulencia.

En organismos patógenos, muchos son los genes de virulencia que están bajo el control de la temperatura y su expresión depende una gran diversidad de mecanismos moleculares como son cambios en el super-enrollamiento del DNA ((Hurme *et al.*, 1998), en proteínas sensoras (Newman *et al.*, 2002), o bien en termómetros de RNA. A continuación se describen ejemplos de este último tipo de elementos regulatorios basados en moléculas de RNA.

1.6.3.1. ¿Lisis o lisogenia?: El desarrollo del fago lambda.

Los primeros estudios que demostraban evidencia de regulación genética mediada por la temperatura, vía mRNA al estructurarse este, fueron realizados en el fago lambda (Figura 6). Altuvia y colaboradores (1989) observaron que la formación de estructuras secundarias en el

mRNA determinaba la tasa de traducción del gen *cIII* del fago, de una manera dependiente de la temperatura. El producto del gen *cIII* está involucrado en el proceso de “decisión” en el cual el fago puede entrar al ciclo lítico o al ciclo lisogénico. *cIII* sirve como un sustrato alternativo de degradación de la proteína FtsH (la cual degrada principalmente a *cII*, un regulador central en el ciclo lisogénico). Cuando la proteína *cIII* se encuentra en altas concentraciones, *cII* se estabiliza promoviéndose el ciclo lisogénico. Se predijeron y se comprobó experimentalmente la existencia de dos estructuras alternativas en el mRNA de *cIII* que respondían de manera diferencial al calor. A temperatura fisiológica (37°C) se presenta una estructura extendida que cubre 84 nucleótidos no traducidos y los primeros 41 nucleótidos codificantes, dejando accesibles las regiones que comprenden al SD y al codón de inicio AUG al ribosoma; en esta conformación hay producción de la proteína *cIII*, favoreciéndose el ciclo lisogénico. Al aumentar la temperatura a 45°C el equilibrio se desplaza hacia la otra estructura excluyente más estable, en donde el SD y el codón de inicio se encuentran ocultos e inaccesibles en una estructura de “pasador”. El bloqueo de la traducción del competidor *cIII* resulta en el decremento de la concentración de *cII*, ya que esta proteína se degrada al ser el sustrato de la proteasa FtsH iniciándose el ciclo lítico (Altuvia et al., 1989). La “decisión” del fago lambda de presentar lisogenia o lisis está controlada por una red transcripcional compleja y por mecanismos postraduccionales en respuesta a distintas señales. El fago tenderá a proliferar y vivir de manera óptima sólo si las células hospederas están sanas y se integran al cromosoma cuando las condiciones de crecimiento son pobres (Narberhaus et al., 2006). Sin embargo, cuando existe daño por radiación UV o por choque térmico severo, resulta mucho más benéfico para el fago escapar del hospedero.

Este termómetro de RNA como se verá más adelante, es el único que enciende o activa la traducción a una temperatura menor a la del choque térmico. Otra de sus características excepcionales es que al parecer no opera de forma gradual al disminuir la temperatura, sino que es un verdadero “switch” encendido/apagado basado en dos conformaciones mutuamente excluyentes.

1.6.3.2. Regulación de VirF por elementos de RNA en *Yersinia pestis*.

Un segundo ejemplo de un termómetro de mRNA involucrado en virulencia está involucrado en la regulación de LcrF (VirF) en *Yersinia pestis*. Este activador transcripcional regula a un gran conjunto de genes de virulencia y aumenta su concentración celular a medida que la temperatura intracelular también lo hace. La eficiencia de traducción de *lcrF* es más alta a 37°C que a 26°C, mientras que la temperatura no tiene efecto sobre su transcripción (Hoe y Goguen, 1993). Dado lo anterior se propuso como mecanismo de regulación de este gen, un posible termómetro de RNA cuya estructura secundaria en el mRNA de *lcrF* fuera capaz de secuestrar al sitio de unión SD. De hecho la desestabilización térmica de la estructura tallo-asa se predice que ocurre a

temperaturas fisiológicas entre los 33° y 37°C. Este termómetro aun no ha sido probado experimentalmente (Figura 6).

1.6.3.3. Regulación de PrfA por elementos de RNA en *Listeria monocytogenes*.

Un mecanismo análogo al anteriormente descrito en *Y. pestis* se ha visto que controla la expresión de genes de virulencia en la bacteria patógena *Listeria monocytogenes*, en la cual el activador de dicho sistema es PrfA. El gen *prfA* se transcribe a 30°C y 37°C, sin embargo únicamente se traduce en la última temperatura. Cossart *et al.*, (2002) predijeron la formación de una estructura tallo-asa en la secuencia líder de dicho gen, que comprendía al menos 120 nt y el sitio de unión del ribosoma. Este termómetro fue probado experimentalmente, realizando mutaciones que lograron desestabilizarlo, reduciendo o eliminando la termorregulación de *prfA*. La formación de la estructura en este mRNA se demostró por ensayos de movilidad en gel, los cuales mostraban que la estructura de tallo que se formaba alrededor de la secuencia SD se abre a 37°C, lo cual no ocurre a 30°C. Finalmente, este grupo observó por expresión heteróloga, que la región líder de *prfA* puede conferirle a *E. coli* la capacidad de regulación térmica (Cossart, 2002). Este termómetro de RNA consiste de una sola estructura tallo-asa (Figura 6).

1.6.3.4. FourU un termómetro más en *Salmonella enterica*.

Recientemente se ha descrito un nuevo termómetro en *Salmonella enterica* que se encuentra en la región UTR 5' del gene que codifica para la HSP AgsA (aggregation-suppression protein) es única en su especie ya que, en lugar de unirse a proteínas desnaturalizadas, se une a la región líder del mRNA de *agsA* regulando su traducción. Este sitio de unión constituye un termómetro de RNA llamado FourU debido a que son cuatro nucleótidos consecutivos de uracilo los que se aparean con la región del SD de *agsA*. Este termómetro difiere respecto a ROSE en que únicamente consiste de 2 estructuras tallo y asa, y en el pasador II se halla el SD y el anti-SD. (Ver sección 2. ROSE: el termómetro de RNA como modelo de estudio). De manera similar a los estudios hechos en *rpoH* y ROSE, las mutaciones en la región del SD afectan la traducción *in vivo*. Los cambios en la conformación de FourU son dependientes de la temperatura, lo que controla el acceso del ribosoma al SD (Figura 6) (Waldminghaus *et al.*, 2007b). En adición a esta regulación traduccional, la transcripción de *agsA* está controlada por σ^{32} .

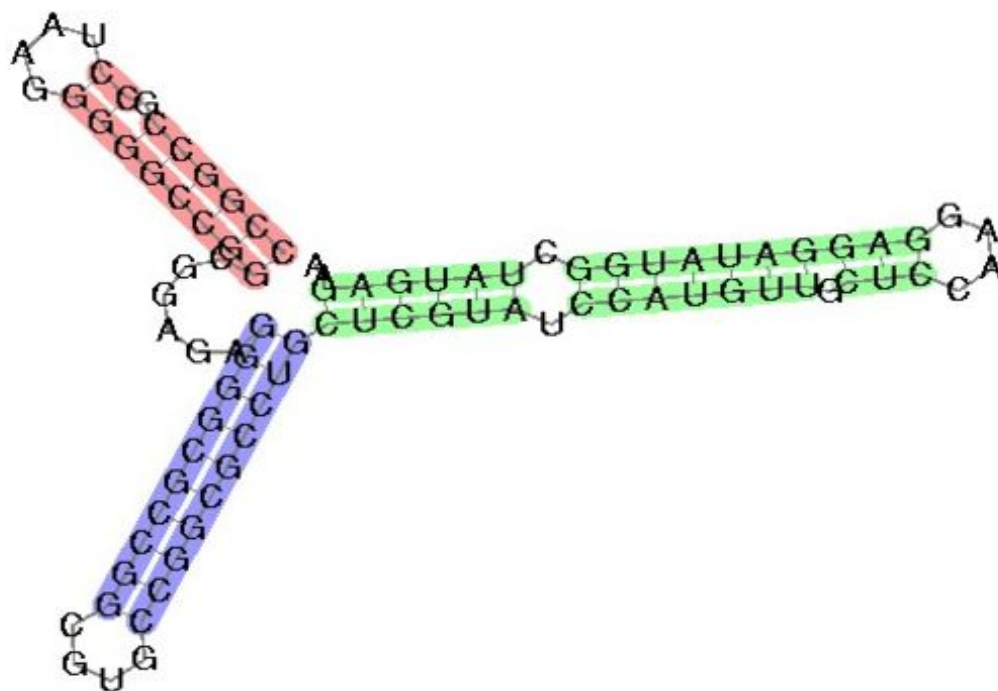
Como se mencionó anteriormente, una estructura similar compuesta de cuatro residuos de uracilo que se aparean con la secuencia SD (AGGA) se predijo en el caso de la región río arriba del gen *lcrF* en *Yersinia pestis*. La verificación experimental de la funcionalidad de este riboswitch y el de FourU indicaría el descubrimiento de nuevos termómetros en donde no necesariamente se requiera de la estructura complejas de RNA como las existentes en ROSE, sino únicamente del apareamiento del SD.

2. Represión de la expresión de los genes de choque térmico por ROSE: el termómetro de RNA modelo.

El termómetro de RNA más conocido y representado en una amplia variedad de organismos es ROSE; fue descrito por primera vez en *Bradyrhizobium japonicum*, un simbiote fijador de nitrógeno en la soya (Nocker *et al.*, 2001). ROSE se encuentra en la región intergénica de genes de choque térmico que codifican para chaperonas pequeñas de tipo α cristalinas de la clase A, cuya distribución se restringe a α y γ Proteobacterias. Se han predicho al menos 40 elementos ROSE distribuidos en 21 especies bacterianas. La resolución de la estructura secundaria *in silico* indican que este termómetro se compone de 3 a 4 “pasadores” (70 a 120 nucleótidos), en donde el cuarto contiene la secuencia SD y el codón de inicio AUG, incluyendo un residuo G sobresaliente, enmascarados por el apareamiento imperfecto entre nucleótidos con la región próxima 3'. La secuencia funcional del pasador IV se encuentra altamente conservada y el motivo consenso es U (U/C)GCU (6-9)AGGA; mientras que el resto de la secuencia que forma los otros pasadores es variable (Figura 7) (Waldminghaus *et al.*, 2005).

A través de mutaciones en ROSE, Nocker *et al.*, demostraron su potencial como termosensor. Usando una construcción de ROSE-HSP fusionado a nivel traduccional a *lacZ*, se observó que su expresión era bastante baja a 30°C y que el crecimiento celular a 37°C incrementa la expresión del gen reportero; la sustitución de bases en la región del SD y el codón de inicio incrementa la expresión de la fusión a 30°C. Como es de suponerse, las mutaciones compensatorias que permiten que las interacciones tipo Watson-Crick sigan existiendo entre un par de nucleótidos, restauran la represión y las mutaciones en regiones desapareadas son neutrales, de tal modo que la estructura secundaria más allá que la secuencia es estrictamente necesaria para monitorear la temperatura correctamente (Nocker *et al.*, 2001).

Estudios de resonancia magnética nuclear demuestran que lo primero en ocurrir es la desestabilización a los 37°C de la región adyacente a la secuencia SD, haciéndola accesible parcialmente a la subunidad ribosomal 30S, y a medida que se incrementa la temperatura la interacción mRNA-30S es favorecida por la liberación completa del SD y el codón de inicio AUG, generando el acceso del tRNA formil-metionina al sitio A de la subunidad ribosomal 30S. El pasador funcional de ROSE se caracteriza por presentar una estructura helicoidal con una serie de apareamientos entre las bases no canónicas, lo cual aunado a una red de puentes de hidrógeno facilitan la liberación del SD en un rango de temperatura fisiológica (Chowdhury *et al.*, 2006).



Bra.jap.	CGCCGC	CGCAGG	CGGC	UCU.	CCGGGCGCC.	UGAC.	GGGUCGG	CGAAUCCAGAGAC	GGCACC	GG.	UCGUGU.	CCG
Bra.jap.	CGCCGC.	GACAAG	CGGU.	...	CCGGGCGCC.	CUAG.	GGGCCCGGCG.	...	GAGACGGGCGCGCGGA.	...	GGUG.	UCCG
Bra.jap.	UGCCGC.	GACAAG	CGGU.	...	CCGGGCGCC.	UUCG.	GGGGCCCGGCG.	...	GAGACGGGCGCGCGGA.	...	GGUG.	UCCG
Bra.jap.	GCCACG	CGGGAG	CGUGG	CG.	CCGGAUGCC.	UCUC.	GGGUCCGG.	...	CAGACG.	CAUCGGG.	CAGAA.	UCCG
Bra.jap.	GCCGCG.	GA.	GAGGUGG	GUC	CCGACGCC.	UCAU.	GGGUCGGGAACGAC.	UGAGAC	GGCACC	GG.	UCGUGU.	CCG
Mes.lot.	...	...	...	...	UCGGGUCGC.	GCAUAA	GGGGCCGAUGU.	...	GUCAGGGAGCGCCAU.	...	CUU.	CUUG
Mes.lot.	...	...	...	...	UCGGG.GC.	UUCG.	G.CCCGACA.	...	GAUUGGGCGUCGCAG.	AGACG.	CUUG	...
Agr.tum.	...	...	...	...	CCAGCCGCU	GAGGUC	AGGGGCGUGGCC.	...	UUGUUGGACG.	...	CGC.	CGU
Agr.tum.	...	...	...	...	CCAGCCGCU.	ACGAC	GGGGCGUGGACU.	...	UGCUGGGAGCGCCG.	...	CCUUU.	CG
Rhi.sp..	...	...	...	...	CCGACCGCCA.	AAA.	UGGCUGAGCCUG.	...	GUCAGGGAGCGCCGAG.	...	CUU.	CUUG
Rhi.sp..	...	...	...	...	GCCGAUGCC.	AAUU.	GGGUCGGCAUG.	...	GUCAGGGAGCGCCACG.	...	CUU.	CUUG
Sin.mel.	...	...	...	...	AGAGCCGCCU.	GUA.	AGGGGCGUCGCA.	...	GUCGAGGAG.	...	CUCCGUU.	CU
Glu.oxy.	...	...	...	...	CCGGCCGCC.	AAA.	GGGGUCGGCA.	...	AAAGAGAAU.	...	AUC.	AG
	<<<<<<	...	>>>>>>	...	<<<<<<	<<<	>>>>>>>>>>	...	<<<<<<<<<<<<	...	>>>>	

Bra.jap.	GUGCC.G...	CUUGUA.	ACCAUUUG	CUUG	CGUG	GAGGAUC	UGGCUAUG	CGCA	Prev
Bra.jap.	ACGCCUG...	CUCGUA.	CCCAUCUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Bra.jap.	ACGCCUG...	CUCGUA.	CCCAUCUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Bra.jap.	AUGU.UG...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Bra.jap.	GUGCC.G...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Mes.lot.	GCGUUC...	CUCGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Mes.lot.	GACGCUG...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Agr.tum.	CUUUCAGCGCU.	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Agr.tum.	GCGCUUC...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Rhi.sp..	GCGCUUC...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Rhi.sp..	GCGCUUC...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Sin.mel.	CUUCGGGCGCUC	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
Glu.oxy.	UUUUGAU...	CUUGUA.	UCCAUAUUG	CUUG	CGUG	GAGGAUUUGG	CUAUGAGGA		Prev
	>>>>>>	<<<<<<	<<<<<<	<<<	>>>>>>>>>>	>>>>>>			Prev

Figura 7. Alineamiento múltiple y estructura secundaria del elemento ROSE presente en α Proteobacteria. El alineamiento muestra las regiones conservadas de ROSE y cada color corresponde a un pasador en la estructura (4 pasadores en α Proteobacteria), la región naranja corresponde al pasador en donde se encuentra la región SD apareada y el codón de inicio AUG, el equivalente a la región señalada en verde en la estructura secundaria predicha, se observa que esta región es la que presenta mayor conservación debido a su capacidad de aparear al SD a través de una estructura complementaria en donde se encuentra el residuo altamente conservado de G sobresaliente. Asimismo los otros pasadores en la estructura secundaria son necesarios para mantener la función y la estabilidad de dicho elemento (ver texto). Tomado de (<http://www.sanger.ac.uk/cgi-bin/Rfam>)

Como se ha visto, existen pocos casos bien descritos de termómetros de RNA en bacterias, y la mayoría de ellos se presentan como casos poco conservados, encontrándose en una sola especie bacteriana, controlando algún mecanismo muy particular para la adaptación y sobrevivencia de la misma. Caso destacable comprende al elemento ROSE distribuido en un amplio grupo de proteobacterias.

Para caracterizar a ROSE en α y γ Proteobacteria (Waldminghaus *et al.*, 2005) fue necesaria una búsqueda bioinformática previa, basada en estructura de aproximadamente 120 genomas de arqueobacterias y bacterias, específicamente en los genes de HSPs pequeñas. No se consideraron los distintos serotipos de una misma especie para una misma proteína, asimismo se observó que algunas bacterias no presentan en el genoma HSPs pequeñas. De tal modo que se tomaron 110 nucleótidos río arriba del codón de inicio anotado de los genes de choque térmico, y se calculó la estructura utilizando el programa mfold (Zuker, 2003). Se encontró un total de 27 estructuras tipo ROSE provenientes de 18 genomas en otras α Proteobacteria, lo cual indicaba que este elemento no estaba restringido a Rhizobiaceae y además se encontraba en miembros γ Proteobacteria: enterobacterias y otras bacterias representativas como *Vibrio* y *Pseudomonas*.

Por otra parte en la revisión hecha por Narberhaus *et al.* (2006) se propone que el hecho de que termómetros tipo ROSE solamente se encuentran en α y γ Proteobacteria no excluye la posibilidad de encontrarlos en otros organismos que hagan uso de termómetros basados en RNA con una secuencia y estructura diferentes a las descritas. Lo cual, en última instancia, nos sugiere que es posible el descubrimiento de nuevos termómetros en una amplia diversidad filogenética de procariontes, arqueobacterias y probablemente en algunos eucariontes.

En contraste con los riboswitches, todos los casos de termómetros de RNA conocidos son únicos o se encuentran limitados a genes ortólogos o parálogos; por ejemplo, los elementos ROSE se presentan río arriba de genes que codifican para proteínas pequeñas de choque térmico. Las comparaciones filogenéticas han revelado que la conservación de los termómetros se da en términos de estructura más allá de la secuencia (Waldminghaus *et al.*, 2007a). En cuanto a aspectos estructurales en ROSE, el tallo-asa que en sentido estricto contiene el termómetro del RNA se caracteriza por algunos residuos altamente conservados, incluyendo la región SD y la secuencia complementaria que lo bloquea a bajas temperaturas (Figura 8).

El corte de ROSE manteniendo únicamente el tallo-asa IV no suprime la represión a 30°C pero sí reduce su inducción por calor *in vivo* (Chowdhury *et al.*, 2003). Aquellas mutaciones que

desestabilizan la parte superior del tallo-asa III aumentan la expresión a bajas temperaturas, y las mutaciones compensatorias en la misma región reestablecen la represión. De tal modo que para el monitoreo adecuado se requiere de la estructura completa e intacta del mRNA ROSE.

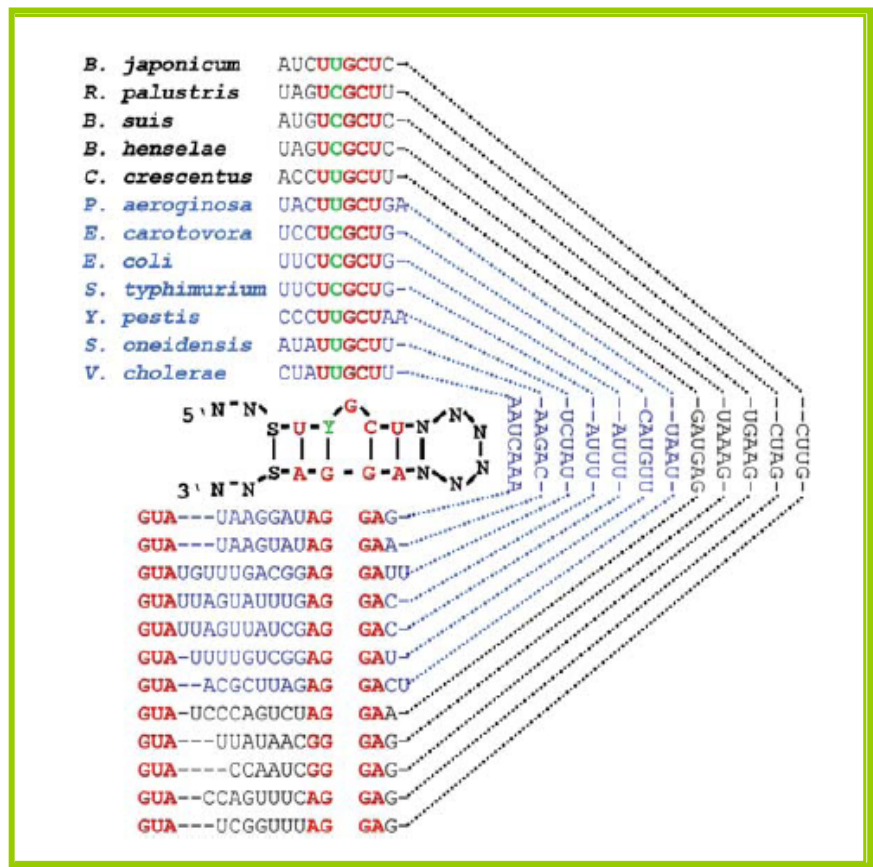


Figura 8. Secuencia y estructura del pasador en algunas secuencias con el termómetro de RNA ROSE. Secuencias representativas en Alpha y Gamma Proteobacteria. La secuencia consenso es la que se encuentra en el centro (N=G, C, A, U; S= G ó C; Y= C ó U). Se observa el residuo G sobresaliente, siendo crucial para el termo-monitoreo. Tomado de (54).

Una característica muy importante en la estructura de ROSE es el hecho de la presencia del residuo G sobresaliente que se encuentra en la región de apareamiento de la secuencia SD. Dicho residuo se encuentra altamente conservado y Nocker *et al.*, (2001) encontraron que la eliminación de este reprime la traducción a altas y bajas temperatura, formándose un pasador más estable que no se abre dentro de un rango de temperatura fisiológica, lo que finalmente compromete la termosensibilidad de ROSE.

El hecho de que el residuo G sobresaliente sea un parámetro de relevancia para la capacidad de respuesta a la temperatura se explica por el apareamiento imperfecto entre las bases nucleotídicas garantizando la fusión dentro de un rango de temperatura apropiado. Este rango es estrecho, ya que la inducción de las proteínas de choque térmico se da de 30°C-37°C; por tanto el residuo G puede facilitar la fusión mediada por temperatura en el mRNA ROSE (Figura 8).

Interesantemente al sustituir dicho residuo por un residuo A no altera la expresión y su termosensibilidad se mantiene de manera muy parecida a ROSE silvestre (Nocker *et al.*, 2001). Los experimentos realizados para demostrar lo anterior consisten en fusiones entre ROSE y el reportero *lacZ*: monitoreando la actividad β -galactosidasa, no se observó la actividad del gene reportero a temperaturas altas entre los 40 o 42°C, debido a que la enzima es inestable bajo esas condiciones, por lo que es posible montar otros sistemas más eficientes que permitan monitorear la expresión a temperaturas mayores (Figura 9).

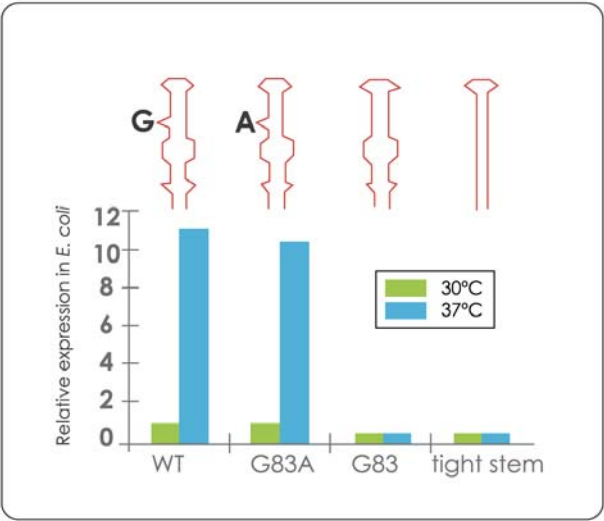


Figura 9. Expresión de las fusiones variantes ROSE-*lacZ*. La construcción mutante en la que se cambia el residuo G sobresaliente por uno A, es capaz de reestablecer la función de termosensor de ROSE, la expresión genética de una mutante en cuyo pasador no se encuentra ningún residuo abate la expresión de *lacZ* en ambas temperaturas, al igual que la construcción mutante de ROSE no estructurado en donde el pasador es estrecho. Tomado de (42).

Excepto por la secuencia SD y el codón de inicio AUG, existen pocas características distintivas para la búsqueda en general de sensores de RNA. Por ejemplo los mRNAs responsivos a vitaminas que se conocen hasta ahora presentan secuencias consenso (B_{12} box, *thi* box, elemento RFN) pero se desconocen sitios de reconocimiento para otras señales. Particularmente, el descubrimiento de nuevos termosensores de RNA mediante el análisis de secuencias puede ser complicado por el hecho de que el apareamiento entre bases con el sitio de unión al ribosoma puede estar ubicado a lo lejos en secuencias dispersas río arriba o dentro de la región codificante (Narberhaus, 2002). De tal modo, se requiere de una inspección exhaustiva en todos los genes conocidos de choque térmico, de virulencia, así como de factores sigma, extendiendo ésta a las secuencias de genomas completos.

Justificación

Los sensores de RNA que hasta ahora se conocen han sido descubiertos recientemente en un tiempo relativamente corto, demostrando la potencialidad de regulación basados en este tipo de moléculas. Dada la descripción de nuevos mecanismos moleculares presentes durante el estrés térmico, y en particular aquellos que explotan dicho potencial, es posible realizar el análisis comparativo entre las distintas alternativas que los organismos han desarrollado, a través de la compilación de la información que ha sido generada recientemente.

El conocimiento de estructuras de mRNA implicadas en mecanismos de regulación genética en respuesta a la variedad de condiciones ambientales en las que se desarrollan las bacterias, es el primer paso para comprender la adaptación y la selección de estos mecanismos a nivel evolutivo. La estructura secundaria en el caso de ROSE está conservada en varios organismos que viven en condiciones de vida muy distintas, siendo sorprendente encontrarlos en organismos de vida libre, bacterias rizobiales, patógenas, extremófilas, entre otras. Por todo lo anterior, es importante la caracterización funcional de nuevos termómetros en organismos como las arqueobacterias que presentan una diversidad en cuanto a la explosión de los lugares en los cuales pueden habitar; dilucidando cierta conservación a nivel de secuencia y quizá lo más importante la conservación de la funcionalidad a nivel de estructura secundaria.

Considerando la prevalencia del termorregulador ROSE y la falta de evidencia que explique la conservación estructural, se hace patente la necesidad de análisis teóricos y experimentales que aborden el estudio de dicho termómetro. Evolutivamente parece ser importante el hecho de que ROSE se encuentre conservado en todas las bacterias α y γ Proteobacteria. Esto a pesar de que un residuo de adenina tendría un efecto similar al residuo G sobresaliente sobre la expresión genética, siendo así ¿Por qué no se ha seleccionado otro residuo? ¿Por qué el residuo de G sobresaliente es tan relevante en este contexto? Los experimentos que demuestran la relevancia del residuo G sobresaliente así como la sustitución por adenina fueron ensayos de la actividad de β -galactosidasa, teniendo como reportero *lacZ* (Nocker *et al.*, 2001) como se había mencionado anteriormente, lo cual no demuestra lo que ocurre con las siguientes generaciones bacterianas al transcurrir el tiempo, por lo que en lo inmediato es posible sustituir los residuos, ignorando lo que ocurre en las siguientes generaciones bacterianas: ¿Se gana y se mantiene ese residuo en las siguientes poblaciones? ¿Se pierde y se reestablece la condición silvestre? ¿Cómo sucedió entonces la evolución de este elemento? ¿Es posible a partir de ROSE obtener variantes sintéticas que termo-respondan de manera similar a la forma silvestre? Dado lo anterior, nuestro interés es encontrar nuevas variantes funcionales de ROSE sintéticas que termo-respondan de manera similar a la forma silvestre; esto a través de una mutagénesis con oligonucleótidos

degenerados diseñados en la estructura ROSE. Esperando tantas variantes como mutaciones existen por nucleótido es necesario desarrollar un método eficiente que nos permita seleccionar aquellas variantes que funcionen de manera similar al termómetro silvestre.

Hipótesis

- Termómetros de RNA no descritos anteriormente estarán presentes y ampliamente distribuidos tanto en bacterias como en arqueobacterias; y por su función tenderán a regular a un grupo de genes ortólogos (proteínas de choque térmico, factores de virulencia entre otros relacionados, en los cuales no necesariamente se conozca la función del gen)
- Mutaciones al azar en el elemento ROSE silvestre generarán variantes sintéticas similares en función al mismo elemento: 1) a temperaturas bajas formarán una estructura de apareamiento del SD, y 2) al inducir el choque térmico esta estructura se abrirá desenmascarando el SD y por lo tanto se dé la traducción del (os) gen (es) río abajo.
- El método de selección mediante una construcción genética será efectivo para discriminar y reducir de manera significativa aquellas mutantes que no se comportan como ROSE silvestre (siempre abiertas o siempre cerradas)

Objetivos

- Llevar a cabo de manera general una revisión sobre el mecanismo de regulación durante el estrés por choque térmico.
- Elaborar una búsqueda bioinformática de elementos reguladores –específicamente termómetros de RNA- con los genomas contenidos en la base de datos del National Center for Biotechnology Information (NCBI)
- Redescubrir termómetros de RNA conocidos (en particular ROSE) para probar la efectividad del método.
- Descubrir nuevos candidatos de termómetros de RNA para dirigir posteriores análisis computacionales y experimentales.
- Realizar un análisis general acerca de los elementos obtenidos con mayor conservación en las regiones intergénicas regulatorias, sean o no regiones tipo termómetro de RNA.
- Encontrar nuevas variantes funcionales de ROSE sintéticas, a través de una mutagénesis al azar con oligonucleótidos degenerados de la estructura ROSE.
- Desarrollar un método eficiente que permita seleccionar a las variantes que termomontoreen de manera similar al termómetro ROSE silvestre y que no han sido seleccionadas naturalmente.
- Construir un sistema genético doble (selectivo/contra-selectivo) para evaluar la expresión del (os) gen (es) reportero(s) de las mutantes funcionales.

- Sintetizar y clonar el termómetro ROSE silvestre de *E. coli* K12 en el sistema genético, para probar la efectividad de la construcción.
- Realizar ensayos de crecimiento a diferentes temperaturas con las clonas mutantes para seleccionarlas de acuerdo a los criterios anteriormente descritos.

1.1. Metodología y Desarrollo Bioinformático

Básicamente el procedimiento consistió en obtener grupos de regiones intergénicas ortólogas de genes asociados al choque térmico, así como factores sigma, sobre grupos de organismos bacterianos y en arqueobacterias. A los conjuntos de secuencias se les aplicó un algoritmo de descubrimiento de motivos llamado MEME (por sus siglas en inglés: Multiple Expectation maximization for Motif Elucidation) (Bailey y Elkan, 1994). Los motivos identificados por MEME son expresados en términos de matrices de probabilidad. Dichas matrices son utilizadas por el programa contraparte de MEME llamado MAST (por sus siglas en inglés: Motive Alignmemt and Search Tool) (Bailey y Gribskov, 1998), para identificar regiones intergénicas que contengan los elementos de regulación en todos los genomas que se encuentran completamente secuenciados. El anterior proceso tiene como objetivo identificar al conjunto original de secuencias y adicionalmente nuevas secuencias intergénicas que pudieran presentar el motivo o motivos sobre-representados. Con este nuevo conjunto extendido de secuencias, el proceso de identificar motivos conservados utilizando MEME se repite. El resultado de rondas sucesivas de definición de motivos seguido de la identificación de secuencias intergénicas con dicho motivo(s), tiende a obtener matrices de probabilidad “refinadas” que tienden a tener mejor representatividad que la matriz construida en el primer paso del estudio, dejando atrás la limitante de los grupos ortólogos iniciales. Así se encuentra una serie de motivos que pueden ser elementos conservados regulatorios aunado a una lista de los posibles genes regulados por los anteriores (Abreu-Goodger, 2005). El análisis se centra en aquellos motivos que se encuentran cercanos a la región SD o aquellos que estructuralmente presentan un tallo-asa, en donde dicha región está apareada (Figura 10). De manera general para la obtención de secuencias a analizar y comparar se generó un programa en el lenguaje de programación PERL: “0_extrae_iMUR_COG_filogenia.pl” (Ver Apéndice 1).

1.1.1. Grupos de genes ortólogos (COGs) asociados al choque térmico por calor y factores sigma.

La búsqueda de elementos sensores de temperatura basados en RNA en los distintos genomas es posible, suponiendo que algunos genes ortólogos en genomas lejanos son regulados por un elemento homólogo común a todos ellos o al menos a una fracción significativa de los mismos. Para encontrar a estos elementos de regulación se buscaron patrones y/o motivos conservados en las regiones de regulación de los genes ortólogos. Con tal motivo, se empleó la base de datos COG (Clusters of Orthologous Groups of proteins) que consta de 4,873 grupos de familia de genes ortólogos, de la cual se extrajeron aquellos COGs relacionados con el choque térmico y factores sigma, a través de su búsqueda por palabras clave: “shock” y “sigma”. De esta

búsqueda se seleccionaron 13 COGs asociados al choque térmico y 19 a factores sigma (Tabla 1 y Tabla 2).

COGs relacionados al choque térmico (palabra clave: shock)	
COG0071	Molecular chaperone. Small heat shock protein. Posttranslational modification, protein turnover, chaperones.
COG0234	Co-chaperonin GroES (HSP10). Posttranslational modification, protein turnover, chaperones.
COG0326	Molecular chaperone (HSP90 Family). Posttranslational modification, protein turnover, chaperones.
COG0443	Molecular chaperone. Posttranslational modification, protein turnover, chaperones.
COG0466	Class III heat shock ATP dependent Lon-protease, bacterial type. Posttranslational modification, protein turnover, chaperones.
COG0459	Chaperonin GroEL (HSP 60 Family). Posttranslational modification, protein turnover, chaperones.
COG1219	ATP-dependent protease Clp, ATPase subunit. Posttranslational modification, protein turnover, chaperones.
COG1278	Cold shock proteins. Transcription.
COG1281	Disulfide bond chaperones of the HSP33 family. Posttranslational modification, protein turnover, chaperones.
COG0576	Molecular chaperone GrpE (heat shock protein). Posttranslational modification, protein turnover, chaperones.
COG1842	Phage shock protein A (IM30), suppresses sigma54-dependent transcription.
COG1420	Transcriptional regulator of heat shock gene. Transcription.
COG0789	Involved in transcriptional regulation (repression) of heat shock proteins. Transcription.

COGs relacionados con factores sigma (palabra clave: sigma)	
COG0568	DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32).
COG1191	DNA-directed RNA polymerase specialized sigma subunit.
COG1366	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor).
COG1508	DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog.
COG1595	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog.
COG1842	Phage shock protein A (IM30), suppresses sigma54-dependent transcription.
COG2172	Anti-sigma regulatory factor (Ser/Thr protein kinase).
COG2208	Serine phosphatase RsbU, regulator of sigma subunit.
COG2390	Transcriptional regulator, contains sigma factor-related N-terminal domain.
COG2747	Negative regulator of flagellin synthesis (anti-sigma28 factor).
COG3026	Negative regulator of sigma E activity.
COG3073	Negative regulator of sigma E activity.
COG3086	Negative regulator of sigma E activity.
COG3160	Regulator of sigmaD.
COG3806	Anti-sigma factor
COG5662	Predicted transmembrane transcriptional regulator (anti sigma factor).
COG4650	Sigma54-dependent transcription regulator containing an AAA-type ATPase domain and a DNA-binding domain.
COG4941	Predicted RNA polymerase sigma factor containing a TRP repeated domain
COG5260	DNA polymerase sigma

Tabla 1 y Tabla 2. COGs obtenidos de la base de datos local COG a partir de la búsqueda por palabras clave.

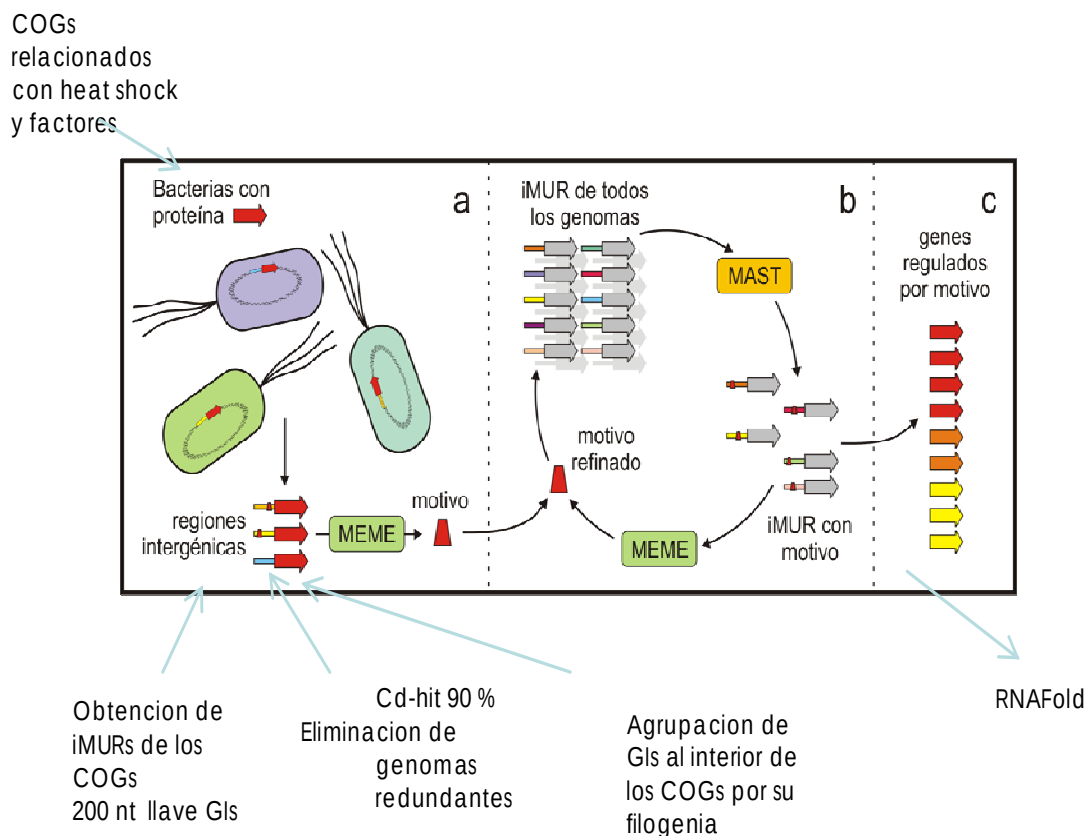


Figura 10. Resumen del método para detectar elementos regulatorios conservados. a) Se toman todas las regiones intergénicas de los COGs (flecha roja) relacionados con las palabras clave “shock” y “sigma” respectivamente, y con MEME se encuentran motivos sobre-representados. b) Cada motivo es refinado al ciclar entre buscarlo con MAST en todas las regiones intergénicas (iMUR) de genomas completamente secuenciados y redefinir el motivo con MEME a partir de las nuevas secuencias encontradas. c) Al final, se obtiene una lista de genes regulados por el motivo, incluyendo miembros de otros COGs además del inicial. Se realiza un análisis de estructura secundaria con RNAfold. Esquema modificado de (1).

1.1.2. Regiones de regulación intergénicas.

Con el subconjunto de COGs definido anteriormente, se extrajeron todas las regiones regulatorias intergénicas de los genomas totalmente secuenciados al momento del estudio. La lista de dichos genomas se encuentra listada en la Tabla 3, de la presente tesis. Este conjunto de secuencias intergénicas conocidas como iMUR (por sus siglas en inglés, intergenic Minimal Upstream Regions, Regiones Mínimas Río-arriba intergenicas). Para evitar tener falsos positivos, es importante que las secuencias a analizar por el programa MEME, no contenga regiones codificantes que pudieran tener regiones conservadas por restricciones funcionales-estructurales impuestos evolutivamente a las proteínas. En adición a la restricción de secuencias exclusivamente no-codificantes, el tamaño de secuencias intergénicas máximo de análisis se estableció en 200 nucleótidos, que es la ventana en la que la mayoría de los termo-riboswitches conocidos es encontrada. Así se obtuvo una clasificación de iMUR por ortología: para cada COG asociado al choque térmico o factor sigma teníamos un conjunto definido iMUR.

1.1.3. Genomas no-redundantes.

Uno de los primeros aspectos que deben ser tomados en cuenta es la gran cantidad de genomas bacterianos secuenciados hasta la fecha, ya que el interés en esta labor se centra en aquellos de importancia industrial, médica, básica y en general económica. Aunado a lo anterior, existen casos en los que se cuenta además con secuencias de genomas filogenéticamente muy cercanos, por ejemplo de distintas cepas tratándose de una misma especie, como en el caso de *E. coli* que se cuenta con 4 cepas secuenciadas o *Salmonella* con 5 cepas secuenciadas al momento de la realización de esta tesis. Contar con estos genomas puede ser de utilidad si se quiere comparar y comprender las diferencias en la patogenicidad de las cepas, sin embargo sería desventajoso incluir todas estas secuencias en un análisis de genómica comparativa como el que pretendemos llevar a cabo, ya que le darían mayor peso al contenido informacional de un pequeño grupo de organismos que se encuentran sobre-representados.

Abreu-Goodger (2005) señala que cuando se hace uso de un algoritmo de descubrimiento de patrones se da por supuesto que las secuencias son independientes, por lo que es necesario generar una serie de secuencias no-redundantes. Para este fin, se empleó el programa CD-HIT (Cluster Database at High Identity with Tolerance, (Li y Godzik, 2006) el cual toma una base de datos de secuencias en formato fasta y devuelve un conjunto de secuencias representativas no-redundantes. Con base en la secuencia más larga de la base de datos ('longest sequence first') se crea un conjunto de secuencias que remueve a aquellas secuencias que se encuentran por debajo de cierto corte de identidad con esta primera secuencia. Adicionalmente en este procedimiento se implementa un método heurístico rápido para encontrar segmentos con alta identidad entre secuencias.

La redundancia entre genomas se eliminó globalmente utilizando todas las secuencias iMUR de la base de datos local iMUR_Genome con un valor de corte de 0.90, lo cual significa que al realizar una comparación pareada, aquellas secuencias que se asemejen en un 90% o mayor, son eliminadas. Como resultado se obtuvo un listado final con 60 genomas no-redundantes. A este se le añadió *E. coli* K12 ya que al ser un organismo modelo y contar con bases de datos completas sobre su regulación es posible analizar de manera exhaustiva las secuencias usando RegulonDB y DBTBS (en el caso de *B. subtilis*) (Tabla 3).

El mismo programa "0_extrae_iMUR_COG_filogenia" que extrae las secuencias iMUR para cada COG, emplea el listado pasado para extraer exclusivamente los iMUR de los genomas no-redundantes por COG. Lo anterior fue la materia prima para la búsqueda de elementos de regulación conservados.

Genomas no redundantes			
Organismo	Phylum	Organismo	Phylum
B_japonicum	Proteobacteria	O_iheyensis	Firmicutes
S_avermitilis	Actinobacteria	M_mazei	Euryarchaeota
M_loti	Proteobacteria	B_suis_1330	Proteobacteria
S_cerevisiae	Fungi	D_radiodurans	Deinococcus-Thermus
S_meliloti	Proteobacteria	L_plantarum	Firmicutes
Nostoc_sp	Cyanobacteria	S_solfataricus	Crenarchaeota
R_etli_CFN_42	Proteobacteria	X_fastidiosa	Proteobacteria
B_anthraxis_Ames_0581	Firmicutes	S_aureus_Mu50	Firmicutes
P_aeruginosa	Proteobacteria	P_aerophilum	Crenarchaeota
E_coli_O157H7_EDL933	Proteobacteria	Halobacterium_sp	Euryarchaeota
A_tumefaciens_C58_Uwash	Proteobacteria	N_europaea	Proteobacteria
S_pombe	Fungi	C_tetani_E88	Firmicutes
P_putida_KT2440	Proteobacteria	A_fulgidus	Euryarchaeota
R_solanacearum	Proteobacteria	C_tepidum_TLS	Chlorobi
V_vulnificus_YJ016	Proteobacteria	P_furiosus	Euryarchaeota
B_bronchiseptica	Proteobacteria	C_burnetii	Proteobacteria
B_thetaiotaomicronn_VPI-5482	Bacteroidetes	W_succinogenes	Proteobacteria
R_palustris_CGA009	Proteobacteria	M_thermoautotrophicum	Euryarchaeota
S_typhi	Proteobacteria	T_maritima	Thermotogae
L_interrogans_serovar_Lai	Spirochaetes	A_pernix	Crenarchaeota
M_acetivorans	Euryarchaeota	M_jannaschii	Euryarchaeota
S_oneidensis	Proteobacteria	B_longum	Actinobacteria
C_violaceum	Proteobacteria	M_leprae	Actinobacteria
G_violaceus	Cyanobacteria	A_aeolicus	Aquificae
Y_pestis_KIM	Proteobacteria	T_volcanium	Euryarchaeota
M_tuberculosis_CDC1551	Actinobacteria	R_conorii	Proteobacteria
B_subtilis	Firmicutes	R_prowazekii	Proteobacteria
B_halodurans	Firmicutes	B_floridanus	Proteobacteria
V_cholerae	Proteobacteria	E_coli_K12	Proteobacteria
C_acetobutylicum	Firmicutes		
C_crescentus	Proteobacteria		
Synechocystis_PCC6803	Cyanobacteria		

Tabla 3. Lista de genomas no redundantes.

1.1.4. Agrupamiento de secuencias iMUR por COG de acuerdo a su filogenia.

Dentro del mismo programa se agruparon las secuencias iMUR por COG en grupos filogenéticos, a partir del *phylum*: 16 *phyla* en el dominio Bacteria, 3 *phyla* en el dominio Archaea y 3 *phyla* en el dominio Eukarya. En tesis previas a ésta, no se había hecho dicho agrupamiento, y la búsqueda era más general lo cual representaba un valor de probabilidad en la conservación de los motivos menor, dando lugar al aumento de posibles resultados falsos positivos. En este trabajo se reduce la cantidad de secuencias por grupo filogénético y por tanto si el motivo se

encuentra representado en la mayoría de éstas, la relación señal-ruido se disminuye considerablemente, así se espera una reducción de dichos falsos-positivos. Aunado a lo anterior, cabe mencionar que por los antecedentes en el descubrimiento de termómetros de RNA, éstos se encuentran conservados al interior de un grupo filogenético y parece ser que no se presenta un mismo elemento en grupos distantes, de tal forma que la búsqueda de motivos por grupos filogenéticos tendría una mayor sensibilidad de encontrar nuevos termómetros de RNA específicos para cada uno de dichos grupos.

Por otro lado, habiendo identificado una firma significativa, es posible indagar acerca de su conservación en grupos filogenéticamente lejanos: si está o no está conservada en otras especies, así como determinar si ese motivo se encuentra río arriba de otros genes adicionales cuya función difiere de los genes que en un principio se buscaron (relacionados con el choque térmico y factores sigma) que en última instancia pudieran estar regulados por un mismo elemento. Esta búsqueda es posible con los programas MEME-MAST (Bailey y Elkan, 1994; Bailey y Elkan, 1998).

1.1.5. Buscando firmas conservadas: MEME y MAST.

MEME (EM Múltiple para Obtención de Motivos, por sus siglas en inglés, donde EM hace referencia al algoritmo para maximizar la expectancia) (Bailey y Elkan, 1994) fue usado para encontrar los motivos conservados en las secuencias iMUR. Como se mencionó anteriormente, el resultado que genera este programa es una o varias matrices específicas de posición que representan a uno o más motivos sobre-representados dentro de un conjunto de secuencias determinadas. La idea fundamental del programa es barrer secuencias calculando la probabilidad que tiene cada posible nucleótido en una ventana de secuencia de distintos tamaños para cada posición. MEME además compara motivos de diferentes tamaños, con el fin de elegir aquel que pudiese tener una mayor relevancia biológica, de tal modo que el usuario no necesariamente conoce el tamaño. En nuestro trabajo se buscaron motivos de al menos 10 nucleótidos de longitud. Asimismo el programa asigna un valor de expectancia a cada uno de los motivos. Solo aquellos con un valor de expectancia mayor al de la probabilidad de encontrarlo azarosamente, son considerados como significativos dado un valor de corte impuesto por el usuario. En nuestro caso, el valor de corte impuesto fue de 10^{-8} . El valor de corte es estricto lo cual permite encontrar elementos reguladores en algunas secuencias al interior del grupo ortólogo. Estos valores de corte fueron escogidos con base al análisis de grupos control con termoriboswitches conocidos (Abreu-Goodger, 2005).

Uno de los parámetros que MEME permite definir es la distribución de motivos; en nuestro estudio se eligió la opción en la cual el motivo se encuentre sólo en algunas secuencias de

entrada, considerando que no todas ellas presentan motivos exclusivamente del tipo de elemento termo-sensor, dando una mayor flexibilidad en la primera búsqueda de motivos. Una vez encontrados los motivos sobre-representados, se puede refinar esta búsqueda a través del programa MAST. Otra opción del programa MEME es el número de motivos que MEME reporta; en nuestro estudio se le pidió encontrar hasta 6 motivos.

En el caso de obtener motivos significativos, éstos se localizaron de manera global en la región intergénica correspondiente y se analizaron para saber si habían sido reportados en trabajos previos -específicamente en el trabajo descrito por (Abreu-Goodger et al., 2004)-, usando el programa RibEx (Explorador de Riboswitches, <http://132.248.32.45/cgi-bin/ribex.cgi>) (Abreu-Goodger y Merino, 2005) en el cual se encuentra la información para cada elemento regulador encontrado, incluyendo el COG, KEGG más significativo, la distribución filogenética del elemento, así como ligas para Rfam (RNA families database of alignments and Covariance Models) para los elementos que ya se encontraban previamente reportados. Para realizar comparaciones se utilizó la opción de ver una imagen con la distribución de motivos y la calificación de todos los genes de los genomas completos en los cuales existe un elemento regulador.

Utilizando los motivos significativos obtenidos anteriormente, hicimos búsquedas con el programa MAST sobre el conjunto total de iMUR, sin agrupación por COG o por filogenia. Lo anterior se realizó con la finalidad de localizar todos los genes que pudiesen ser regulados por el mismo elemento y que no necesariamente sean ortólogos, de tal modo se puede obtener un grupo expandido de genes.

MAST es el complemento de MEME y lo que hace es considerar matrices de peso posición-dependiente que representan a los motivos encontrados obtenidas con el análisis de MEME. El usuario provee una base de datos de secuencias y MAST regresa las secuencias que presentan el motivo, la posición de éste y una calificación de probabilidad. Así, MEME y MAST se pueden usar iterativamente en la búsqueda y refinamiento de motivos sobre-representados de un conjunto de secuencias.

En general, las secuencias en donde se desea identificar un motivo determinado puede presentar una frecuencia no-equilibrada de cada uno de sus elementos (nucleótidos en secuencias de DNA y aminoácidos en secuencias de proteínas), sino presentar algún tipo de sesgo en su composición. Este sesgo puede incidir favorable o desfavorablemente para que los motivos sobre-representados sean identificados. Por ejemplo, encontrar más fácilmente a un motivo rico en ATs en un fondo de secuencias intergénicas con un sesgo composicional de mayor contenido de los nucleótidos A y T. Para corregir la sobre o subidentificación de motivos originado

por el sesgo de fondo, MAST tiene la opción de considerar el contenido composicional del fondo cuando éste es proporcionado por el usuario. Considerando lo anteriormente expuesto, se realizó un programa (programa_piloto_corre_meme.pl) que proporciona a MAST el archivo con las frecuencias de nucleótidos de las regiones intergénicas de los genomas a analizar (Ver Apéndice 1).

Adicionalmente, con el fin de identificar y analizar la congruencia de los genes que se encuentran a través de MAST y el descubrimiento de la función en común de un conjunto de genes, así como la visualización de otros genes que se encuentran en la vecindad tanto río abajo como río arriba, es decir el contexto genómico, se empleó la herramienta web GeConT y GeConT 2 (Ciria et al., 2004; Martínez-Guerrero et al., 2008).

1.1.6. Predicción de la estructura secundaria.

Es importante puntualizar que la señal de regulación que se espera identificar está localizada en una molécula de RNA y depende de la estructura que ésta adopte una vez que ha sido transcrita. Con el propósito de conocer si los motivos identificados pudieran ser parte de elementos relevantes en la estructura secundaria del RNA y potencialmente relacionados con el monitoreo de la temperatura, se empleó el programa RNAfold (Hofacker y Stadler, 2006). A partir de una secuencia de RNA, este programa predice la estructura secundaria de RNA cuya energía libre sea mínima –la estructura de mayor estabilidad termodinámica-, y genera una matriz de probabilidad de apareamiento. La temperatura a la cual se llevaron a cabo las predicciones fue la que el programa tiene por omisión: 37°C. Al analizar la estructura secundaria se identificó si el SD estaba o no formando parte de una región apareada en la molécula de RNA.

Los resultados obtenidos fueron analizados y comparados con lo reportado anteriormente, con el fin de discernir si se trata de información nueva o coincide con lo establecido. Los casos identificados en nuestro estudio que contaban con antecedentes experimentales publicados en la literatura, formaron parte de nuestros controles positivos. Por otro lado, los casos nuevos e interesantes carentes de referencias previamente reportadas constituyeron nuestro conjunto de candidatos para ser verificados experimentalmente en un futuro próximo.

1.2. Resultados y Discusión Bioinformática

1.2.1. Análisis de los COGs relacionados con el choque térmico por calor (“shock”).

Se probaron 16 casos en relacionados con los COGs de choque térmico. Se consideraron únicamente aquellos casos en donde se presentaban motivos conservados con un valor E por arriba de $1e^{-10}$. Debido a esta restricción al final sólo se obtuvieron 9 COGs representativos. A continuación se presenta una tabla con la información sobre los distintos motivos detectados por el método, conocidos y desconocidos (Tabla 4). Para la revisión, análisis y documentación de éstos, se consultó RegulonDB (Salgado et al., 2004), DBTBS (Makita et al., 2004) y RibEx (Abreu-Goodger y Merino, 2005). Uno de los principios bajo los cuales reside esta tesis, es que al realizar el análisis de motivos, el agrupamiento filogenético por *phylum* le daría al método una mayor sensibilidad, así como una mayor especificidad en la búsqueda de elementos reguladores vía RNA. Sin embargo, como se puede apreciar en la Tabla 4, un buen agrupamiento, no es garantía de encontrar únicamente elementos de RNA, aunado a que sino es que para todos los casos estén o no altamente conservados, ya han sido anteriormente descritos. Lo anterior puede deberse a que, en general los termosensores de RNA como se ha visto, se encuentran en grupos filogenéticos particulares (p.ej. ROSE en Alpha y Gamma Proteobacteria) y/o en gran parte de los casos en una especie aislada bacteriana (p.ej. *IcrF* en *Listeria monocytogenes*), sin que necesariamente exista la conservación en secuencia primaria en otras especies.

1.2.1.1. Elementos reguladores conocidos basados en RNA.

En esta sección se describen aquellos resultados en donde encontramos que el motivo obtenido en nuestra búsqueda, coincide con cierto sensor de temperatura basado en RNA descrito en estudios previos, estos casos constituyen el control positivo de la metodología bioinformática.

La búsqueda con la palabra clave “shock” arrojó el motivo asociado al COG0071, (Molecular chaperone, small heat shock protein. Posttranslational modification, protein turnover, chaperones) en el grupo Proteobacteria, dicho motivo corresponde al termómetro de RNA ROSE cuya firma y estructura se encuentran reportadas previamente en la base de datos Rfam (RF00435, <http://rfam.sanger.ac.uk/family?entry=RF00435>) lo cual es un control positivo al validar los resultados obtenidos con el método empleado en la presente tesis. La Figura 11 muestra los operones identificados en nuestro estudio que presentan este motivo, así como el consenso del motivo. Interesantemente, esta firma también fue identificada en nuestro servidor web RibEx y corresponde al elemento RLE0343.

COG	Phylum	Motivos	Elemento regulatorio	Referencias
COG0234 Class I heat shock protein chaperonin <i>GroES</i> (<i>HSP10</i>)	Firmicutes	1 6.3e ⁻²⁶	CIRCE (sitios HrcA)	<i>RibEx</i>
	Proteobacteria	1 6.3e ⁻⁷³		
COG0326 Molecular chaperone HSP90 Family	Proteobacteria	1 6.5e ⁻¹¹	Sitio de unión de sigma 32	RegulonDB
COG0443 Molecular chaperone involved in induction by stress Conditions.	Proteobacteria	3 a) 6.3e ⁻¹⁹ b) 6.3e ⁻¹⁴ c) 3.1e ⁻¹²	a) Sitio de unión de sigma 32 (<i>dnaK</i>) b) No se reporta en RegulonDB (<i>hscC</i> , <i>yegD</i>) c) Sitio de unión de sigma 32 (<i>dnaK</i>)	a) y c) RegulonDB
COG0459 Class I heat shock protein chaperonin <i>GroEL</i> (<i>HSP60</i>)	Proteobacteria	1 4.2e ⁻¹³	No reportado, no se recupera en <i>E.coli</i> K12	
	Euryarchaeota	1 8.12e ⁻²²	¿?	
	Crenarchaeota	1 5.6e ⁻¹¹	¿?	
COG0466 Class III heat shock ATP dependent Lon-protease Bacterial type	Proteobacteria	1 5.3e ⁻²¹	Sitio de unión de sigma 32 y Sigma 70	RegulonDB
COG0789 Involved in transcriptional Regulation (repression) of Heat shock proteins	Firmicutes	1 2.4e ⁻³⁰	Sitio de unión de sigmaA	<i>RibEx</i> DBTBS RegulonDB
	Proteobacteria	2 a) 1.5e ⁻⁷⁵ b) 3.8e ⁻²⁵	a) Sitio de unión de SoxR que actúa como represor b) no se encuentra en <i>E.coli</i> K12 y no recupera a <i>B.subtilis</i>	
COG0576 Molecular chaperone GrpE Heat shock protein	Proteobacteria	1 1.2e ⁻³¹	Sitio de unión de sigma 32	<i>RibEx</i> RegulonDB
COG1278 Cell processes; adaptation	Firmicutes	1 4.6e ⁻¹⁷	Recupera el SD, termosensor para El choque frío y cold box	DBTBS <i>RibEx</i> RegulonDB
Cold shock proteins	Proteobacteria	1 8.6e ⁻²⁷	Terminador traduccional, Termosensor para el choque frío y Cold box	Yamanaka et al., 1999
COG1420 Involved in transcriptional Regulation (repression) of Heat shock gene	Firmicutes	1 1.3e ⁻²²	CIRCE (sitios HrcA) Autorregulación	DBTBS
COG0071 Molecular chaperone	Proteobacteria	1 5.5e ⁻⁴⁶	Repressor of heat shock (ROSE)	<i>RibEx</i> RegulonDB

Tabla 4. Resultados de los motivos encontrados en las secuencias que pertenecen a los COGs asociados al choque térmico.

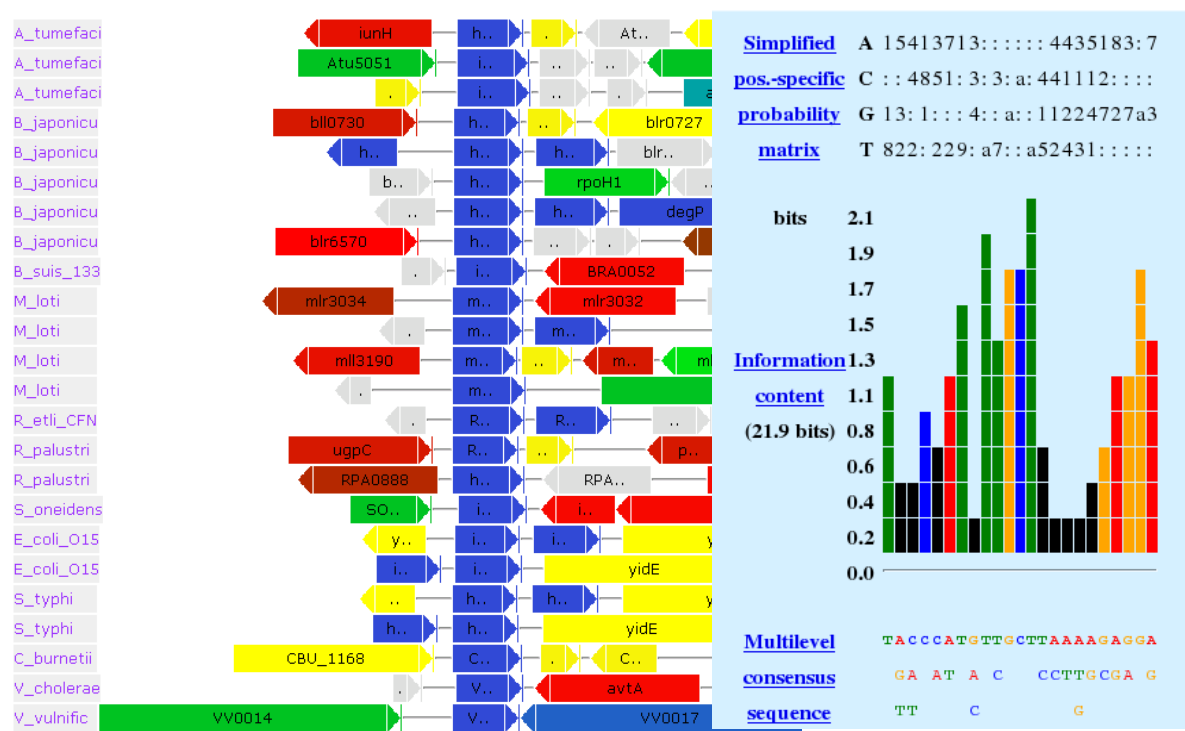


Figura 11. Motivo y operones que presentan a éste del COG0071, pertenecientes al termómetro de RNA ROSE. El motivo se encuentra en las regiones intergénicas de proteínas que codifican para proteínas pequeñas asociadas al choque térmico por calor, presentes en Proteobacteria. Abajo se muestra la representación gráfica del motivo obtenido por MEME, en donde se observa el contenido informacional de acuerdo a la conservación del motivo.

El COG1278 contiene a proteínas que son inducidas durante el choque térmico por frío (Cold Shock Proteins) (Figura 12). La regulación de los genes de las CSP's ocurre generalmente a nivel post-transcripcional; en estudios previos (Yamanaka *et al.*, 1999) se encontró que la región intergénica 5' del mRNA de *cspA*, presentaba un tamaño considerablemente grande (160 nucleótidos) y una estructura particular, por lo cual se especuló que dicha estructura a bajas temperaturas, en particular la región del SD y otra región que se encontraba altamente conservada a la que denominaron upstream box (UB), podría ser distinta a la que se observa a 37°C, temperatura a la cual el SD se encuentra apareado (Figura 13). El cambio estructural puede hacer al mRNA *cspA* más accesible a los ribosomas. En este mismo estudio se sugirió que la UTR 5' podría estar actuando como un termómetro de RNA, de manera similar al termómetro *cIII* en el bacteriófago lambda (Narberhaus, 2002). Esta firma también fue identificada en nuestro servidor web *RibEx* y corresponde al elemento RLE0357.

La estructura secundaria de esta región intergénica 5' es muy peculiar: en la mayoría de los casos por 4 residuos de uracilo, semejante a la estructura del termómetro FourU; cercanamente de la UB se encuentra una estructura que se asemeja a un brazo, la cual consta de una caja y un tallo-asa (Figura 13). Aún no se ha descrito detalladamente cómo ocurre la interacción entre el ribosoma y este elemento de RNA, por lo que esta 5' UTR del gen *cspA* podría estudiarse

experimentalmente con el fin de determinar el mecanismo molecular de la caja UB y si se trata o no de un termómetro de RNA *bona fide*, como los descritos en la revisión hecha en la introducción de esta tesis.

A.

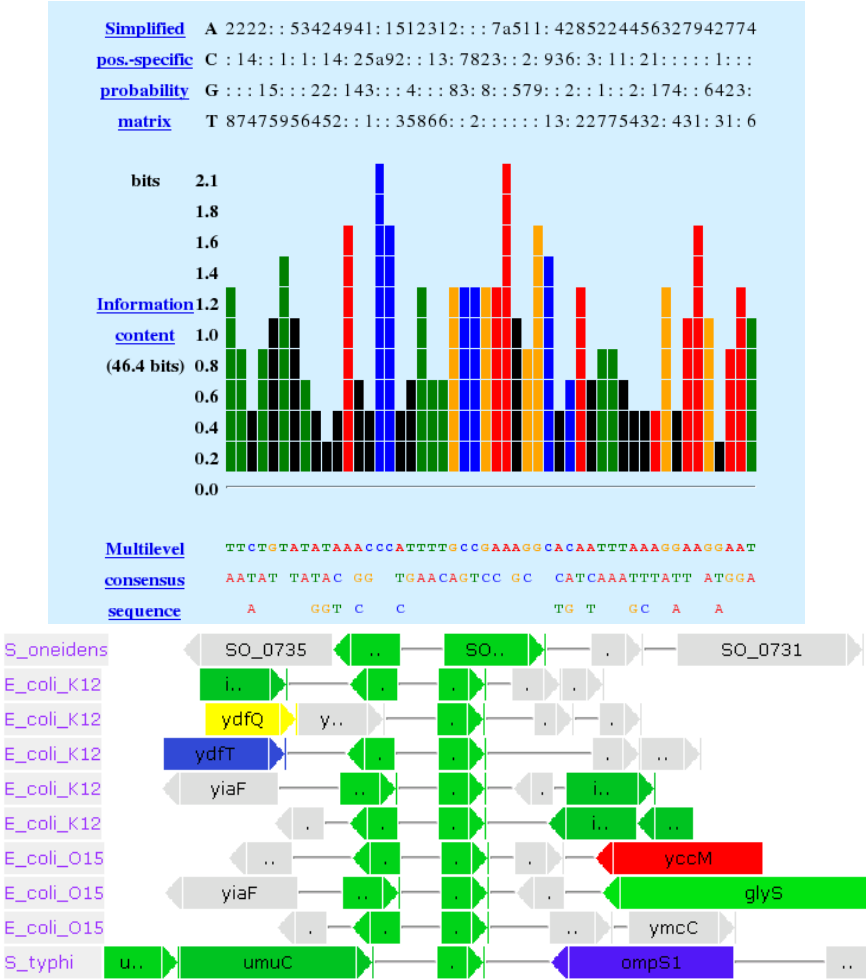
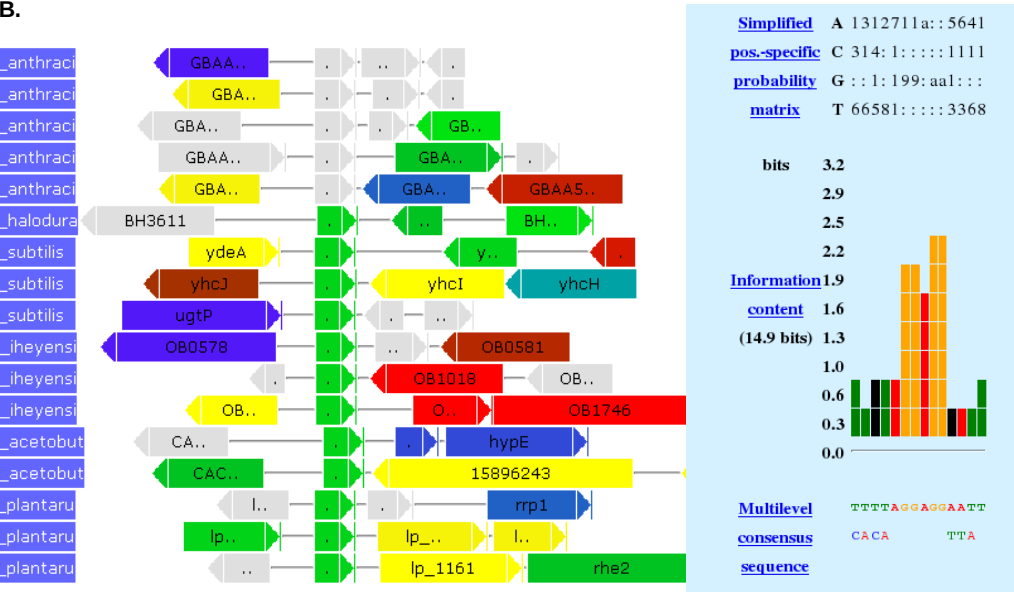


Figura 12. Operones y motivos presentes al interior del COG1278. A) Genes en los cuales se presenta el motivo y representación gráfica de éste para el grupo de Proteobacteria. B) Genes en los cuales se presenta el motivo y representación gráfica de éste para el grupo de Firmicutes, se observa que el motivo en este grupo es muy pequeño y corresponde perfectamente con el sitio SD, además de recuperar *per se* la secuencia de este sitio, en la figura 17 se muestra para algunos casos conservación estructural secundaria similar a la descrita como “upstream box” (UB)

B.



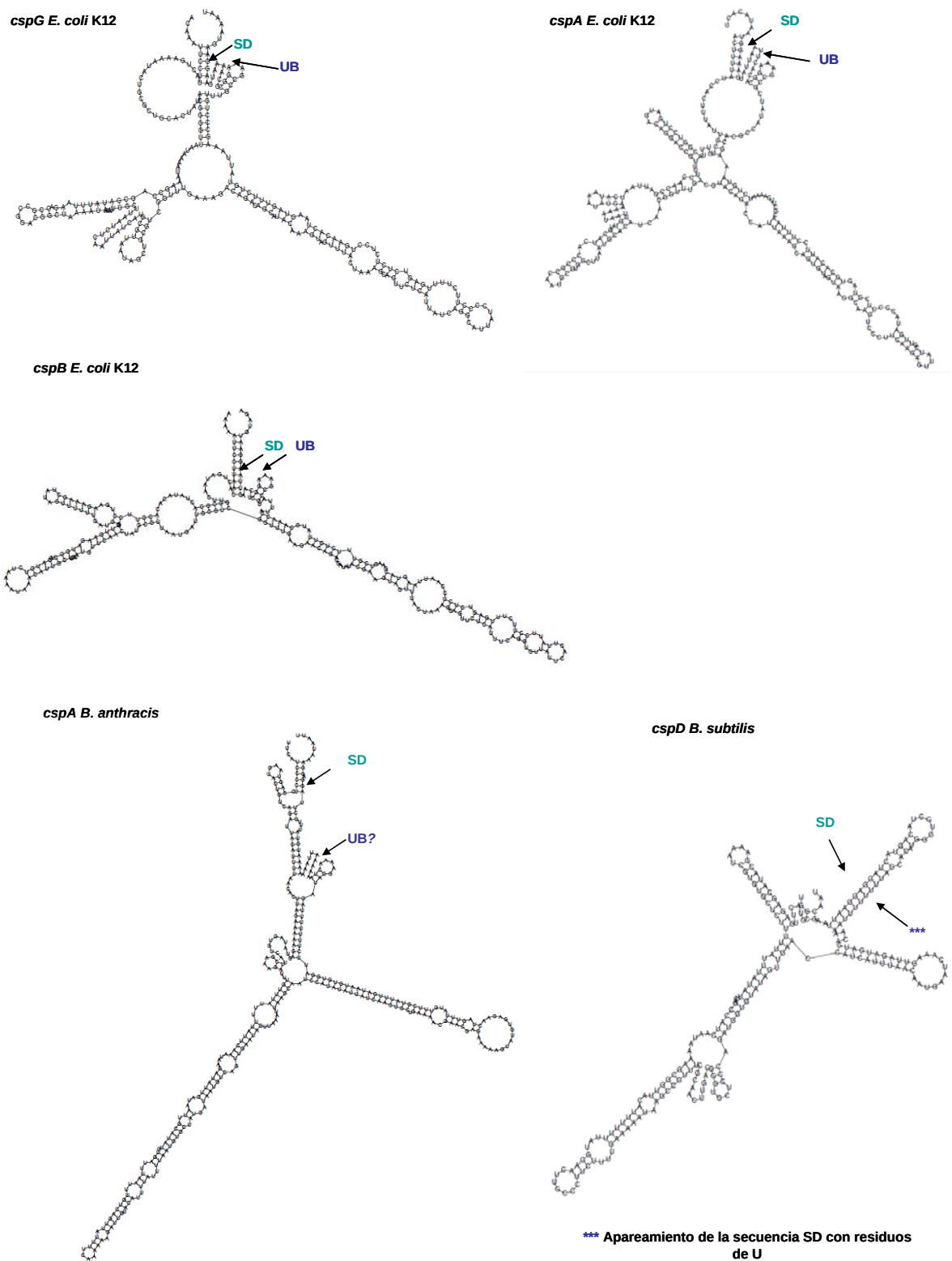


Figura 13. Conservación estructural del elemento formado en la región intergénica 5' de proteínas de choque térmico por frío. En la parte superior se muestran las estructuras secundarias de las regiones intergénicas de los genes *cspA*, *cspB* y *cspG* en *E. coli* K12 en donde se observa el apareamiento de la secuencia SD y la “upstream box” (ver texto). En la parte inferior se muestra la estructura secundaria para 2 regiones intergénicas de CSPs en *B. subtilis* y *B. anthracis*, se observa una vez mas el apareamiento de la secuencia SD, en el caso de *cspA* en *B. anthracis* se observa una estructura parecida a la UB, sin embargo distante al SD. *cspD* en *B. subtilis* muestra apareamiento con cuatro residuos de U, tal como se describe para el termómetro de RNA FourU (ver texto).

1.2.1.2. Sitios de interacción DNA-proteína.

La metodología bioinformática en nuestro análisis, en principio pretendía recuperar únicamente elementos regulatorios basados en RNA, sin embargo las 5' UTRs bacterianas contienen de manera importante sitios de unión a proteínas reguladoras de la expresión genética, por lo que esperábamos encontrar este tipo de señales representadas como motivos conservados; a continuación se describen brevemente dichos resultados, en los cuales los motivos pertenecen al sitio de pegado de alguna proteína reguladora.

En el COG0789, el cual agrupa a reguladores transcripcionales en Proteobacteria, el motivo resultó ser un sitio de unión para la proteína represora SoxR. Respecto al COG0234 (Co-chaperonin GroES HSP10, RLE0003) y COG1420 (Transcriptional regulator of heat shock gene) tanto en Proteobacteria (exceptuando γ Proteobacteria) y Firmicutes, son casos de los más representativos ya que coinciden con un motivo palindrómico denominado CIRCE (Controlling Inverted Repet for Chaperone Expression). CIRCE es una estructura repetida e invertida, los cuales son sitios de unión para la proteína HrcA que reprime a genes asociados al choque térmico por calor, por lo cual anteriormente se pensó que era un termosensor (Yuan y Wong, 1995). Esta proteína se encuentra con una alta conservación en el dominio Bacteria, de tal modo que se especula que la presión de selección para mantener este sistema de regulación es suficiente como para encontrar a CIRCE como un motivo y sitio de unión a DNA sobrerrepresentado.

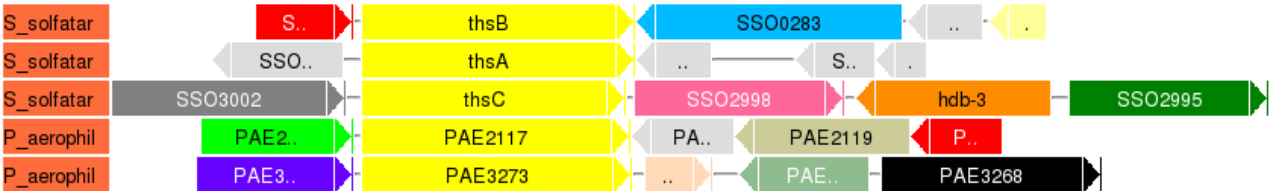
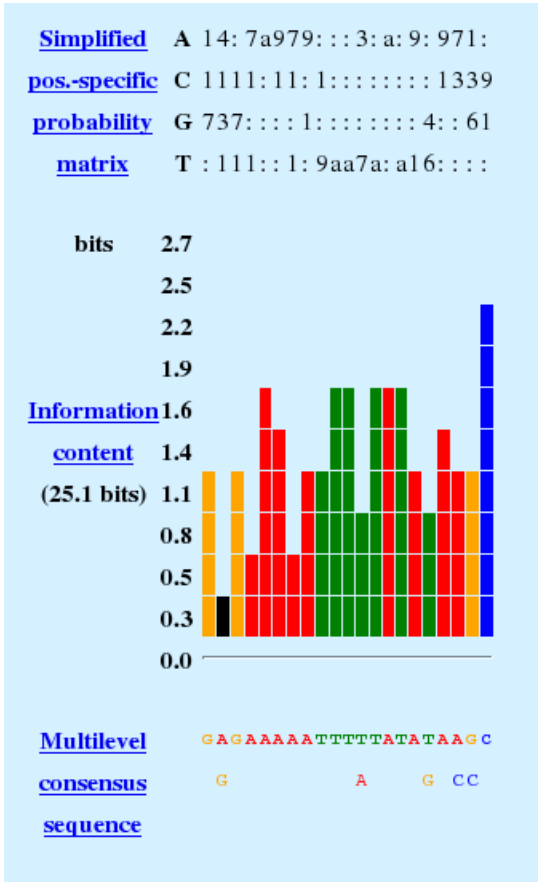
Un caso particular de interacción DNA-proteína corresponde a los sitios de unión para factores sigma, los cuales son los motivos más comúnmente encontrados con este método y los observamos en los siguientes COGs: COG0326 (Molecular chaperone, HSP90 family, Proteobacteria), COG0443 (Molecular chaperone DnaK, RLE0050, Proteobacteria), COG0466 (ATP-dependent Lon protease, bacterial type, Proteobacteria), COG0789 (Predicted transcriptional regulators, RLE0029, Firmicutes) y el COG0576 (Molecular chaperone GrpE, heat shock protein, RLE0235, Proteobacteria). Estos motivos bien caracterizados se encontraron a través de RegulonDB (*E. coli* K12) y DBTBS (*B. subtilis*); la información que se obtiene de estos análisis es valiosa ya que, es a partir de ésta es posible extrapolar los sitios de unión para ciertas proteínas (en este caso factores sigma) en otros genomas para los cuales se desconoce o se carece de la información acerca de la regulación, los posibles factores de transcripción y factores sigma involucrados en ella.

1.2.1.3. Motivos encontrados en arqueobacterias que aún no han sido descritos.

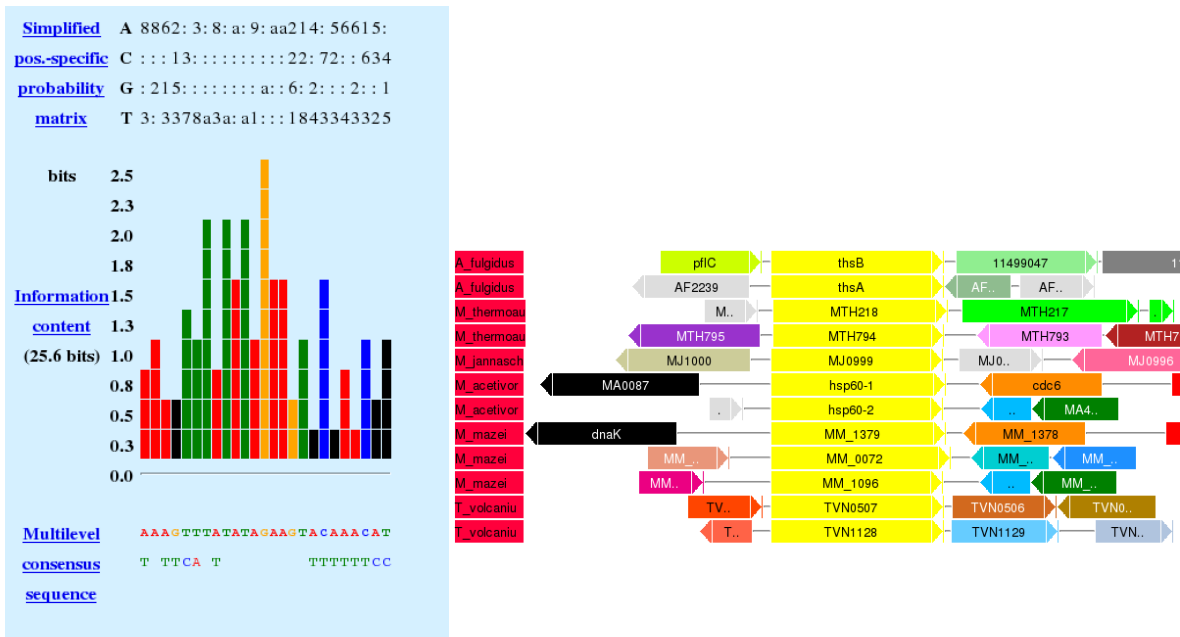
Los motivos al interior del COG0459 - Chaperonin GroEL (HSP60 family) - tanto para Proteobacteria como para Euryarchaeota y Crenarchaeota que encontramos, no han sido descritos en trabajos previos; más aún la metodología no recupera ningún motivo en *E. coli* K12, por lo que no se sabe a que corresponde este motivo en Proteobacteria (Figura 14). En el caso de

Euryarchaeota y Crenarchaeota, resulta más difícil por un lado, pero llamativo por el otro, intentar encontrar la razón por la cual esta firma se encuentra tan conservada, aunado a que el análisis estructural revela para algunos casos el apareamiento de la secuencia SD con una secuencia de residuos de uracilo, por lo cual dilucidar experimentalmente que molécula se pega a esta región y si se trata o no de un elemento termorregulador de RNA, aportaría información suficiente para ampliar este estudio e identificar nuevos elementos de RNA en el dominio Archaea en donde se carece de análisis previos y hasta la fecha no han sido descritos.

Motivo COG0459
Crenarchaeota



Motivo COG0459
Euryarchaeota



Motivo COG0459
Proteobacteria

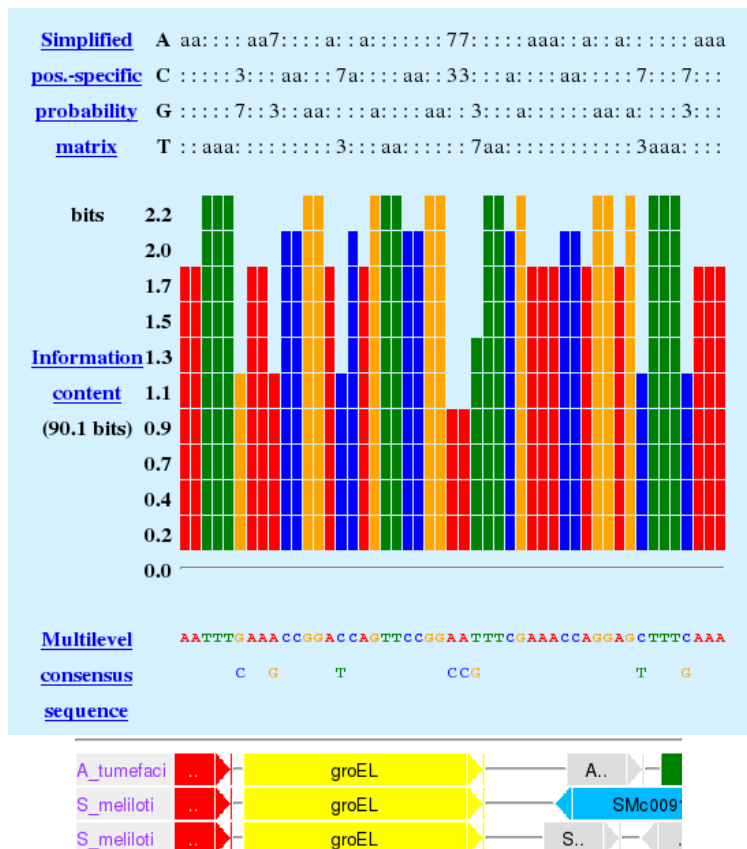


Figura 14. Motivos encontrados en el COG0459. Se muestran los motivos encontrados al interior de este COG y su distribución en los grupos filogenéticos Proteobacteria, Crenarchaeota y Euryarchaeota.

1.2.2. Análisis de los COGs relacionados con factores sigma

La mayoría de las especies bacterianas sintetizan disitintos factores sigma que reconocen diferentes secuencias consenso. Esta variedad de factores sigma da la posibilidad de que las bacterias puedan mantener la expresión genética basal, así como la regulación de la expresión genética en respuesta a cambios ambientales o señales de desarrollo (Wosten, 1998). Resulta evidente pensar la posibilidad de encontrar un termosensor de RNA en el gen *rpoH* que codifica para el factor sigma *bona fide* que se activa cuando la célula es sometida a estrés por calor: σ_{32} . Sin embargo existen factores sigma que se activan bajo otras condiciones en *E. coli*: σ_{70} (RpoD) es el factor sigma "housekeeping" y transcribe la mayor parte de los genes de las células en crecimiento, se encarga de la producción de proteínas necesarias para mantener a la célula viva; σ_{54} (RpoN) el factor sigma que funciona durante la limitación de nitrógeno; σ_{38} (RpoS) el factor sigma de fase estacionaria; σ_{28} (RpoF) el factor sigma flagelar; σ_{24} (RpoE) el factor sigma de estrés extracitoplasmático, así como de estrés térmico extremo; y finalmente σ_{19} (FecI) el factor sigma de citrato férrico que regula al gen *fec* para el transporte de hierro. Así, el hecho de encontrar un termómetro de RNA para un factor sigma, motivó la búsqueda de algún otro elemento regulador basado en RNA que controlara la expresión de otros factores sigma que se activan en respuesta a los diferentes estímulos, teniendo este primer nivel de control basado en RNA de manera inmediata.

Del mismo modo que para los COGs asociados al estrés por choque térmico se presenta en la siguiente tabla aquellos COGs asociados a la palabra clave “shock” que presentaron motivos conservados.

COG	Phylum	Motivos	Elemento regulatorio	Referencias
COG0568 DNA-directed RNA polymerase Sigma subunit	Proteobacteria	1 4.1e ⁻²⁰	Secuencias promotoras, una de Ellas corresponde al sitio de unión Del factor sigma54	RibEx RegulonDB
COG1191 DNA-directed RNA polymerase Specialized sigma subunit	Firmicutes	1 5.4e ⁻¹¹	Recupera la región SD, motivo Pequeño, no hay apareamiento del SD	
COG1595 DNA-directed RNA polymerase Specialized sigma subunit	Firmicutes	1 1.3e ⁻³⁶	En <i>B. subtilis</i> se trata de sitios de Autorregulación en los factores Sigma con función extracitoplásmica	RibEx DBTBS Eiamphungporn y Helmann, 2009
	Proteobacteria	1 7.1e ⁻¹¹³	Sitio de unión de Fur	

Tabla 5. Resultados de los motivos encontrados en las secuencias que pertenecen a los COGs asociados a factores sigma.

Al hacer la búsqueda por palabra clave “sigma” se encontraron 19 COGs entre los cuales se encuentran reguladores transcripcionales de estos factores sigma. Realmente existen 6 COGs que incluyen exclusivamente factores sigma, de los cuales únicamente hay 3 casos de que presentan motivos conservados con un valor de expectación significativo, que a continuación son descritos.

1.2.2.1. COG1191 (DNA-directed RNA polymerase specialized sigma subunit): Un sensor basado en RNA, implicado en la esporulación en *Bacillus subtilis*.

La firma encontrada en el COG1191 en Firmicutes (DNA-directed RNA polymerase specialized sigma subunit) es sumamente pequeña (GGAGGGAAAG) y corresponde a la secuencia SD (Figura 15). Por la naturaleza de esta firma, es posible que se trate de un sitio de pegado de alguna proteína de unión a RNA o DNA, cuyo sitio de pegado sobrelapa al SD, y forma parte de otro motivo de mayor tamaño; no obstante al realizar el análisis con la base de datos sobre la regulación transcripcional de *B. subtilis*, DBTBS, no se encontraron reportes previos sobre ésta. De manera conjunta se realizó el análisis de estructura secundaria de esta región intergénica en *Bacillus subtilis*, *Bacillus halodurans*, *Bacillus licheniformis*, *Bacillus cereus* y *Bacillus anthracis*; se observa la formación de estructuras de tipo tallo-asa que secuestran al SD (sitio de unión al ribosoma), en el caso de las regiones intergénicas de *Bacillus cereus* y *Bacillus anthracis* el apareamiento del sitio SD es parcial, ya que sólo mitad de este se encuentra secuestrado (Figura 16).

Esta firma se encuentra en la 5'UTR del gen *sigG* o *spoIIIG*, el cual en *B. subtilis* codifica para el mismo factor σ^G , que se expresa particularmente en la “pre-espora” una célula pequeña que se forma durante la primera y segunda hora de esporulación, cuando la célula se divide de manera asimétrica, formándose la célula madre y dicha “pre-espora” (Sun et al., 1991; Chary et al., 2005) (Figura 17).

El gen *sigG* se encuentra dentro del operón *spoIIIGA-sigE-sigG* (Figura 18) éste primero se transcribe a partir del promotor río arriba del primer gen *spoIIIGA*, en etapas tempranas de la esporulación (Chary et al., 2005). Sin embargo la traducción de *sigG* es escasa o nula, y se ha especulado que dicha ineficiencia es resultado de la formación de una estructura tallo-asa que secuestra al sitio SD, la cual se encuentra en la región intergénica de *sigG* (Masuda et al., 1988; Sun et al., 1991). Al mover *sigG* a un *locus* ectópico, la formación de las esporas continúa, de tal forma que la transcripción desde el promotor de *spoIIIG*, no es crítica para la expresión de σ^G ; el rol de este transcrito (*spoIIIG-sigE-sigG*) aún es incierto (Hilbert y Piggot, 2004).

El *locus sigG*, después de la formación del septo (Figura 17) se transcribe y traduce productivamente desde su propia región promotora, a través del factor σ^F que también está confinado a la “prespora”, la presencia de una región promotora dependiente de σ^F asegura por

un lado la presencia de σ^G y la activación de este último factor únicamente en la “prespora” (Hilbert *et al.*, 2004).

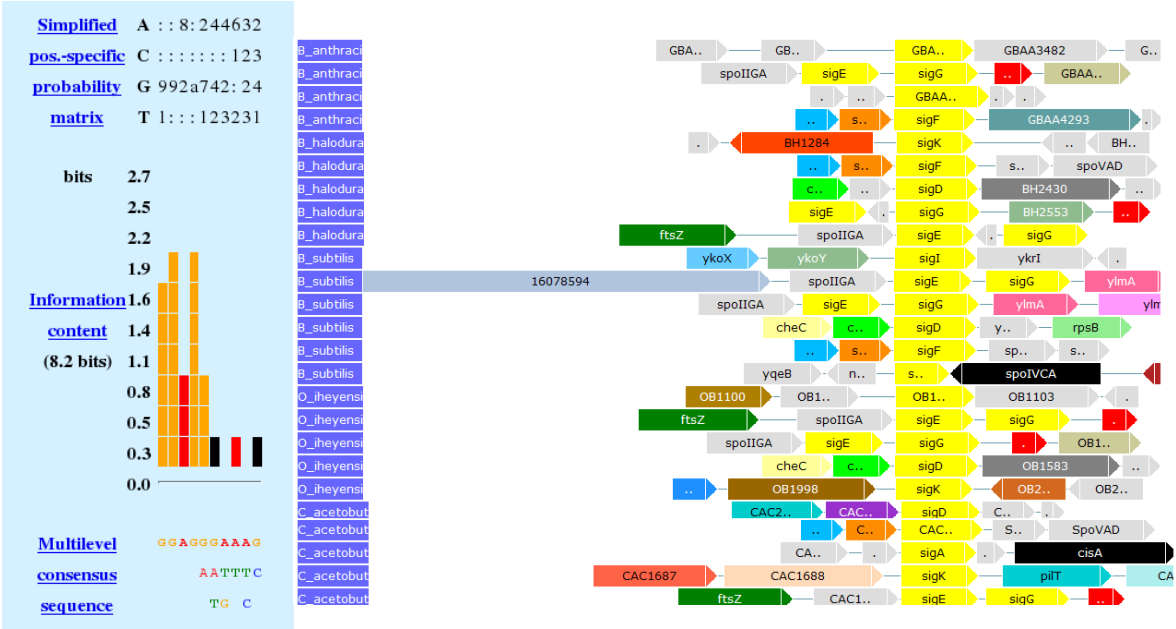
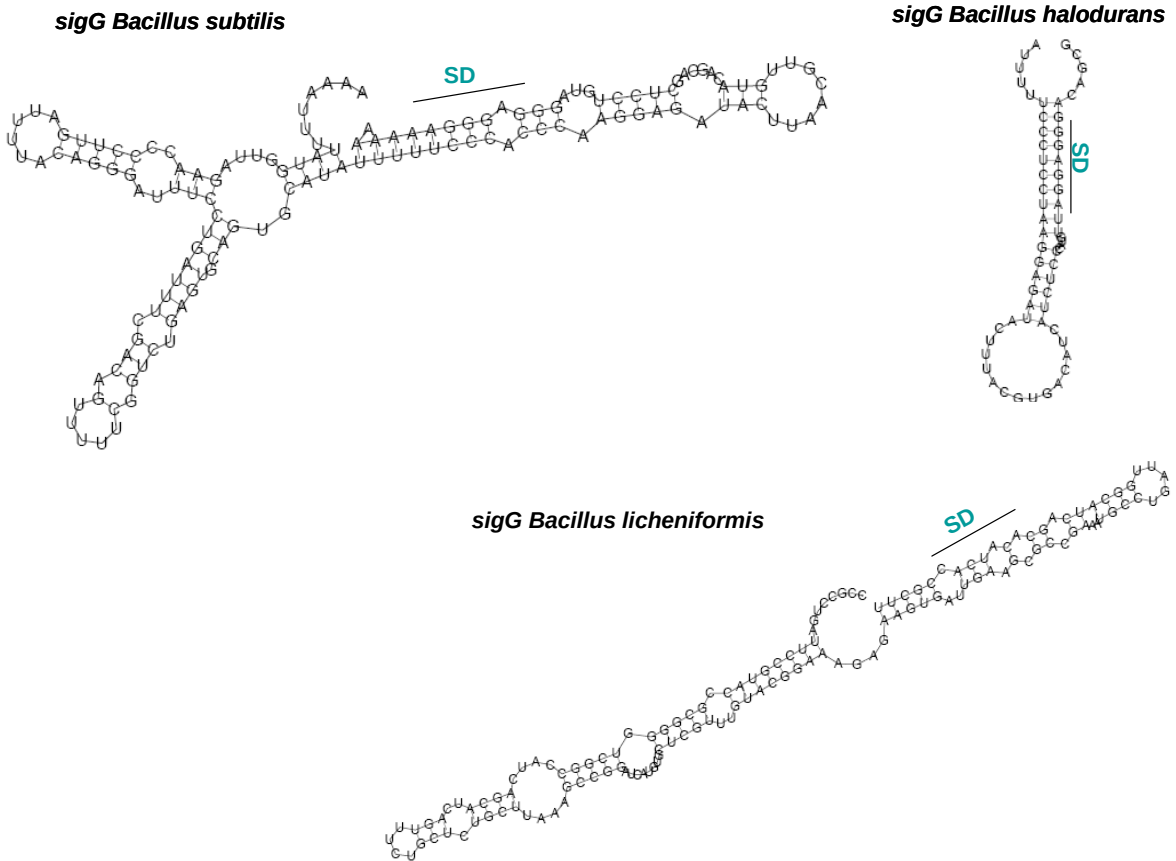


Figura 15. Motivo encontrado en el COG1191 en Firmicutes. Se observa claramente que el motivo corresponde al sitio de unión al ribosoma SD.



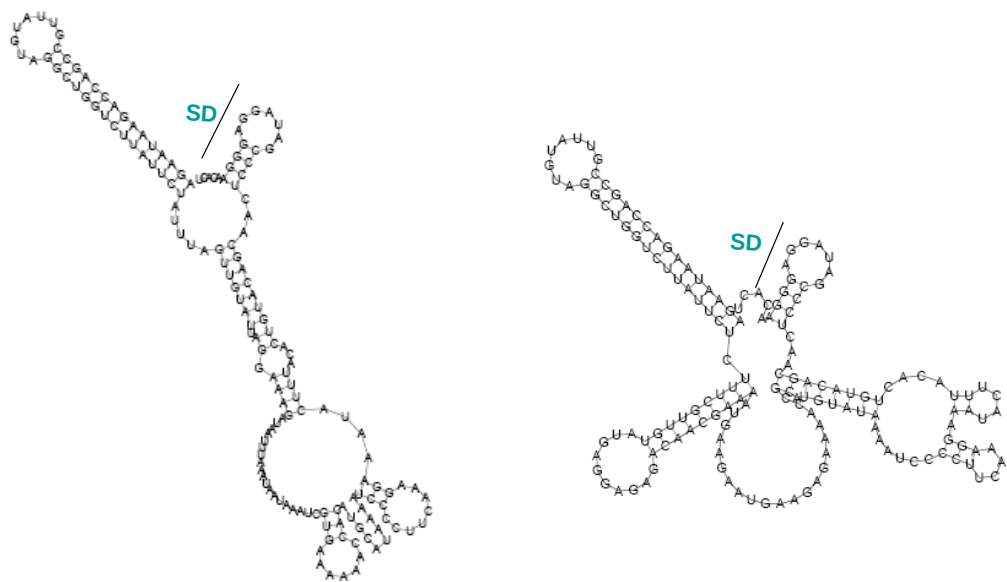


Figura 16. Estructura secundaria de la región intergénica 5' del factor σ^G en *Bacillus*. En la página anterior se muestran las estructuras secundarias de las regiones intergénicas del gen *sigG* en *B. subtilis*, *B. licheniformis* y *B. halodurans*, el sitio de unión al ribosoma SD, se encuentra enmascarado en una estructura de tipo tallo-asa. En esta página se muestra la estructura secundaria para las regiones intergénicas del mismo gen en *B. cereus* y *B. anthracis*, se observa el apareamiento parcial de la región SD.

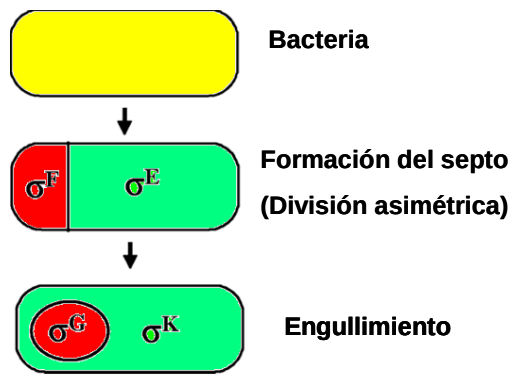


Figura 17. Representación esquemática de las etapas de esporulación, mostrando el lugar y actividad de los factores sigma específicos durante este proceso. Entre los procariontes, la formación de esporas en *B. subtilis* se ha convertido en un paradigma en el

estudio de la diferenciación celular. Poco después del inicio de la formación de esporas, la bacteria (amarillo) se divide asimétricamente para dar lugar a la célula pequeña, denominada “prespora” (rojo) y a la célula madre de mayor tamaño (verde). Posteriormente la célula madre engulle (inmersión) a la “prespora”. La “prespora” se desarrolla en la espora madura, mientras que la célula madre se lisa. El proceso de formación de esporas se caracteriza por la activación celular específica y ordenada de cuatro factores sigma. Inmediatamente después de la formación del septo, σ^F se activa en la prespora, dicha activación contribuye a la activación del factor σ^E en la célula madre. No es hasta que finaliza la engullimiento, que σ^G se vuelve activo en la prespora, lo cual a su vez activa a σ^K en la célula madre. Así, la activación de σ^G en la prespora, que ocurre hasta que se completa el engullimiento, depende primeramente de la activación de σ^F en la prespora, y de la activación de σ^E en la célula madre. La activación de los factores sigma está estrechamente coordinada dentro y entre los dos tipos celulares, en un proceso que se conoce como regulación “crisscross”. Tomado y modificado de (9).

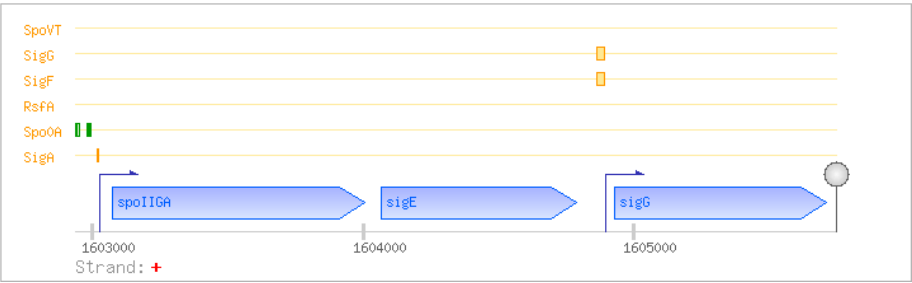


Figura 18. Localización del gen *sigG* (*spoIIIG*) en el operón *spoIIIGA-sigE-sigG*.

La estructura tallo-asa que enmascara la región SD que proponen Masuda et al., (1988) y Sun et al., (1991) corresponde a la firma encontrada en esta tesis. Poco se sabe de la función fisiológica, Masuda et al., (1988) sugieren que podría ser un tipo de atenuación a nivel de la traducción y que debería existir un mecanismo que interrumpiera esta estructura tallo-asa. Asimismo hacen una importante observación sobre los nucleótidos del inicio de la transcripción en la región intergénica de *sigG*, los cuales forman parte del asa de la estructura; lo que significa que únicamente la mitad del tallo se transcribe, previniendo la formación de la estructura tallo-asa y en última instancia no habría control a nivel de la traducción, permitiendo la síntesis proteica de *sigG*, a través de su propia región promotora (Masuda et al., 1988). Entonces la atenuación a nivel de la traducción, ocurre cuando *sigG*, se transcribe desde el promotor de *spoIIIG*, y está presente la estructura tallo-asa que secuestra el sitio de unión al ribosoma de *sigG*, por lo que la traducción es abatida (Figura 19).

La cuestión relevante e inexplorada que podría estudiarse experimentalmente es dilucidar cuál es la función de este tipo de atenuación en la regulación de la expresión de *sigG* durante la esporulación, así como la importancia del transcrito *sigG* que proviene de la transcripción a partir del promotor de *spoIIIG*; finalmente explorar la posibilidad de encontrar alguna proteína implicada en el “planchado” de esta estructura tallo-asa como mecanismo alternativo, o si es suficiente que el inicio de la transcripción de la región promotora este a la mitad de dicha estructura para la traducción de *sigG*.

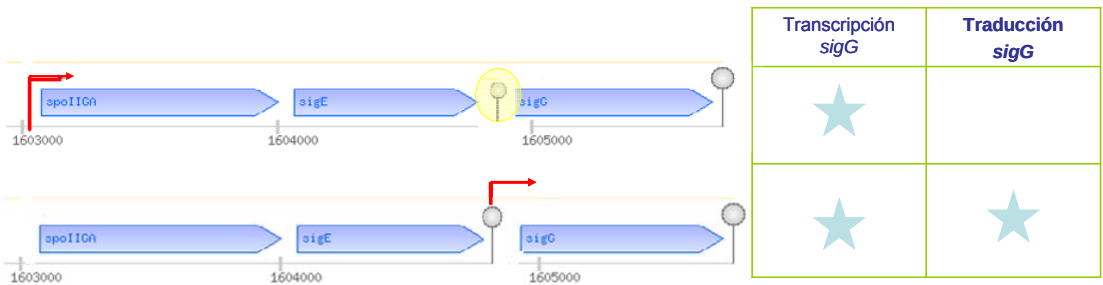


Figura 19. Atenuación a nivel de la traducción en *sigG*. *sigG* puede transcribirse desde el promotor que se encuentra al inicio del operón (parte superior), sin embargo no se traduce y por tanto no hay producto proteico (σ^G), la razón de este hecho es la posible formación de una estructura secundaria de tipo tallo-asa que secuestra el sitio de unión del ribosoma en la región intergénica de *sigG*. Por el contrario, cuando *sigG* se transcribe desde su propio promotor (parte inferior) la traducción ocurre, esto se debe posiblemente a que el sitio de inicio de la transcripción se localiza dentro del tallo-asa y sólo la mitad de la estructura se traduce (ver texto). Este posible mecanismo de atenuación y su importancia durante la esporulación sigue siendo desconocido.

1.2.2.2. El caso de los factores sigma del COG0568 (RLE0166) en *Proteobacteria* (DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)).

Los factores sigma que pertenecen a este COG en *E. coli* K12 que presentan el motivo son *rpoS* (sigma38) *rpoH* (sigma32) y *rpoD* (sigma70) (Figura 20). Al realizar la búsqueda en RegulonDB, se encontró que por lo común estos genes presentan varias regiones promotoras, además de presentar diferentes sitios de unión para una serie de factores transcripcionales como CRP, ArcA, LexA y CytR. A pesar de ello, ninguno de estos sitios corresponde a la firma encontrada: en el caso del gen *rpoD* no se encuentra región alguna que coincida con el motivo (Figura 21a), para el gen *rpoS* se reporta una región promotora inferida por el curador de esta base de datos (Figura 21b) (Gama-Castro et al., 2008) careciendo de la información sobre el factor sigma correspondiente; y para el último gen, *rpoH* se encontró interesantemente que correspondía a la región promotora dependiente del factor sigma 54 (Pallen, 1999) (Figura 21c); sin embargo a la fecha aún se desconocen las condiciones específicas que inducen el inicio de la transcripción del promotor de *rpoH* (Janaszak et al., 2007). El consenso reportado para la unión del factor $\sigma 54$ es YTGGCACCG-4N-TTGCWNN y se encuentra en la posición -12 y -24 respecto al inicio de la transcripción (Barrios et al., 1999).

La subunidad $\sigma 54$ no es esencial para el crecimiento de *E. coli* bajo condiciones de laboratorio, sin embargo es necesaria para la expresión de genes involucrados en diversos procesos celulares tales como el metabolismo de nitrógeno, transporte de ácido dicarboxílico, formación del pilus, síntesis de formato deshidrogenasa, metabolismo de compuestos aromáticos (xileno, tolueno), así como la expresión del operón *pspA-E* (un operón dependiente del choque térmico) y el gen *ibpB* que codifica para una pequeña proteína de choque térmico (Barrios et al., 1999).

En *E. coli* se estiman 30 promotores que dependen del factor $\sigma 54$ (funciona principalmente durante la limitación de nitrógeno). Dicha cantidad es pequeña si se contrasta con el hecho de que existen secuencias abundantes que se asemejan a los promotores canónicos de $\sigma 54$.

Cabe la posibilidad de que el motivo encontrado en *rpoS* y *rpoD* sean posibles regiones promotoras para la unión de $\sigma 54$, ya que al comparar los resultados generados por MEME parecen ser sitios similares al consenso de $\sigma 54$ reportado previamente, aunque se observa cierta variación entre los nucleótidos (ver Figura 22).

Motivo COG0568
Proteobacteria

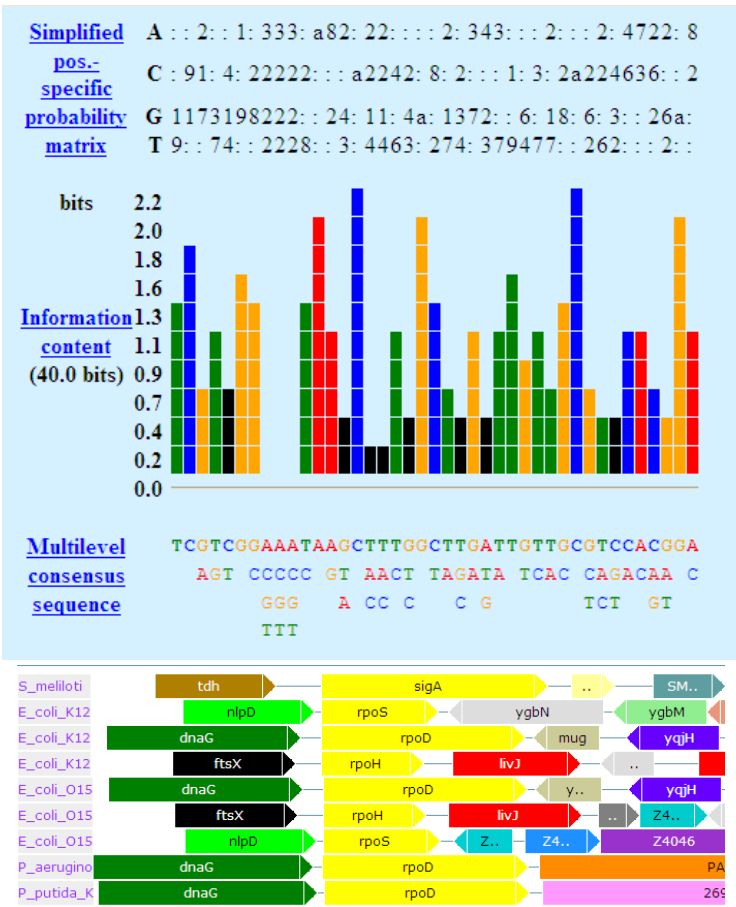


Figura 20. Motivo encontrado en el COG0568 en Proteobacteria. Distribución del motivo en los diferentes operones y representación gráfica del motivo; en *E. coli* K12 el motivo se encuentra en las regiones intergénicas de los genes que codifican para los factores sigma 70 (“housekeeping”) sigma 32 (asociado con el choque térmico por calor) y finalmente sigma 38 (involucrado en la fase estacionaria y en la limitación de nutrientes).

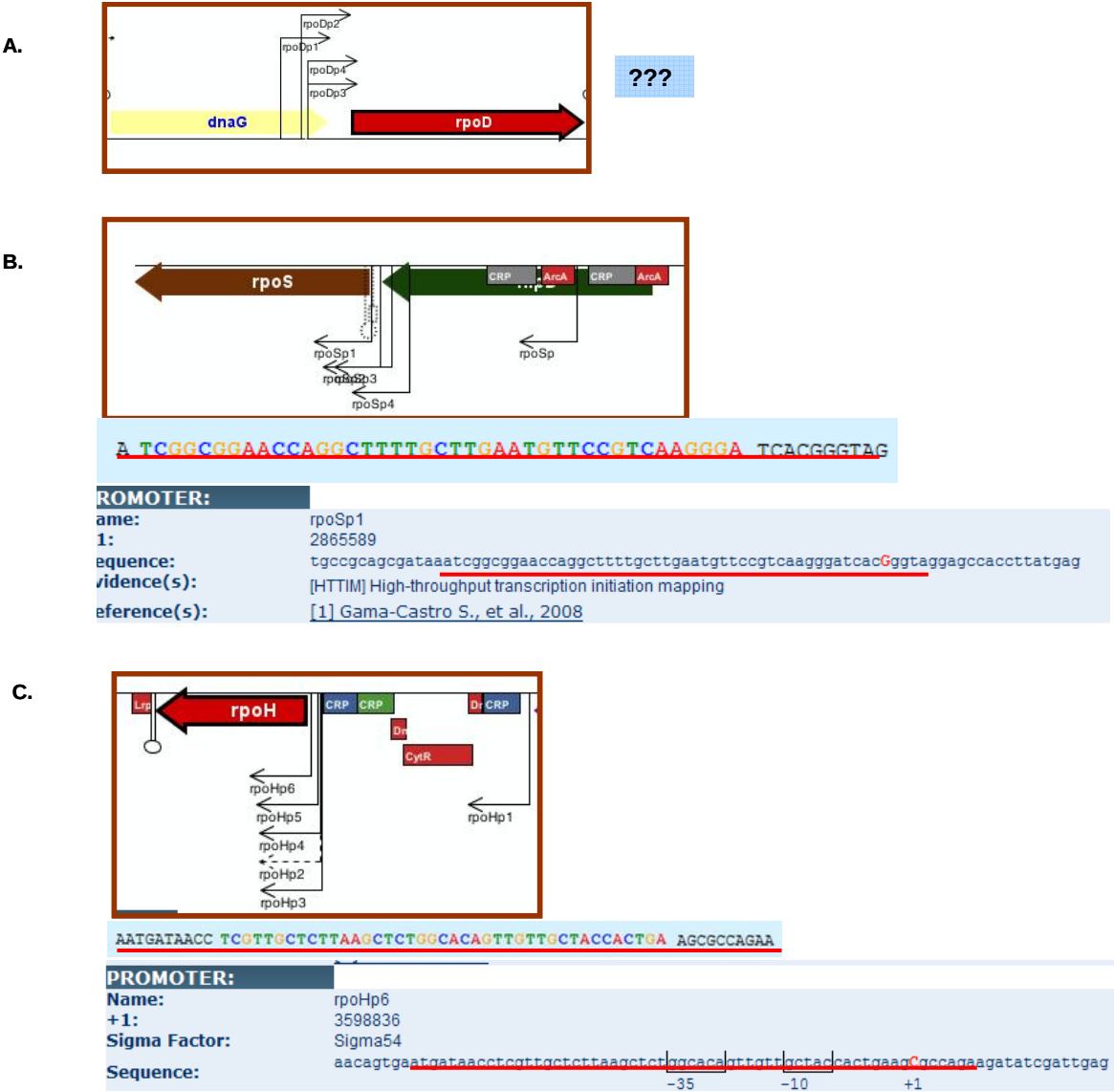


Figura 21. Análisis de los motivos encontrados en el COG0568 (DNA-directed RNA polymerase, sigma subunit, sigma70/sigma32) para *E. coli* K12. a) El motivo encontrado para el gen *rpoD* no se encontró en la base de datos RegulonDB. b) El gen *rpoS* presenta una región promotora inferida por el curador de RegulonDB, cuya secuencia coincide con el motivo encontrado con la metodología empleada en esta tesis. c) En el caso de *rpoH* el motivo corresponde a un sitio de unión ya descrito para el factor $\sigma 54$, de este hecho parte la hipótesis de que es probable que para los otros dos genes (*rpoD* y *rpoS*) sea también un sitio de unión para dicho factor (ver texto).

	NAME	START	P-VALUE	SITES
	15803258	2	2.55e-15	A TCGGC GGAACCAGGCTTTTGCTTGAATGTTCCGTCAAGGGA TCACGG
rpoS	16130648	2	2.55e-15	A TCGGC GGAACCAGGCTTTTGCTTGAATGTTCCGTCAAGGGA TCACGG
	15803973	48	3.04e-14	AATGATAACC TCGTTGCTCTTAAGCTCTGGCACAGTTGTTGCTACCACTGA AGCGCC
rpoH	16131333	48	3.04e-14	AATGATAACC TCGTTGCTCTTAAGCTCTGGCACAGTTGTTGCTACCACTGA AGCGCC
	15803609	19	1.19e-13	CTCGTAATTA TCGTTGGCGGTAAACAACCGTTGGATTTAGCGTTAACGGC TGAAGG
rpoD	16130963	19	1.19e-13	CTCGTAATTA TCGTTGGCGGTAAACAACCGTTGGATTTAGCGTTAACGGC TGAAGG
	26987129	7	2.80e-13	GGCCCA TGCTCGGGTATAATCCTCGGCTTGTTTTTCGCCGCCAAGA CCTTCA
	15595773	19	2.36e-12	CTCAGCCAG GCATCAGGTATAATCCTCTGCTTATTTTTCGCCGCCAAGA CCTTCA
	15966067	20	5.24e-12	TCCGTTACG TCAGGGGAAATAATCGGTGGCGGGTACGCGGCGTGCCCGGA CCTCCG
Secuencia consenso para RpoN (Barrios et al., 1999)				
YTGGCACG-N4-TTGCWNNw				
Secuencia consenso para RpoN (Pallen, 1999)				
TGGCACNNNNNTTGCW				

Figura 22. Hipótesis acerca de nuevos sitios de unión para RpoN en *E. coli* K12. Se muestran los motivos encontrados en los genes *rpoH*, *rpoD* y *rpoS*. En *rpoH* es muy claro que el motivo corresponde al sitio de unión de RpoN ($\sigma 54$). En el caso de los otros genes sólo algunos nucleótidos se comparten entre la secuencia consenso y los nucleótidos de la firma, por lo cual no es claro si se trata o no de un sitio de unión para RpoN ($\sigma 54$), aunque existe cierta evidencia a favor, al menos en el caso del gen *rpoS* en *B. burgdorferi* en donde este factor sigma 54 se une a *rpoS* (ver texto) (Y = C ó T; W = A ó T).

Se ha reportado en la espiroqueta *Borrelia burgdorferi* evidencia de que RpoN ($\sigma 54$) controla al gen *rpoS* uniéndose a la secuencia río arriba -24/-12 del ORF de este gen, controlando en dicho organismo la expresión de factores de virulencia (Smith et al., 2007). Esta predicción resulta un ejemplo interesante que apoyaría la hipótesis de la existencia de diferentes procesos de regulación concertada de varias biomoléculas que más allá de actuar solamente bajo condiciones específicas están relacionados entre si cuando la célula se encuentra bajo estrés, ya sea por calor, limitación de nutrientes, fase estacionaria e incluso bajo condiciones normales. Si alguna de estas predicciones son confirmadas, se puede encontrar la relación entre redes de regulación importantes: entre el regulón RpoD y el regulón RpoN, el regulón RpoS y el regulón RpoN. Es posible comprobar si se tratan o no de sitios de unión para dicho factor de manera experimental.

1.2.2.3. COG1595: Factores sigma que involucrados durante el estrés extracitoplasmático. Dos casos diferentes: Firmicutes y Proteobacteria.

Al interior del COG1595 (DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog) (RLE0075) se agrupan aquellos factores sigma relacionados con la regulación de las proteínas que residen en la membrana externa o el espacio periplásmico, de tal modo que se conocen con el nombre de factores sigma con función extracitoplasmática (por sus siglas en

inglés: ECF Extra Citoplasmic sigma Factors). Este “cluster” filogenético incluye a reguladores del estrés periplásmico y la respuesta al choque térmico por calor (σ^E en *E. coli*), transporte de hierro (FecI en *E. coli*), sistema de flujo de iones metálicos (CnrH en *Alcaligenes*), secreción de alginato (AlgU/T en *Pseudomonas aeruginosa*) y la síntesis de carotenoides localizados en la membrana en *Mixococcus xanthus* (CarQ). La única característica unificadora de estos diferentes procesos fisiológicos es que todos ellos involucran funciones de la cubierta celular (transporte, secreción, estrés extracitoplasmático). Respecto al número de factores sigma ECF se encuentran dos en *E. coli*, siete en *B. subtilis*, diez en *M. tuberculosis* y sorprendentemente ~50 en *S. coelicolor* (Helmann, 2002; Eiamphungporn y Helmann, 2009).

Debido a que el exterior de bacterias invasivas está en contacto con el hospedero y necesita responder al ambiente de este, los factores sigma ECF también regulan los genes asociados con la virulencia bacteriana (Brooks y Buchanan, 2008). Tal es el caso de la actividad de σ^E en *E. coli* involucrada en la uropatogénesis (Redford y Welch, 2006) o *P. aeruginosa* con el homólogo de σ^E , AlgU que regula la expresión de la cubierta mucosa que protege a esta bacteria de los antibióticos, estrés oxidativo y el ataque del sistema inmune.

El monitoreo de las condiciones en el periplasma y/o el ambiente exterior se transduce como señal a la maquinaria genética celular, pasando por una o dos membranas bacterianas al citoplasma. Así la señalización a los factores sigma ECF requiere de proteínas integrales de membrana para establecer esta comunicación. La similitud en cuanto a secuencia, así como estudios funcionales sugieren que la mayoría de estos factores se encuentran regulados por mecanismos similares a la activación del sigma de la respuesta a estrés del periplasma σ^E o la activación del mecanismo del factor sigma de respuesta al citrato férrico, σ^{FecI} (FecI), ambos en *E. coli*. (Brooks y Buchanan, 2008)

El análisis con la metodología desarrollada en la presente tesis muestra que existen al menos 2 motivos conservados en la región de regulación de los genes que pertenecen al COG1595 (DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog) tanto en Firmicutes como en Proteobacteria (Figura 23), como se observa en la Tabla 5, los cuales ya se encontraban previamente descritos, sin embargo son casos demostrativos que generan algunas dudas, las cuales son discutidas en las siguientes secciones.

1.2.2.3.1. Factores sigma involucrados durante el estrés extracitoplasmático en Firmicutes.

B. subtilis posee siete factores sigma ECF parálogos. Los factores ECF mejor entendidos son σ^M , σ^W y σ^X y para los cuales se encontró este motivo antecediendo a sus correspondientes genes codificantes en *B. subtilis* y para otros factores que aún no han sido estudiados: σ^H , σ^V y σ^Y (Figura 23a). Para σ^H es muy claro que se trata de un sitio de unión para AbrB, un regulador

global de la expresión genética que comienza en la fase estacionaria y de esporulación (Strauch, 1995). Es posible que para los otros factores ECF también sea un sitio de unión para este regulador, ya que existe evidencia que sugiere que AbrB y otros reguladores del estado de transición afectan los regulones σ^W y σ^X , sin embargo aún no se esclarecen la relación entre estos reguladores de transición de estado y los regulones de estos factores (Helmann, 2002). El resultado que encontramos en la presente tesis podría ayudar a consolidar la idea anterior y más aún a comprender y esclarecer las interacciones y el mecanismo de acción de estos factores, de los cuales se sabe aún muy poco. Aunado a lo anterior, dentro del mismo motivo, para los otros factores sigma (distintos de σ^H) se reporta el motivo para cada uno de estos factores como un sitio de pegado para si mismo, se ha postulado que estos factores se autorregulan (Helmann, 2002). Si el motivo corresponde a un sitio de autorregulación para cada factor sigma y es el mismo motivo en todas las secuencias esto podría sugerir que estos factores pueden interactuar entre ellos. Helmann (2002) señala bajo los mismos términos de similitud en secuencia, que existen preguntas importantes por resolver: si estos sitios se tratan verdaderamente de autorregulación o no, y en este sentido ¿Cuál es la especificidad del promotor para cada uno de estos factores y en qué son diferentes? Y si cabe la posibilidad de que los factores sigma ECF puedan regularse entre si. Para la primera cuestión Helmann indica que existe ya una evidencia usando microarreglos de DNA en donde se demuestra la autorregulación; además se propone que haciendo una inspección de la secuencia río arriba de los genes de factores sigma ECF ubicando el motivo de autorregulación se pueden buscar nuevos factores sigma de esta clase, no obstante en nuestra búsqueda no se encuentra el motivo en otras secuencias distintas, ni en *B. subtilis* ni en otros genomas.

Respecto a la segunda cuestión es necesario esclarecer si la expresión genética de los factores sigma ECF esta coordinada en respuesta a los distintos estímulos extracitoplasmáticos y más aún si estos pueden interactuar entre si, así como su papel durante dichos procesos ¿Por qué existen 7 factores sigma involucrados y de qué manera lo hacen? de tal modo este es un caso que contribuye de buena manera a apoyar hipótesis expuestas en trabajos previos y resulta interesante para seguir analizándolo de manera experimental.

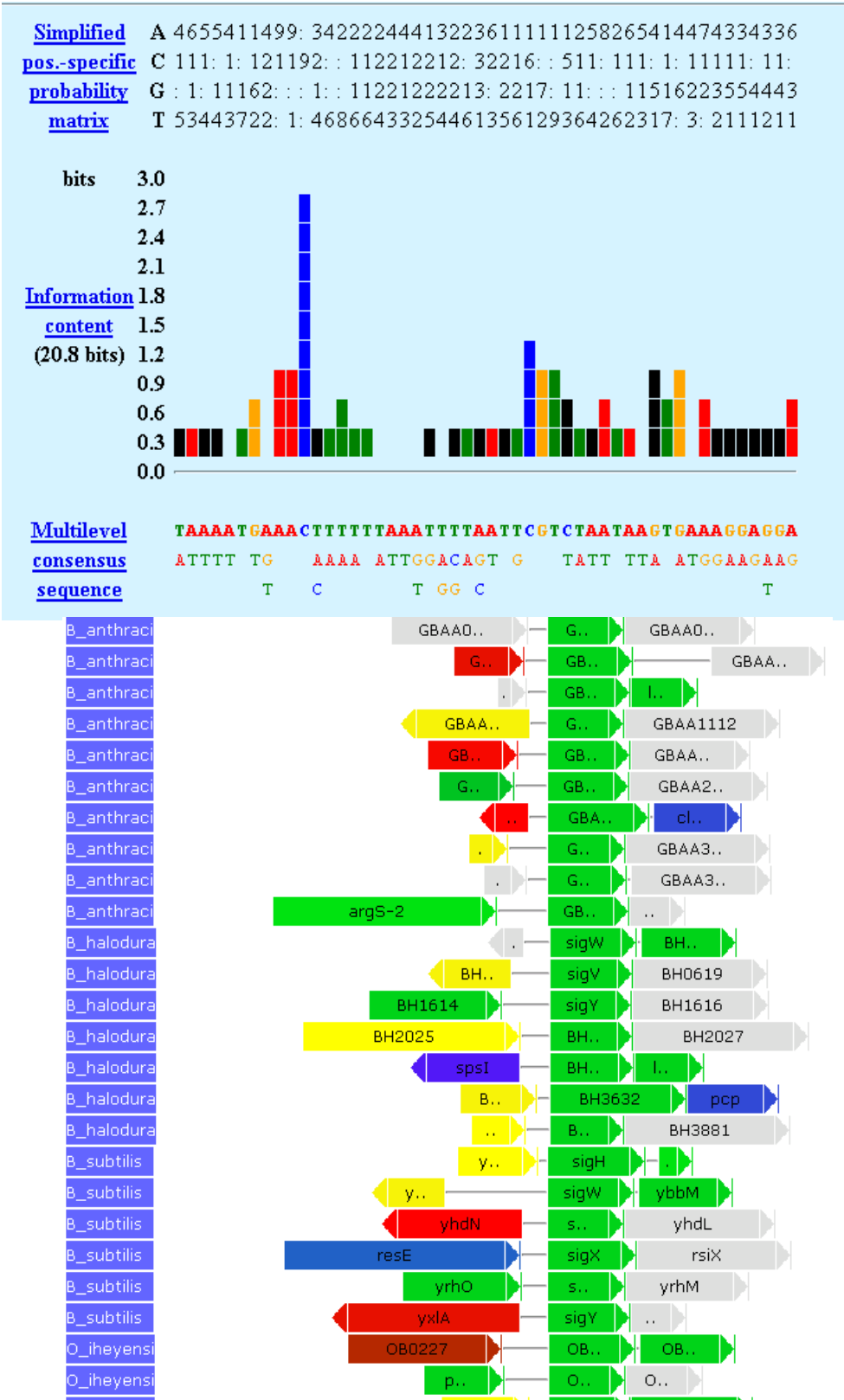
1.2.2.3.2. Un factor sigma involucrado en la expresión de genes relacionados con el transporte de citrato férrico en *Proteobacteria*.

Un caso que desde el inicio de este proyecto generó interés fue el de este motivo en *Proteobacteria* (Figura 23b), ya que este se encontraba con un valor de expectancia sumamente alto y se especulaba como un motivo que debía ser de gran relevancia biológica, fue el del COG1595 (DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog) para el gen que codifica para otro factor sigma ECF: *fecI*.

Motivo COG1595

Firmicutes

A.



Motivo COG1595
Proteobacteria

B.

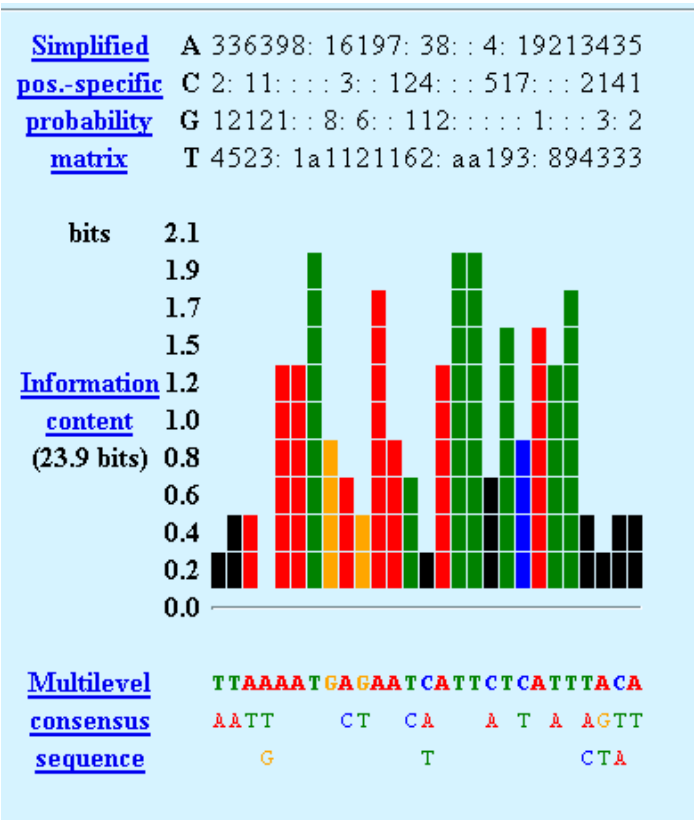


Figura 23. Motivos encontrados al interior del COG1595. a) y b) Representación y distribución del motivo en los operones del grupo Firmicutes y en el grupo Proteobacteria respectivamente; para ambos grupos se trata de motivos con un alto valor de expectancia y por tanto con una gran conservación. En Firmicutes se trata de sitios de unión del mismo factor (autorregulación) y en Proteobacteria de sitios de unión para Fur (ver texto).

En general para las bacterias Gram negativas, resulta un problema poder introducir hierro a la célula debido a la membrana adicional, además de que el hierro *per se* es un recurso escaso y que se encuentra de manera insoluble en ambientes aeróbicos. De tal suerte que estos organismos han desarrollado sistemas complejos para detectar y adquirir al hierro en iones metálicos. Una de las estrategias más común son los sideroforos, los cuales forman complejos con el hierro y posteriormente estos complejos son transportados primeramente hacia el periplasma en donde se pueden unir a una proteína secundaria de unión a hierro. Después el hierro (sideróforo férrico) es transportado a través de la membrana interna al citoplasma (Brooks y Buchanan, 2008)

Aunque los miembros del sistema de adquisición de citrato férrico no son proteínas esenciales, deben tener alguna ventaja porque este sistema en realidad es un sistema bastante costoso, no sólo por la energía involucrada en la síntesis proteica, sino en el transporte del hierro por cada una de las membranas que requiere forzosamente de energía obtenida de la hidrólisis de ATP; una vez que la célula ha adquirido suficiente hierro, la producción de la maquinaria implicada en el transporte de hierro es apagada por el represor Fur (Ferric uptake repressor) en condiciones de abundancia de hierro Fur forma un complejo con este metal, teniendo la forma activa del represor de la expresión genética de los genes del transporte de hierro. Caso contrario es el del factor sigma Fecl que es activado para reclutar a la RNA polimerasa y unirse al operón *fec*, incrementando la expresión de los genes de adquisición de citrato férrico (Pressler et al., 1988).

Se encontró que justamente el motivo tan conservado constituía un sitio de unión para la proteína Fur, y es de esperarse tal conservación, ya que como señala Abreu-Goodger (2005) los sistemas que modifican la concentración de iones Fe^{+} necesitan estar regulados por las razones antes expuestas entorno a la carencia y la necesidad pero por otro lado las concentraciones elevadas de este ión son mortales.

Fur no sólo se encarga de reprimir a los genes que se encargan del aumento en la concentración de hierro celular, sino que se encarga de activar proteínas almacenadoras de hierro de manera no directa, a través de reprimir a su vez un RNA pequeño RyhB, el cual mantendría inactivas a estas proteínas (Masse y Gottesman, 2002). Durante la búsqueda de este motivo que corresponde al sitio de unión para Fur, fue sorpresivo identificar a muchos más genes de diferentes COGs en diferentes especies bacterianas (amplia distribución filogenética) y que estos COGs se encontraran sumamente relacionados en cuanto a su función (Tabla 6). Por lo descrito anteriormente parece ser de gran relevancia biológica en el mundo bacteriano controlar de manera correcta la adquisición y concentración de hierro, al nivel de cada uno de los componentes de esta compleja maquinaria. El análisis del COG1595 resulta un caso representativo en donde la presión de selección ha sido suficiente para conservar un mecanismo de regulación de manera conservada y extensa.

Tabla 6. COGs obtenidos a partir del MAST con el motivo del COG1595 para Proteobacteria

COGs relacionados al transporte de hierro recuperados por el COG1595	
COG3712	Fe dicitrate sensor
COG1629	Putative ferric siderophore receptor
COG0735	Ferric uptake regulator, FUR family. Fe / Zn
COG1918	Fe transport system protein A. FeoA
COG0370	Ferrous iron transport protein B. FeoB
COG0614	Iron compound ABC transporter type Fe3+.
COG4594	ABC type Fe3+ citrate transport system
COG0716	Flavodoxins
COG0803	ABC type metal ion transport periplasmic component
COG0577	ABC type antimicrobial peptide transport, permease, defense
COG3182	Uncharacterized iron-regulated membrane protein
COG0609	ABC cobalamine Fe3+siderophore transporter
COG4559 / 58	ABC cobalamine Fe3+siderophore transporter
COG4772 / 73	Outer membrane receptor for ferric coprogen and ferric rhodotorulic acid, siderophore
COG1840	ABC Fe3+ siderophore transporter, periplasmic binding protein
COG0834	ABC aminoacid transport / signal trasnduction systems, periplasmic
COG0783	DNA protection during starvation conditions. DNA binding ferritin
COG3721	Putative heme iron utlization protein. Inorganic ion transport
COG4606	ABC type iron enterochelin transport, permease component
COG4592	ABC type Fe2+ enterobactin transport system
COG4264	Siderophore synthetase component
COG1120	ABC type cobalamin / Fe3+ siderophores transport
COG2906	Bacterioferritin-associated ferredoxin
COG2193	Bacterioferritin cit b1
C0G0444	ABC type dipeptide oligopeptide/nickel transport

2.1. Estrategia Experimental

El planteamiento inicial era obtener un gran número de mutaciones a lo largo de los nucleótidos que conforman la estructura tallo-asa que enmascara el SD en el termoriboswitch ROSE, la idea fue seleccionar a aquellas mutantes que funcionaran de manera análoga a ROSE, es decir que a 30°C el termómetro de RNA estuviera cerrado (apagando la traducción) y a temperaturas mayores (37-42°C) estuviera abierto permitiendo la traducción.

Para esto se diseñó un sistema reportero con doble selección, el cual comprende la fusión de la estructura mutante a una proteína reportera, cloranfenicol acetil-transferasa (CAT), que a su vez está fusionada a otro gen, *sacB*. Para asegurar que la expresión genética de ambos genes, *cat* y *sacB*, estuvieran bajo la misma regulación traduccional de nuestras variantes del elemento ROSE, se realizó una fusión traduccional CAT-SacB con un linker de 8 aminoácidos los cuales quedan en fase y fusionados al gen *sacB* (Figura 24). Este último gen funcionaría como contra selección ya que la expresión de SacB en un medio de crecimiento con sacarosa produce un intermediario tóxico para bacterias Gram-negativas: levano-sacarosa, un azúcar de alto peso molecular (alrededor de 60,000 moléculas de D-fructosa) (Figura 25), que se acumula en el periplasma de *E. coli* (Briones, 2008).

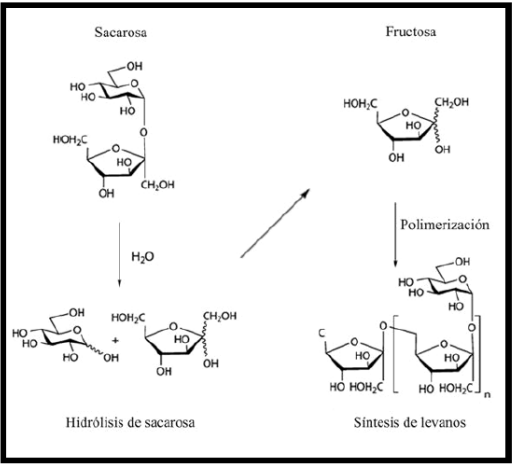


Figura 25. Reacciones catalizadas por la levansacarasa SacB. Tomada de (70).

La primera parte dentro de la estrategia experimental consistió en construir el control positivo del sistema genético usando la estructura de ROSE silvestre de *Escherichia coli* K12 para regular la expresión de la fusión traduccional de los genes *cat* y el gene *sacB*, (Figura 24). Esta fusión *cat-sacB* fue utilizada para comprobar que la estructura ROSE, colocada al principio de la fusión, monitoreaba las fluctuaciones en la temperatura de la manera antes mencionada (para ver una explicación detallada ver Figura 28).

Es importante notar que una de nuestras hipótesis de trabajo es que el producto peptídico resultante de la fusión de Cat y SacB unido por el péptido conector Gly-Gly-Ser-Ser, será funcional simultáneamente para las actividades de cloranfenicol acetil-transfera y levan-sucrasa. Esta hipótesis está basada en diversos reportes en donde el gen *cat* es utilizado como reportero de plegamiento cuando se encuentra fusionado a un segundo gene en estudio (Maxwell et al., 1999;Seitz et al., 2007).

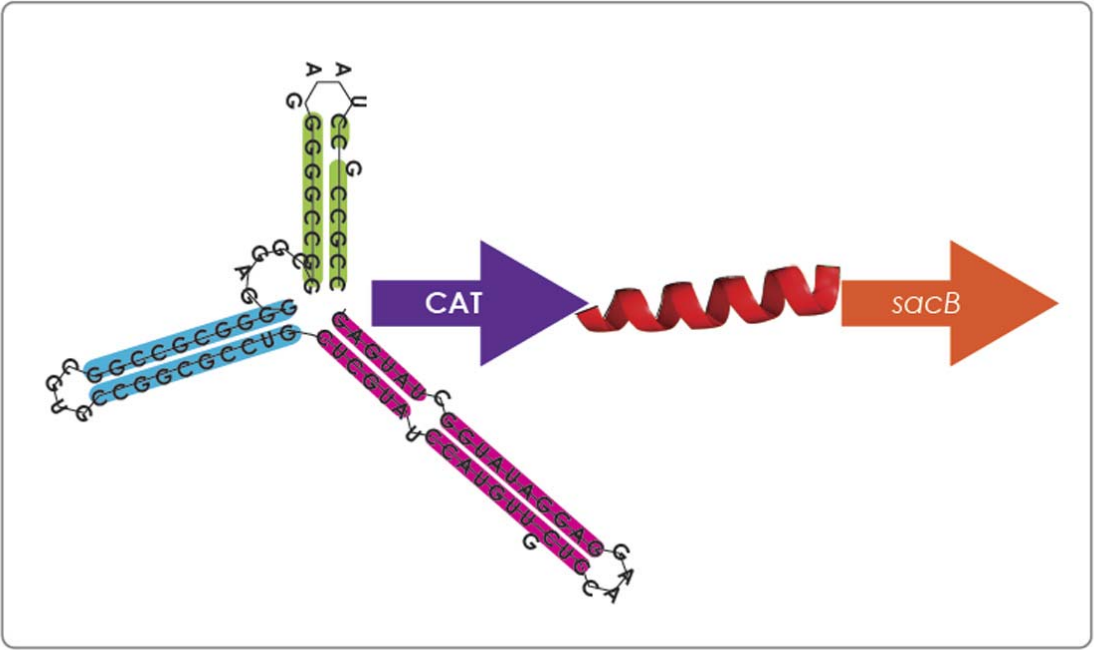


Figura 24. Fusión traduccional bajo el control de ROSE. En rojo se muestra el linker entre los genes *cat* y *sacB*.

2.2. Metodología Experimental

2.2.1. Amplificación de los insertos por PCR: ROSE, cat y sacB.

Las reacciones de PCR se llevaron a cabo en 50 o 100 µL con agua tetradestilada estéril, Buffer mix de reacción AccuPrime™ Pfx 10X (incluye los dNTP's) (Invitrogen), MgSO₄ 50 mM, templado a una concentración de 20-100 ng/µL, 1.25 U de DNA polimerasa AccuPrime™ Pfx (Invitrogen) y finalmente los oligonucleótidos a una concentración final de 20 pmol/µL. La temperatura de hibridación de los oligonucléotidos para todos los casos se determinó mediante un gradiente por PCR.

Las muestras fueron incubadas a 94°C por 5 minutos, 25 ciclos de 94°C 40 segundos, Tm específica de cada par de oligonucleótidos (Tabla 7) 40 segundos y 68°C dependiendo del número de bases del producto de PCR esperado se calculó el tiempo con la relación 1 minuto por 1 kilobase y una incubación final de 68°C por 7 minutos. Los productos de PCR se purificaron por el kit de purificación de PCR (Roche).

Gen	Tm (°C)	Oligo complementario/reverso complementario
ibpA ROSE	58	CATCTGTTGATCGTGGGTGTTGGCCT
		CAT AATCAATAGCTCCTGAAATCAGCGAG
Cat	58	GAGAAAAAATCACTGGATATACCACCGT
		AGAACCGCCACCGC CCCCCGCCCTGCCACTCATCG
sacB	60	GGTGGGGGCTCG AACATCAAAAAGTTTGCAAACAAGCAC
		TTGCCTGCTTTTATATAGTTCATATGGG

Tabla 7. Oligonucleótidos empleados en esta tesis.

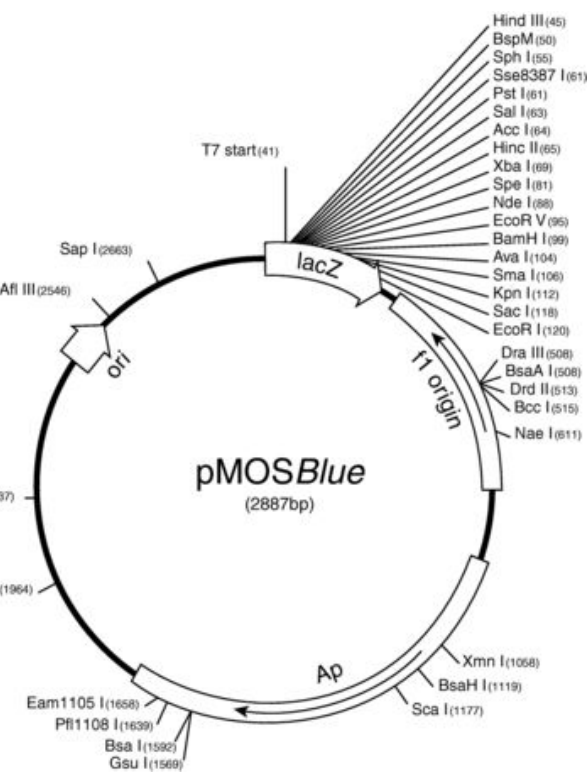
2.2.2. Reacción de fosforilación-ligación entre ROSE y cat.

Se llevó a cabo la reacción de fosforilación del producto de PCR ROSE en un volumen final de 25 µL, adicionando 10 µL con agua tetradestilada estéril, Buffer 10X T4 DNA Ligasa (New England Biolabs) (este buffer es compatible con esta reacción de fosforilación y se puede usar en ambas reacciones fosforilación/ligación), DNA 70-100 ng/µL, 1 U de T4 DNA kinasa. (New England Biolabs) La reacción sin la adición de enzima se incubó a 70°C 5 minutos, se enfrió a 4°C 1:30 minutos y se le agregó la enzima para iniciar la reacción de fosforilación, la cual fue incubada a 37°C 2 horas y finalmente se inactiva la enzima a 65°C 20 minutos.

Posteriormente se llevó a cabo la reacción de ligación, entre los productos ROSE y cat en una relación molar 1:1; a la reacción de fosforilación se le agregó nuevamente agua tetradestilada estéril para llevar la reacción a un volumen final de 50 µL, la cantidad proporcional de Buffer 10X T4 DNA Ligasa (New England Biolabs), el DNA que se desea ligar 70-100 ng/µL, en este caso el gen cat y 2 U de T4 DNA ligasa (New England Biolabs) y se incubó la reacción de ligación a 25°C 12 horas. El producto de ligación ROSE-cat se purificó por el kit de purificación de PCR (Roche).

Se llevó a cabo un PCR de este producto para amplificarlo en la dirección correcta (esperando un fragmento de aproximadamente 970 nucleótidos), con los oligonucleótidos complementario para amplificar ROSE y el reverso complementario para *cat*, con una T_m de 58°C; bajo las mismas condiciones descritas previamente y se purificó de la misma manera. Este producto se mandó a secuenciar para verificar la dirección así como la inexistencia de mutaciones que modificaran el marco de lectura o introdujeran mutaciones no deseadas.

2.2.3. Clonación del producto ROSE-cat en el vector pMOS-Blue.



El producto de PCR fue clonado usando el kit de clonación pMOS-Blue para extremos romos (Amersham Pharmacia Biotech) el cual precisamente está diseñado para la clonación de cualquier tipo de fragmentos de DNA, estén purificados o no, aunque la eficiencia aumenta cuando son purificados. El inserto es fosforilado mediante un cóctel enzimático, seguido de una breve incubación en calor, el producto es ligado en el vector pMOS desfosforilado y romo. La transformación en células competentes altamente eficientes resulta en un 75% de colonias recombinantes con el inserto. El vector se encuentra linearizado en el sitio de restricción *EcoRV* del gen *lacZ* por lo cual una clonación y selección exitosa resulta en la formación de colonias blancas en un medio con x-gal (Figura 26).

Figura 26. Vector de clonación pMOS-Blue (Amersham) empleado para la construcción.

El inserto se fosforila mediante la reacción pK en la cual se adiciona agua tetradesilada estéril para llevar la reacción a un volumen final de 10 µL, Buffer 10X pK, DTT 100 mM (agente reductor), cóctel de enzimas pK, inserto ROSE-cat 100-500 ng/µL. La reacción se incubó a 22°C 40 minutos, después se centrifugó a 13,000 rpm 20 segundos para colectar el contenido que se encuentra en la parte de superior del tubo, se inactivó la reacción a 75°C 10 minutos y se enfrió en hielo 2 minutos.

Después se lleva a cabo la reacción de ligación con agua tetradesilada estéril la suficiente para llevar la reacción a un volumen final de 12 µL, el producto de la reacción pK (ROSE-cat fosforilado), vector pMOS-Blue 50 ng/µL. Se incubó a 22°C 2 horas. La ligación se purificó con el

kit de purificación de PCR (Roche). 5 μ L de reacción de ligación se transformaron en 100 μ L de células electrocompetentes de la cepa *E. coli* DH5 α . Se recuperaron en 1 mL de medio SOC a 37°C 1 hora y se sembraron en cajas de LB sólido con ampicilina 200 μ g/mL, X-gal 50mg/mL e IPTG 24 mg/mL a 37°C ON.

Se seleccionaron las colonias blancas y se les purificó el plásmido por el kit de purificación de plásmidos (Wizard), usando éstos como templado se realizaron PCRs para comprobar y amplificar el inserto ROSE y *cat*. Como paso final del protocolo, se mandaron a secuenciar con el oligonucleótido universal T7.

2.2.4. Apertura del vector pMOS-Blue ROSE-cat por PCR, Clonación de *sacB* a través de la fosforilación-ligación.

Se puede amplificar mediante un PCR el inserto y vector que se desean ligar. Previo a la ligación el inserto es fosforilado. Por lo general los oligonucleótidos para amplificar fragmentos de PCR carecen de fosfatos en el extremo 5' y el vector no puede recircularizarse a menos que sea ligado al inserto fosforilado (Rodríguez, 2007) (Figura 27).

El vector se linearizó mediante PCR inverso que consiste en amplificar al vector de clonación mediante el uso de oligonucleótidos por PCR en direcciones divergentes. Este último se amplificó con una enzima diferente que permite amplificar fragmentos de gran tamaño con alta fidelidad, la reacción de PCR se llevó a cabo en 100 μ L con agua tetradestilada estéril, Buffer de amplificación *Pfx* 10X (Invitrogen), dNTP's 10mM, MgSO₄ 50 mM, 1.25 U de DNA polimerasa Platinum^R *Pfx* (Invitrogen) y finalmente los oligonucleótidos a una concentración final de 20 pmol/ μ L y templado a una concentración de 20-100 ng/ μ L (vector pMOS-Blue ROSE-cat cuya secuencia fuera correcta). Los oligonucleótidos usados fueron el oligonucleótido complementario diseñado específicamente para la apertura del plásmido y el oligonucleótido reverso complementario para amplificar *cat*.

La muestra fue incubada a 94°C por 7 minutos, 25 ciclos de 97°C 30 segundos, Tm 63°C 40 segundos y 68°C 1:45 minutos y una incubación final de 68°C por 7 minutos. El producto de PCR se purificó por el kit de purificación de PCR (Roche).

Posteriormente se fosforiló el producto de PCR *sacB* usando los mismos reactivos y bajo las mismas condiciones de incubación que se describieron previamente para la fosforilación de ROSE y se llevó a cabo la reacción de ligación, entre los productos pMOS-Blue ROSE-cat y *sacB* fosforilado en una relación molar 2:1, como se describió en la reacción de ligación entre ROSE y *cat*. El producto de ligación pMOS-Blue ROSE-cat-*sacB* se purificó por el kit de purificación de PCR (Roche).

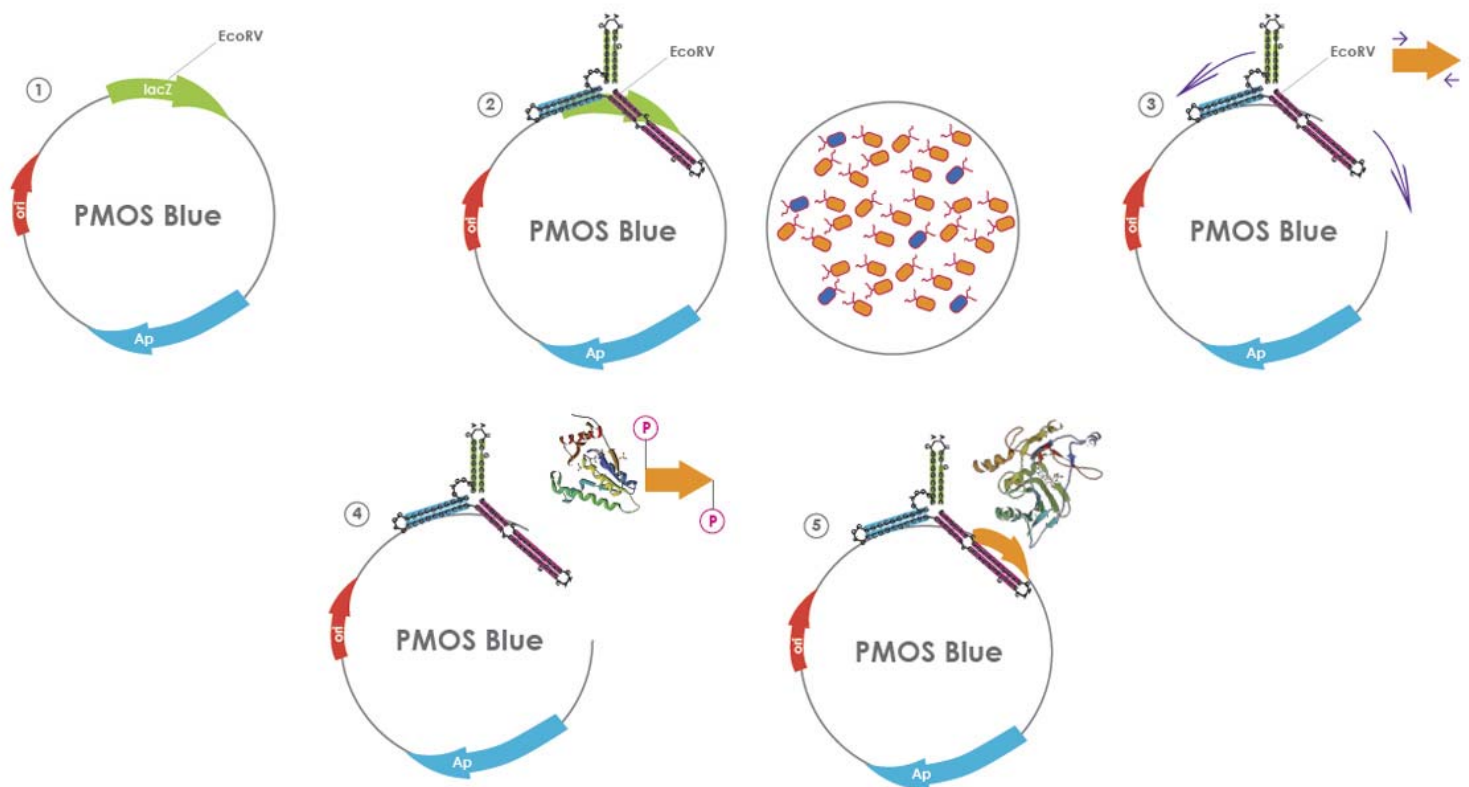


Figura 27. Protocolo general de clonación empleado en esta tesis. 1) El kit de clonación incluye al vector pMOS-Blue el cual tiene el gen de resistencia a ampicilina y *lacZ*, por el cual se podrán seleccionar las colonias recombinantes. 2) El uso de este kit permite clonar un fragmento en el polilinker específicamente en *EcoRV*, sobre el gen de *lacZ*, las colonias con el plásmido de interés son blancas y son seleccionadas. 3) Una vez que se ha clonado el fragmento deseado es posible clonar otro a partir de abrir el plásmido haciendo un PCR inverso con los oligonucleótidos que lo amplifiquen en su totalidad, y otro PCR que amplifique el fragmento de interés (se muestra en naranja). 4) El fragmento a clonar es fosforilado por la T4 polinucleótido kinasa. 5) Se lleva a cabo una reacción de ligación entre el fragmento fosforilado y el vector, posterior a esto se puede seleccionar el plásmido por las cualidades propias del fragmento o por la resistencia a ampicilina de pMOS-Blue. Este mismo esquema general puede seguirse para cualquier fragmento que se quiera introducir a continuación.

2.2.5. Ensayos de crecimiento para probar y seleccionar la construcción de interés.

Para probar que ROSE wt controlaba la expresión de forma dependiente a la temperatura de la construcción genética, se hicieron ensayos funcionales basados en el crecimiento bacteriano. Una vez que se obtuvo la primera construcción pMOS-Blue ROSE-cat con la secuencia correcta, se transformaron 10 μ L de este vector, en 100 μ L de células electrocompetentes de la cepa *E. coli* DH5 α , las cuales se recuperaron en 1 mL de medio SOC 1 hora 37°C y se sembraron 100 μ L, en cajas con medio LB sólido con cloranfenicol 25 μ g/mL a 30°C, 37°C y 42°C ON; como control de este experimento se transformó en 100 μ L de células electrocompetentes de la cepa *E. coli*

DH5 α 5 μ L de plásmido pBR329, el cual también contiene el gen de resistencia a cloranfenicol, se recuperó en medio SOC 1 hora 37°C, incubando a las mismas 3 temperaturas ON.

La construcción completa pMOS-Blue ROSE-*cat-sacB* (10 μ L) se transformó en 100 μ L de células electrocompetentes *E. coli* DH5 α y se recuperó en medio SOC a 37°C 1 hora, se sembraron 100 μ L en medio LB sólido con ampicilina 200 μ g/mL a 37°C ON. Posteriormente se tomaron colonias que fueron resembradas en tres medios distintos: i) LB sólido con cloranfenicol 25 μ g/mL, ii) LB sólido con sacarosa 6% (wt/vol) (concentración final) y iii) LB sólido con cloranfenicol y sacarosa a las mismas concentraciones finales; y finalmente LB sólido con ampicilina 200 μ g/mL. Estas siembras fueron incubadas a 30°C, 37°C y 42°C ON.

Aquellas colonias que crecieron de la manera que se esperaba (Figura 28) fueron seleccionadas y crecidas en 2mL de medio LB líquido con ampicilina 200 μ g/mL ON a 37°C. Posteriormente se llevó a cabo una dilución con las colonias candidatas de 100 μ L de células por 900 de medio líquido LB, hasta la cuarta dilución 100 μ L de células por 900 μ L LB, creciendo las células en cajas de LB sólido con cloranfenicol 25 μ g/mL a 30°C, 37°C y a 42°C ON. Las diluciones se hicieron para reducir el número de colonias esperadas, facilitando la visualización de colonias en las cajas, y principalmente para poder cuantificar el número de colonias presentes y así poder establecer una relación entre el número de colonias por caja crecida a estas tres temperaturas y la eficiencia de la construcción genética.

A la par de los ensayos de crecimiento funcionales a las colonias candidatas se les purificó el plásmido por lisis alcalina haciendo minipreps según Sambrook y colaboradores (1989). Estos plásmidos se usaron como templado para llevar a cabo PCRs que amplificaran al inserto *sacB*, la región *cat-sacB* y ROSE-*cat-sacB*, usando los respectivos oligonucleótidos, con el fin de verificar si se encontraba el inserto *sacB*, así como su direccionalidad. Los plásmidos con el inserto verificado por PCR, fueron purificados a través del kit de purificación de plásmidos (Wizard) y secuenciados en la unidad de secuenciación del Instituto de Biotecnología. UNAM. La secuenciación se realizó con diferentes oligonucleótidos para poder obtener toda la secuencia de la región que incluye a ROSE-*cat-sacB*.

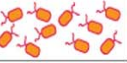
ROSE wt	30°C	37-42°C	siempre abierto	30°C	37-42°C	siempre cerrado	30°C	37-42°C
LB agar + cm	X		LB agar + cm			LB agar + cm	X	X
LB agar + sacarosa 6%		X	LB agar + sacarosa 6%	X	X	LB agar + sacarosa 6%	X	X
LB agar + cm + sacarosa 6%	X	X	LB agar + cm + sacarosa 6%	X	X	LB agar + cm + sacarosa 6%	X	X

Figura 28. Modo de selección con la construcción pMOS-Blue ROSE-cat-sacB de aquellas mutantes que funcionan en termo-monitoreo de forma análoga a ROSE. En la primera tabla se muestran los resultados que se esperan con la expresión de la construcción pMOS-Blue ROSE-cat-sacB, cuando ROSE es la forma silvestre. En presencia de cloranfenicol a 30°C no hay crecimiento debido a que ROSE se encuentra cerrado y así, la fusión traduccional *cat-sacB* no se traduce por lo que no hay resistencia al antibiótico. A 42°C el termómetro se abre y el ribosoma puede unirse al SD del mensajero y de este modo se traduce la fusión traduccional *cat-sacB* que confiere resistencia al antibiótico permitiendo que las bacterias pueden crecer. En un medio con sacarosa a 30°C hay crecimiento bacteriano, debido a que el termómetro se encuentra cerrado y no se da la expresión de la fusión traduccional *cat-sacB* y las células no mueren porque no hay contraselección; por el contrario a 42°C las células no crecen debido a que el gen contraselectivo se expresa y se encuentra en presencia de sacarosa que es tóxica para las bacterias recombinantes. Lo mismo ocurre con las células al colocarlas en un medio adicionado con cloranfenicol y sacarosa en ambas temperaturas, no se espera crecimiento. Al realizar el banco de mutantes, se desea seleccionar sólo aquellas que se comportan como en teoría deberían de comportarse las construcciones con ROSE silvestre, de modo tal que la construcción genética sirve para reducir y seleccionar efectivamente a éstas, las siguientes dos tablas se refieren al modo en el que ocurre esta selección para el caso de las mutantes defectivas en el termo-monitoreo. En el caso de presentar el termómetro abierto a ambas temperaturas (siempre abierto) (tabla central) en un medio con cloranfenicol, crecerán a las 2 temperaturas, pero al adicionar sacarosa al medio, el sistema funcionará contra-seleccionándolas ya que estas expresan *sacB*, de tal modo que estas mutantes no se seleccionan desde un principio. En el caso de las mutantes de ROSE que a ambas temperaturas se encuentran cerradas, tampoco crecerán desde un principio, debido a que nunca ocurre la traducción de la fusión traduccional *cat-sacB*.

2.2.6. Otras consideraciones metodológicas

La síntesis de oligonucleótidos y las secuenciaciones se realizaron en la Unidad de Síntesis del Instituto de Biotecnología, UNAM. Para visualizar productos de PCR o purificarlos por gel, se emplearon geles de agarosa DNA (Invitrogen) al 1% (en relación inversa al tamaño de la secuencia que se desea visualizar) en solución amortiguadora TAE 1X (Sambrook *et al.*, 1989) Las electroforesis se efectuaron a 100V. Las células electrocompetentes *E. coli* DH5α se prepararon según Sambrook *et. al.* (1989) usando medio YENB (bacto yeast extract 0.75%, bacto nutrient Broth 0.85%). A continuación se presenta su fenotipo: huA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17.

2.3. Resultados y Discusión Experimental

2.3.1. Amplificación de los insertos por PCR: ROSE, *cat* y *sacB*.

En *E. coli* K12, ROSE forma parte de la región intergénica de una proteína pequeña de choque térmico del gen *ibpA* (GI: 16131555). Se mapeó el promotor en dicha región y se mandó a sintetizar el par de los oligonucleótidos para amplificar 300 nucleótidos que incluyen la región que comprende a ROSE y el codón de inicio AUG. El templado usado para amplificar ROSE por PCR fue el DNA cromosomal total de *E. coli* K12.

El gen *cat* codifica para una enzima homotrimérica de 25 KDa la cual inactiva al cloranfenicol, éste último es un antibiótico que se une a la subunidad grande del ribosoma (50s) y bloquea a la peptidiltransferasa. En primer instancia se empleó este gen para seleccionar a las clonas que presentes en cloranfenicol adicionado al medio, fueran resistentes y que termomontorearan correctamente (comportamiento ROSE wt), de tal modo que *cat* funciona como gen de resistencia y adicionalmente como un gen reportero.

El casete de resistencia a cloranfenicol (656 nucleótidos) se obtuvo del plásmido pBR325. El oligonucleótido reverso complementario se diseñó para que la región amino terminal de *cat* tuviera parte del linker traduccional Gly-Gly-Ser-Ser, cuidando que los codones codificantes de estos aminoácidos tuvieran una frecuencia de uso alto en *E. coli*.

SacB es un enzima levanosucrasa que polimeriza al fructano, un producto del catabolismo de la sacarosa, que en el periplasma de las bacterias Gram negativas resulta ser tóxico. Así, cuando las bacterias son crecidas en presencia de sacarosa la expresión de SacB es tóxica para éstas (Edwards et al., 1998). Usamos el gen *sacB* para llevar a cabo una contra-selección: en un banco de estructuras ROSE mutantes, este sistema se puede usar con el fin de eliminar aquellas mutantes que siempre termomontorearan incorrectamente debido a que el termómetro de RNA siempre se encontrara abierto, mientras que el gen *cat* en este doble sistema selectivo/contra-selectivo es el primer filtro de aquellas estructuras mutantes que siempre se encuentren cerradas. (Para una explicación detallada ver Figura 28).

sacB (1421 nucleótidos) se amplificó tomando como molde al plásmido pRE112. De la misma forma que se diseñaron los oligonucleótidos para *cat*, el oligonucleótido complementario presenta la otra parte repetida del linker (Gly-Gly-Ser-Ser). Se eliminó el codón de inicio de este segundo gene ya no es más necesario porque se encuentra fusionado traduccionalmente a un primer gene; en este caso *sacB*. El oligonucleótido reverso complementario a *sacB* se diseñó unos cuantos pares de bases después de su correspondiente codón de término.

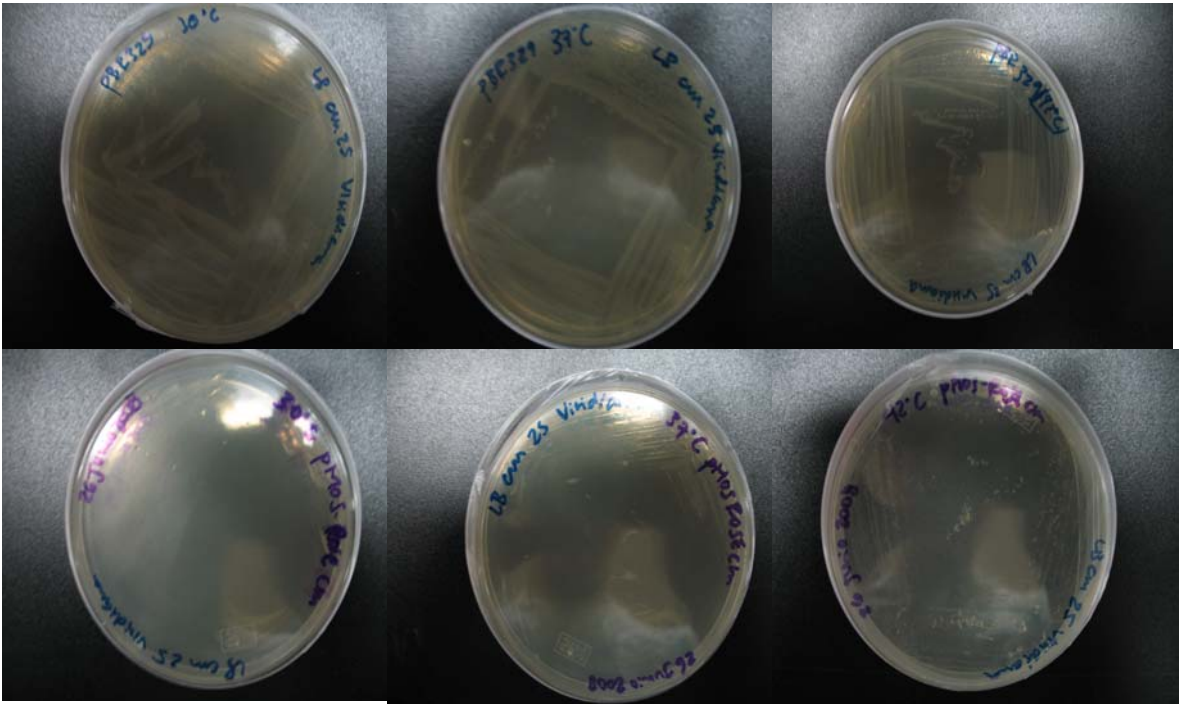
2.3.2. La expresión de la construcción pMOS-Blue ROSE-cat está termorregulada por ROSE.

La clonación del fragmento ROSE-cat usando el kit de clonación pMOS-Blue fue exitosa, resultando en un gran número de colonias blancas y sólo un 10% de colonias azules. Por lo cual se seleccionaron 10 colonias, de las cuales al mandar a secuenciar solamente una tenía el inserto completo sin mutaciones, ni inserciones. Al crecer esa colonia en LB sólido y líquido en presencia de cloranfenicol 25 µg/mL a 30°C, 37°C y 42°C, efectivamente se observaba que no crecían a 30°C, medianamente a 37°C y a 42°C crecían normalmente. El crecimiento se comparó con un control que consistía en el mismo experimento pero con células transformadas con el plásmido pBR329, el cual también presenta el gen de resistencia a cloranfenicol que se expresa de manera constitutiva. Estos resultados, aunado a la secuencia de nucleótidos, llevan a la conclusión de que la expresión genética en dicha construcción se encuentra regulada a nivel de la traducción, a través de cambios en la conformación termo-dependientes en ROSE (Figura 29).

En estudios previos con el termómetro ROSE, como ya se había mencionado anteriormente, se realizaban fusiones de ROSE, usando como sistema reportero *lacZ* (Waldminghaus et al., 2005). Uno de los problemas radicaba en que este sistema no es capaz de monitorear la expresión genética a temperaturas arriba de 40°C. Así que ROSE sólo había sido probado a 30°C y a 37°C en donde la expresión es menor que a la temperatura real de choque térmico por calor para *E. coli*. A partir del sistema desarrollado en nuestro estudio, se pudo determinar que la estructura de ROSE que se forma a los 42°C contribuye a la expresión adecuada y total del gen reportero, en nuestro caso *cat* cuyo producto peptídico es estable a estas temperaturas.

Otra de las observaciones interesantes a mencionar es que durante el diseño con oligonucleótidos por PCR de nuestra construcción genética, se fusionó ROSE al gen *cat* directamente, siendo que los reportes y los análisis experimentales previos hechos con ROSE hasta la fecha, consideran de relevancia debido a que se observa una expresión mayor del gen reportero al colocar los primeros nucleótidos del gen *ibpA*, el cual es regulado originalmente por ROSE, seguidos de la secuencia nucleotídica del gen que funciona como reportero. En nuestro estudio no se colocaron estos primeros nucleótidos, en parte para probar que los termómetros de RNA son elementos fácilmente intercambiables que pueden ser colocados como regiones regulatorias de cualquier gen, no importando la secuencia que se coloque río abajo.

Nuestra construcción al parecer es prueba de que no es necesario colocar los primeros nucleótidos de *ibpA*, para tener una eficiencia traduccional mayor o una regulación óptima, y que en última instancia es suficiente la estructura ROSE para controlar de manera dependiente del calor la expresión de genes río abajo de dicho elemento.



ROSE wt	30°C	37-42°C
LB agar + cm	X	

Figura 29. Crecimiento bacteriano de las células con la construcción genética pMOS-Blue ROSE-cat a 30, 37 y 42°C. Figura Arriba. Resultados del experimento de crecimiento que demuestran que en nuestra construcción es el elemento ROSE termorregula a *cat*. Fotografías. Arriba. Crecimiento de colonias transformadas con el plásmido pBR329 a las tres temperaturas, se observa crecimiento en todas las cajas (Control positivo). Abajo. Ensayo funcional en células transformadas con la construcción genética pMOS-Blue ROSE-cat. A 30°C no se observa crecimiento a 37 y 42°C se observa crecimiento. Todas las cajas contienen LB agar con cloranfenicol.

2.3.3. Construcción genética pMOS-Blue ROSE-cat-sacB

Al transformar el plásmido pMOS-Blue ROSE-cat-sacB en células electrocompetentes de *E. coli* DH5α se obtuvieron 41 colonias, las cuales fueron resembradas en diferentes medios (LB sólido) y temperaturas, ON (tal como se describe en la metodología experimental). Los resultados se muestran en la Tabla 8.

De 41 colonias, únicamente 5 de ellas (Tabla 8, señaladas en naranja) presentan parte del comportamiento esperado, esto es que a una temperatura de 30°C en un medio de crecimiento que contiene cloranfenicol y en cloranfenicol con sacarosa, no se presenta crecimiento porque el termómetro de RNA se encuentra cerrado y por la misma razón en un medio de crecimiento que contiene sacarosa puedan crecer debido a que no ocurre la expresión de *sacB*. Por el contrario a 42°C crecen en cloranfenicol, y no así en cloranfenicol con sacarosa, ya que a esta temperatura

ROSE se abre y se da la traducción de *cat* y *sacB*, este último gen, como ya se ha mencionado produce a la proteína SacB, la cual participa en la degradación de la sacarosa generando un intermediario tóxico para las células y consecuentemente estas mueren.

Al crecer estas colonias en un medio con sacarosa a las tres temperaturas esperábamos que a 30°C crecieran, ya que el termómetro de RNA, se encuentra cerrado y no se expresa la proteína SacB, y que por el contrario a temperaturas de 37°-42°C, no se observara crecimiento ya que el termómetro se encuentra abierto y SacB se produce y como se mencionó anteriormente, en presencia de sacarosa se generan intermediarios tóxicos para las células. Sin embargo al realizar los ensayos de crecimiento en el medio con sacarosa, en ninguna de las 41 colonias crecidas a las tres temperaturas de 30°, 37° y 42°C, se observó el comportamiento contra-selectivo esperado, siempre se observó crecimiento.

Estos ensayos de crecimiento se repitieron al aislar el subconjunto de las 5 colonias, creciendo a éstas por duplicado (Tabla 9). Como se puede observar, la única colonia que se pudo mantener fue la colonia 8, el resto no creció de la manera como lo habían hecho originalmente desde el primer experimento, hecho que nos hace pensar que al sembrar y repetir los ensayos de crecimiento no se realizó de manera adecuada y no se logró inocular la colonia, cabe señalar que aunque la tabla muestra resultados por duplicado, la siembra de estas colonias se repitió al menos 10 veces y los resultados siempre fueron variables; así durante su constante manipulación se trabajó con una gran cantidad de cajas a diferentes temperaturas, con distintos componentes en el medio LB, considerando las repeticiones hechas para cada una de estas caja, así como el numero de diluciones, es posible que se generaran mutaciones espontáneas, que provocan dichas variaciones en el crecimiento, y por tanto no exista la reproducibilidad en estos ensayos.

Asimismo se observa que la colonia 8 siguió creciendo en sacarosa indiscriminadamente, no obstante, en medio con cloranfenicol creció de acuerdo a lo que se esperaba bajo el control de ROSE (a 30°C no creció, mientras que a 37°C-42° hubo crecimiento), lo cual nos hizo pensar que si estaba presente la resistencia CAT y que ésta se regulaba a nivel de la traducción por ROSE. Cuando esta misma colonia fue crecida en cloranfenicol con sacarosa, el crecimiento fue igual al que esperábamos en un medio con sacarosa, esto es, que a 30°C el termómetro de RNA estuviéase cerrado y SacB no se expresará, y por tanto las células sobrevivieran, y a temperaturas mayores no se detectara crecimiento debido a que el termómetro de RNA se encuentra abierto, produciéndose SacB y con esto los levamos tóxicos para *E. coli*. Una posible explicación para el crecimiento en el medio con sacarosa, y el crecimiento en medio con sacarosa y cloranfenicol, es que la expresión contraselectiva de *sacB* en un medio con sacarosa, por si sola, no este sujeta a una presión que seleccione en este caso al plásmido, como la que ocurre por lo general con el uso de resistencias a algún antibiótico; la mayoría de los estudios que han empleado al gen *sacB* como un marcador de contra-selección incorporan un gen de resistencia a algún antibiótico, y en

última instancia crecen el cultivo en un medio que contiene tanto sacarosa, como el antibiótico de resistencia.

Sin embargo los resultados anteriormente descritos no fueron concluyentes para comprobar la funcionalidad de la fusión a nivel de la traducción de ROSE-cat-sacB, y en última instancia para demostrar la contra-selectividad de la expresión genética de SacB.

Colonias candidatas. Construcción genética pMOS-Blue ROSE-cat-sacB												
Colonia	30°C				37°C				42°C			
	cm	cm + sac	sac	ap	cm	cm-sac	sac	ap	cm	cm- sac	sac	ap
1	*	*	*	*	*	*	*	*	*	*	*	*
2	*	*	*	*	*	*	*	*	*	*	*	*
3	-	-	*	*	*	-	*	*	-	-	*	*
4	-	-	*	*	-	-	*	*	-	-	*	*
5	-	-	*	*	-	-	*	*	-	-	*	*
6	-	-	*	*	-	-	*	*	-	-	*	*
7	-	-	*	*	-	-	*	*	-	-	*	*
8	-	-	*	*	-	-	*	*	-	-	*	*
9	-	*	*	*	-	-	*	*	-	-	*	*
10	-	-	*	*	-	-	*	*	-	-	*	*
11	*	*	*	*	*	*	*	*	*	*	*	*
12	-	-	*	*	*	*	*	*	*	*	*	*
13	-	-	*	*	-	-	*	*	-	-	*	*
14	*	*	*	*	*	*	*	*	*	*	*	*
15	-	-	*	*	-	-	*	*	-	-	*	*
16	*	*	*	*	*	*	*	*	*	*	*	*
17	*	-	*	*	*	*	*	*	*	*	*	*
18	-	-	*	*	-	-	*	*	-	-	*	*
19	*	*	*	*	*	*	*	*	*	*	*	*
20	-	-	*	*	*	-	*	*	*	-	*	*
21	*	-	*	*	*	*	*	*	-	*	*	*
22	-	-	*	*	-	-	*	*	-	-	*	*
23	-	-	*	*	-	-	*	*	-	-	*	*
24	-	-	*	*	-	-	*	*	-	-	*	*
25	-	*	*	*	-	*	*	*	*	*	*	*
26	-	*	*	*	-	*	*	*	*	*	*	*
27	-	-	*	*	-	-	*	*	-	-	*	*
28	-	-	*	*	-	*	*	*	-	-	*	*
29	-	-	*	*	-	-	*	*	-	-	*	*
30	-	*	*	*	-	*	*	*	-	*	*	*
31	-	-	*	*	-	*	*	*	-	*	*	*
32	-	*	*	*	*	*	*	*	*	*	*	*
33	-	-	*	*	-	-	*	*	-	-	*	*
34	-	-	*	*	-	-	*	*	-	-	*	*
35	*	*	*	*	*	-	*	*	*	*	*	*
36	-	-	*	*	*	-	*	*	*	*	*	*
37	-	-	*	*	-	-	*	*	-	-	*	*
38	-	*	*	*	*	*	*	*	*	*	*	*
39	-	-	*	*	-	-	*	*	-	-	*	*
40	-	*	*	*	*	*	*	*	*	*	*	*
41	-	-	*	*	-	*	*	*	-	*	*	*

Tabla 8. cm = LB sólido + cloranfenicol 25 µg/mL; cm + sac = LB sólido + cloranfenicol 25 µg/mL y sacarosa 6%; sac = LB sólido + sacarosa 6%; ap = LB sólido + ampicilina 200 µg/mL; * = crecimiento de la colonia; - = no crecimiento de la colonia.

	30°C			37°C			42°C		
Colonia	cm	cm + sac	sac	cm	cm-sac	sac	cm	cm- sac	sac
6	--	--	**	--	_*	**	--	--	**
7	--	--	**	--	--	**	--	--	**
8	--	**	**	**	--	**	**	--	**
10	--	--	**	--	--	**	--	--	**
20	--	_*	**	--	--	**	--	--	**

Tabla 9. Resultados por duplicado. cm = LB sólido + cloranfenicol 25 µg/mL; cm + sac = LB sólido + cloranfenicol 25 µg/mL y sacarosa 6%; sac = LB sólido + sacarosa 6%; * = crecimiento de la colonia; - = no crecimiento de la colonia.

A la par de que se realizaron los ensayos funcionales con este grupo de colonias (5), se llevaron a cabo PCRs para determinar por tamaño si se encontraba el gen *sacB*, y otros para amplificar la región comprendida de *sacB* a ROSE y la región entre *sacB* y *cat* con el fin de verificar la direccionalidad de la construcción genética (Figura 30).

En las colonias analizadas, las únicas que presentaban el inserto *sacB* se encontraba tanto eran la colonia 8 como en la 10, en el caso del PCR *sacB*-ROSE y *sacB*-*cat* sólo se observan las bandas esperadas para la colonia 8 (Figura 30). De tal modo que la colonia 8 se mandó a secuenciar para verificar que la construcción genética no tuviera mutaciones, el sentido del gen *sacB* y ensamblar toda la secuencia, la cual comprendía desde ROSE hasta *sacB*. Los oligonulceótidos usados para la secuenciación fueron el U19 y el T7, ya que son oligonucleótidos universales de gran utilidad que amplifican el vector pMOS-Blue por fuera esto es flanqueando los extremos que se encuentran a lado del inserto. Otros oligonucleótidos empleados fueron el oligonucleótido complementario para amplificar *cat* y el complementario para *sacB*.

La secuenciación con el oligonucleótido U19 permitió determinar que ROSE y un fragmento de *cat* se encuentran en fase y sin mutaciones; el oligonucleótido complementario de amplificación de *cat* dio como resultado el fragmento restante de *cat* y un fragmento pequeño de *sacB*, se verificó la orientación y la ausencia de mutaciones, así como la presencia correcta de los nucleótidos del linker traduccional; con el oligonucleótido complementario para *sacB*, así como con el T7 se comprobó la dirección correcta del gen y que en este no se encontraran deleciones, ni mutaciones (Figura 30). De tal modo se comprobó que la construcción genética se encuentra en fase y sin mutaciones, ni inserciones aparentes.

Los experimentos, tanto de PCR como la secuenciación, demuestran que la construcción genética se encuentra convenientemente constituida tal como había sido diseñada en un principio: con la secuencia esperada y con los tamaños correspondientes a cada uno de los insertos, sin embargo estos resultados no son del todo congruentes con la funcionalidad de esta

construcción, así se hace patente la necesidad de realizar experimentos adicionales para verificar los resultados.

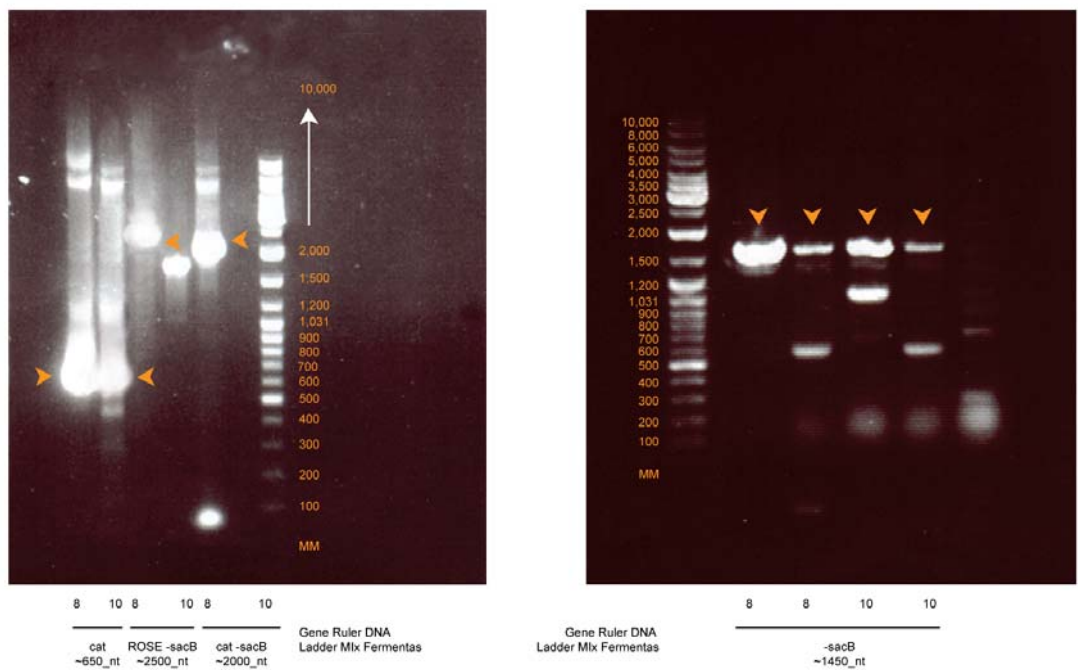


Figura 30. Amplificaciones para verificar la presencia y orientación de la construcción genética pMOS-Blue ROSE-cat-sacB. Se aprecian los diferentes fragmentos obtenidos en las distintas amplificaciones, las flechas indican aquellos resultados congruentes con los tamaños esperados. La colonia 8 presentó el tamaño esperado para todas las amplificaciones, y la presencia de *sacB*. La colonia 10 es recombinante también al tener al gen *sacB*, sin embargo no tiene la orientación esperada a partir del tamaño de los fragmentos obtenidos por PCR

Un factor muy importante a considerar, sobre las inconsistencias entre el crecimiento esperado y el observado (la funcionalidad de la construcción genética) fue el diseño del linker el cual se planeó para que su tamaño fuese de 8 aminoácidos entre CAT y SacB, éste puede ser muy pequeño considerando que se quiere unir un fragmento de tamaño medio (~650 nt) seguido de una proteína de mayor tamaño como es SacB (~1450 nt). Parte de la consideración del pequeño tamaño del linker, es que éste puede no ser lo suficientemente largo o flexible como para dar soporte a esta fusión proteica, o simplemente que este linker afecte el plegamiento de SacB. Pese a lo anterior, los ensayos funcionales sugieren que si se da la expresión y la producción de CAT bajo la regulación de ROSE, lo que no ocurre es la expresión genética (contra-selección) de *sacB* en el medio que únicamente contiene sacarosa, lo que podría explicarse debido a que el plásmido no se está seleccionando realmente, y requiere de la resistencia a antibiótico, en este caso cloranfenicol, como ocurre en el caso del medio con cloranfenicol y sacarosa en donde si ocurre el comportamiento que se esperaba.

Sería necesario en todo caso construir un plásmido control en donde sólo se tuviera a *sacB* bajo la regulación de ROSE o bien bajo otro promotor, y hacer los ensayos de crecimiento correspondientes con el fin de demostrar la función de *sacB* como gen contra-selectivo. O bien es posible invertir el orden de los genes para determinar que sucede cuando el siguiente gen o fragmento a traducir se encuentra después del linker y es más corto. Por último, si no existen resultados concluyentes se debe principalmente al tamaño del inker, se podría llevar a cabo una construcción genética que considere un linker de mayor longitud entre ambos genes.

Discusión General y Conclusiones

El método bioinformático resultó ser lo suficientemente robusto al detectar como control positivo a ROSE y otros elementos regulatorios basados en RNA, descritos previamente. La idea de agrupación por grupos filogenéticos (p.ej. Proteobacteria, Firmicutes) resultó ser una buena estrategia para la búsqueda debido a que hay mecanismos de regulación específicos y los termómetros de RNA parecen ser un buen ejemplo de ello, que sólo ocurren en grupos filogenéticos pequeños. Se llevó a cabo un análisis cuidadoso de nuevos elementos obtenidos conservados que pudieran resultar interesantes y que a la posteridad pueden ser verificados y estudiados experimentalmente, como es el caso de los motivos encontrados en arqueobacterias.

Es relevante señalar que los grupos filogenéticos Proteobacteria y Firmicutes, son dos grupos que cuentan con una gran representación de especies dentro del grupo, no así para los demás *phyla* por lo que al secuenciarse otros genomas será probable que existan otros mecanismos de regulación vía termómetros de RNA diferentes a los que se conocen actualmente y que puedan ser o no específicos para cada grupo. Se aprecia que existe una cantidad considerable de señales que no corresponden siquiera a elementos que actúen a nivel de RNA entre los resultados que se obtuvieron, y esto puede deberse como se ha señalado a la poca conservación y distribución de este tipo de moléculas (termorriboswitches).

Se debe considerar que por ejemplo en el caso de ROSE, la conservación se da a nivel de estructura primaria (secuencia) y por ende en estructura secundaria, lo cual determina y es crucial para la funcionalidad de este elemento. Sin embargo la relación anterior no es bidireccional ya que no siempre se cumple que la conservación estructural secundaria conlleve la conservación de la estructura primaria (como el caso de los termómetros de RNA FourU; de tal modo que la metodología bioinformática en cuanto a conservación estructural puede ser mejorada, por un lado, llevar a cabo una búsqueda a nivel de motivos (firmas) en las secuencias no garantizaba la conservación en estructura secundaria y por otro hacer un análisis manual para cada secuencia con el motivo mediante RNAfold estaba limitado a la conservación de la estructura primaria. Programas como Infernal (Inference of RNA Alignment, por sus siglas en inglés, (Nawrocki y Eddy, 2007) permiten realizar búsquedas con base a la estructura secundaria del RNA, aunado a la similitud en secuencia, a través de modelos de covarianza que dan una puntuación a la combinación del consenso en secuencia y otro consenso en estructura secundaria, así es posible identificar homólogos de RNA en donde se conserve principalmente la estructura secundaria, más allá de tan sólo la estructura primaria.

Un objetivo fundamental que propició la realización de esta tesis fue resolver el aspecto entorno a si existen nuevos termorriboswitches que aún no hayan sido descritos por ejemplo en organismos extremófilos; más aún una nueva búsqueda basada en un enfoque estructural podría resolver esa cuestión; y si existen termómetros de RNA que puedan regular la expresión genética

de otros genes que no necesariamente estén involucrados en el choque térmico por calor, por frío o en la virulencia.

Es relevante hacer notar que análisis *in silico* basados en estudiar la evolución y conservación de mecanismos particulares (en este caso de regulación vía RNA) con un enfoque dirigido a la gran diversidad de genomas que permiten extrapolar, explicar y comprender dichos mecanismos, como el que usado en esta tesis, tienen un gran potencial y serán de gran utilidad en un futuro, dada la inmensa cantidad de información que seguirá generándose a partir de la cantidad de genomas disponibles que son producto de la implementación de técnicas de secuenciación masiva.

La tendencia observada en diferentes organismos bacterianos del uso de moléculas de RNA como elementos sensores de la temperatura, permite pensar que pudieran existir ciertas ventajas selectivas en comparación a posibles tipos de regulación basadas en reguladores proteicos. Aunque la verificación experimental de cualquier hipótesis que trate de explicar esta tendencia es requerida, se pueden proponer las siguientes ventajas que representa una la regulación genética basada en elementos de RNA en relación al de proteínas activadoras/represoras en respuesta a cambios de temperatura intracelular:

1. **Respuesta rápida.** Debido a que estos RNAs se encuentran en *cis*, la traducción se inicia hasta que la región SD se encuentra accesible al ribosoma, y así la regulación de la traducción se efectúa justo antes de que las proteínas reguladoras entren al juego. Tal es el caso de *rpoH*, cuya expresión primero es regulada por un termómetro de RNA o el caso de *IcrF* que codifica para un activador que induce la expresión de un grupo genes de virulencia.
2. **Gasto energético menor.** La síntesis de unos cuantos nucleótidos del RNA (el tamaño promedio de termoriboswitches no es superior a los 100 nucleótidos) pudiera representar un gasto energético inferior al de la síntesis y traducción de los potenciales reguladores peptídicos de varias decenas de aminoácidos.
3. **No se necesitan para la transcripción.** Los transcritos constitutivos con función reguladora son importantes cuando la célula se encuentra comprometida. Esto es relevante después del choque térmico, cuando $\sigma 70$ es termolábil, se inactiva y tiene que ser reemplazado por $\sigma 32$. La traducción de transcritos de *rpoH* ya sintetizados, los cuales están disponibles inmediatamente para los ribosomas, sobrepasando la necesidad de que sean sintetizados *de novo*.

4. **Reversibilidad.** La mayoría de los sistemas reguladores basados en proteínas deben de ser apagados cuando concluyen la respuesta a estrés, lo cual generalmente requiere de circuitos complejos que regulen la síntesis, actividad o estabilidad de los factores reguladores. El sistema de regulación en el caso de los RNA se reduce al plegamiento del RNA, el cual puede ser regulado de acuerdo a las condiciones del entorno.
5. **Simplicidad.** Una de las características más intrigantes del RNA como molécula sensora es su simplicidad. Son independientes del establecimiento de la relación entre los factores reguladores y sus sitios blanco, el único requerimiento es la misma molécula de RNA para el gen que será controlado.
6. **Regulación en cualquier momento después de la transcripción.** Si se considera que un mRNA es traducido varias decenas de veces antes de su degradación, la termorregulación mediante elementos de RNA en *cis* pudiera efectuarse en cualquier momento posterior al cambio de temperatura y no exclusivamente al inicio de la transcripción del gen blanco.

Respecto a los termómetros de RNA *rpoH* y ROSE, la idea de la regulación de la expresión genética a través de una respuesta tan inmediata basada en el RNA, hace pensar que es posible la construcción de un sistema minimalista, en donde la operación de un único elemento, en este caso el pasador que presenta la región del SD y algunas veces el codón de inicio, son suficientes para llevar a cabo la función de regulación de la traducción por el RNA.

En el caso de *rpoH* se han hecho algunas variantes sintéticas mínimas que se acercan al comportamiento silvestres; sin embargo, en algunas construcciones que perdieron segmentos internos, se ven alteradas ciertas propiedades regulatorias (Morita *et al.*, 1999). De manera similar ocurre a ROSE, ya que algunos cambios en los nucleótidos que están fuera de la estructura involucrada en el apareamiento del SD, modifican la regulación (Nocker *et al.*, 2001) de tal forma que el la estructura secundaria conservada que presentan ciertos termómetros de RNA pudieran tener una implicación funcional importante.

Existen al menos dos razones por las cuales se explica la funcionalidad de otras regiones aledañas al SD y codón de inicio: la primera es la propuesta de estas regiones como una “palanca” que le confiere cierta inestabilidad a la apertura de la región termosensible durante el choque térmico; y la otra es que estas regiones se comportan como “cierres” promoviendo el plegamiento de la región termosensible (Narberhaus, 2002).

Una vez más como en el caso de *rpoH*, la cuestión de la dualidad en la regulación de los genes pequeños de choque térmico, esto es por una parte a nivel transcripcional por σ_{32} , y por otra traduccional con ROSE, se explica en parte porque ROSE puede monitorear directamente

alguna desviación en la temperatura, mientras que el regulón σ^{32} integra las señales asociadas al estrés por choque térmico (una de ellas podría ser la concentración de proteínas desnaturalizadas). Estas señales podrían inducirse independientemente de las variaciones en la temperatura por una sobreexpresión proteica o el tratamiento con antibióticos.

Ya se ha mencionado sobre la importancia de la estructura secundaria en estos elementos, debido a que determina en buena parte el funcionamiento de los termómetros de RNA, sobre este aspecto (de Smit y van Duin, 1994) demostraron que los pasadores de mRNA con la secuencia SD afectan la eficiencia traduccional únicamente si su energía libre es menor que $-6\text{kcal}\cdot\text{mol}^{-1}$, lo anterior es un argumento en contra, en términos termodinámicos de la existencia de estructuras tallo-asa tan estables que en un rango de temperatura fisiológico puedan abrirse del todo. Sin embargo los termómetros de RNA son estructuras complejas y modulares, que a través de toda una serie de características estructurales son funcionales con una alta eficiencia.

ROSE es un ejemplo de ello, es muy claro que existen ciertos residuos que constituyen la estructura que comprometen la función de este elemento. En este proyecto no fue posible responder a la pregunta acerca de la conservación del residuo G sobresaliente, y si existían o no otras estructuras similares en donde el residuo de G, no fuese necesario para el termo-monitoreo. Sin embargo se puede especular sobre la conservación de dicho residuo y la presión de selección que lo mantiene así para todas las estructuras ROSE, aunque experimentalmente el cambio de este residuo por otro de A representa la misma respuesta del termómetro, una posible respuesta *a priori* se puede encontrar en el argumento en que ambos residuos son purinas que presentan dos anillos, lo cual estructuralmente puede mantener correctamente el tallo-asa que aparea el SD y su función en ambos casos. Aun sigue abierta la cuestión de si este cambio de residuos puede mantenerse y conservarse en las siguientes generaciones bacterianas, o si la presión de selección para el residuo G es tan alta como para revertir las mutaciones.

En lo que concierne a la discusión de los resultados experimentales es posible que lo que ocurra con la fusión traduccional esté relacionado aparentemente con su tamaño de aproximadamente 2100 nucleótidos, considerablemente grande —el doble del tamaño promedio—, si consideramos que el tamaño de un gen promedio en *E. coli* es de 1000 nucleótidos, por lo que quizás la eficiencia traduccional se vea afectada y la proteína no sea producida en su totalidad, o bien que el plegamiento tridimensional producto de la fusión tenga impedimentos de tipo estérico, siendo estas hipótesis producto de la especulación que deberá de ser verificado posteriormente.

Finalmente, se diseñó una estrategia de selección positiva para la regulación traduccional a través del termómetro de RNA ROSE en la que se demostró que el gen regulado (*cat*) permanece inactivo a bajas temperaturas (30°C) y activo a temperaturas mayores, dentro de un rango fisiológico (37- 42°C). Dicha estrategia funcionó como un control positivo, demostrando que esta construcción genética presenta la capacidad de termo-monitoreo vía un termómetro de RNA.

Como parte de la metodología experimental se clonó en pMOS-Blue, ROSE-CAT-sacB, como un sistema genético que por un lado selecciona de manera positiva y por otro es contraselectivo con el fin de realizar un banco de mutantes; sin embargo aún hacen falta experimentos adicionales que demuestren la funcionalidad de esta construcción para su uso en el futuro. Cabe mencionar que la complejidad de nuestra estrategia de regulación que implica la fusión traduccional de dos proteínas diferentes, es resultado del tipo de regulación traduccional de ROSE. Para el caso de modelos de regulación transcripcional, el desarrollo experimental pudiera ser mucho más sencillo, en términos de que la fusión de los genes de selección positiva y negativa, sería a nivel transcripcional colocandolos de forma contigua favoreciendo la transcripción de los mismos en una misma unidad transcripcional (operón). En lo que a nuestro conocimiento concierne, nuestra estrategia experimental de selección-contraselección desarrollada en éste estudio, es única y pudiera ser empleada para el análisis de nuevos elementos de regulación.

Perspectivas

Con el material expuesto en este proyecto de tesis aún se pueden realizar un sinfín de experimentos, implementaciones, mejoras y darle continuidad al mismo. Desde la perspectiva bioinformática, como se ha mencionado anteriormente es posible llevar a cabo un análisis más allá de los motivos encontrados en la secuencia de nucleótidos, esto es basado en la conservación estructural, ya se ha propuesto el uso del paquete de análisis “infernal” para tales fines e incluso se pueden buscar exhaustivamente motivos estructurales ya descritos y compararlos contra otras secuencias con el fin de encontrarlos regulando genes para los que aún no hay información. Los resultados bioinformáticos generaron predicciones que podrían ser interesantes para su comprobación experimental, es posible usar el sistema reportero construido en este proyecto, para verificar dichos casos.

Es de suma importancia finalizar cuidadosamente el análisis funcional de la construcción genética pMOS-Blue ROSE-*cat-sacB*, y en caso de que esta funcione como se había planteado desde un principio entonces es viable llevar a cabo los siguientes puntos:

- Sintetizar oligonucleótidos degenerados de ROSE, para la construcción de un banco de mutantes que se selecciona con la construcción genética realizada en esta tesis.
- Analizar las secuencias de las estructuras mutantes de ROSE y establecer si es probable o no generar nuevas variantes funcionales sintéticas de un termómetro ya conocido.
- Analizar el patrón de expresión genética de los termómetros funcionales mutantes con respecto a la temperatura; aunado a la observación temporal con el fin de determinar la existencia de cambios en la estructura dadas n generaciones.
- En el caso de aquellas mutantes que presenten el residuo A en la posición del residuo G sobresaliente conservado, sería interesante observar que ocurre después de n generaciones, si el residuo revierte a G o si se conserva en función del papel que tiene la presión de selección sobre esta estructura de RNA y particularmente sobre aquellos residuos que son cruciales para el termo-monitoreo adecuado.
- Caracterizar experimentalmente nuevos motivos encontrados en arqueobacterias (Euryarchaeota y Crenarchaeota) en el caso del COG0459 (Class I heat shock protein chaperonin GroEL, HSP60).
- Explorar de manera experimental, el mecanismo de atenuación a nivel de la traducción vía un elemento de RNA, que ocurre en el COG1191 (DNA-directed RNA polymerase specialized sigma subunit) en *B. subtilis*.

Referencias

1. Abreu-Goodger,C. (2005). Señales conservadas en regiones intergénicas bacterianas: riboswitches y más allá. Tesis de doctorado. Instituto de Biotecnología, UNAM.
2. Abreu-Goodger,C. and Merino,E. (2005). RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.*, **33**, W690-W692.
3. Abreu-Goodger,C., Ontiveros-Palacios,N., Ciria,R., and Merino,E. (2004). Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends Genet.*, **20**, 475-479.
4. Altuvia,S., Kornitzer,D., Teff,D., and Oppenheim,A.B. (1989). Alternative mRNA structures of the cIII gene of bacteriophage lambda determine the rate of its translation initiation. *J. Mol. Biol.*, **210**, 265-280.
5. Bailey,T.L. and Elkan,C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28-36.
6. Bailey,T.L. and Gribskov,M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics.*, **14**, 48-54.
7. Barrios,H., Valderrama,B., and Morett,E. (1999). Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res.*, **27**, 4305-4313.
8. Brooks,B.E. and Buchanan,S.K. (2008). Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim. Biophys. Acta*, **1778**, 1930-1945.
9. Chary,V.K., Meloni,M., Hilbert,D.W., and Piggot,P.J. (2005). Control of the expression and compartmentalization of (sigma)G activity during sporulation of *Bacillus subtilis* by regulators of (sigma)F and (sigma)E. *J. Bacteriol.*, **187**, 6832-6840.
10. Chowdhury,S., Maris,C., Allain,F.H., and Narberhaus,F. (2006). Molecular basis for temperature sensing by an RNA thermometer. *EMBO J.*, **25**, 2487-2497.
11. Chowdhury,S., Ragaz,C., Kreuger,E., and Narberhaus,F. (2003). Temperature-controlled structural alterations of an RNA thermometer. *J. Biol. Chem.*, **278**, 47915-47921.
12. Ciria,R., Abreu-Goodger,C., Morett,E., and Merino,E. (2004). GeConT: gene context analysis. *Bioinformatics.*, **20**, 2307-2308.
13. Cossart,P. (2002). Molecular and cellular basis of the infection by *Listeria monocytogenes*: an overview. *Int. J. Med. Microbiol.*, **291**, 401-409.
14. de Smit,M.H. and van Duin,J. (1994). Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *J. Mol. Biol.*, **244**, 144-150.

15. Dernburg,A.F. and Karpen,G.H. (2002). A chromosome RNAissance. *Cell*, **111**, 159-162.
16. Edwards,R.A., Keller,L.H., and Schifferli,D.M. (1998). Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene*, **207**, 149-157.
17. Eiamphungporn,W. and Helmann,J.D. (2009). Extracytoplasmic function sigma factors regulate expression of the *Bacillus subtilis* yabE gene via a cis-acting antisense RNA. *J. Bacteriol.*, **191**, 1101-1105.
18. El Samad,H., Kurata,H., Doyle,J.C., Gross,C.A., and Khammash,M. (2005). Surviving heat shock: control strategies for robustness and performance. *Proc. Natl. Acad. Sci. U. S. A*, **102**, 2736-2741.
19. Gama-Castro,S., Jimenez-Jacinto,V., Peralta-Gil,M., Santos-Zavaleta,A., Penaloza-Spinola,M.I., Contreras-Moreira,B., Segura-Salazar,J., Muniz-Rascado,L., Martinez-Flores,I., Salgado,H., Bonavides-Martinez,C., Abreu-Goodger,C., Rodriguez-Penagos,C., Miranda-Rios,J., Morett,E., Merino,E., Huerta,A.M., Trevino-Quintanilla,L., and Collado-Vides,J. (2008). RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36**, D120-D124.
20. Gottesman,S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.*, **21**, 399-404.
21. Helmann,J.D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol*, **46**, 47-110.
22. Hilbert,D.W. and Piggot,P.J. (2004). Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol. Biol. Rev.*, **68**, 234-262.
23. Hoe,N.P. and Goguen,J.D. (1993). Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *J. Bacteriol.*, **175**, 7901-7909.
24. Hofacker,I.L. and Stadler,P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics.*, **22**, 1172-1176.
25. Hurme,R., Berndt,K.D., Normark,S.J., and Rhen,M. (1997). A proteinaceous gene regulatory thermometer in *Salmonella*. *Cell*, **90**, 55-64.
26. Hurme,R. and Rhen,M. (1998). Temperature sensing in bacterial gene regulation--what it all boils down to. *Mol. Microbiol.*, **30**, 1-6.
27. Janaszak,A., Majczak,W., Nadratowska,B., Szalewska-Palasz,A., Konopa,G., and Taylor,A. (2007). A sigma54-dependent promoter in the regulatory region of the *Escherichia coli* rpoH gene. *Microbiology*, **153**, 111-123.
28. Lai,E.C. (2003). RNA sensors and riboswitches: self-regulating messages. *Curr. Biol.*, **13**, R285-R291.
29. Li,W. and Godzik,A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.*, **22**, 1658-1659.

30. Makita,Y., Nakao,M., Ogasawara,N., and Nakai,K. (2004). DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res.*, **32**, D75-D77.
31. Mandal,M., Boese,B., Barrick,J.E., Winkler,W.C., and Breaker,R.R. (2003). Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, **113**, 577-586.
32. Mandal,M. and Breaker,R.R. (2004). Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451-463.
33. Martinez-Guerrero,C.E., Ciria,R., Abreu-Goodger,C., Moreno-Hagelsieb,G., and Merino,E. (2008). GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic Acids Res.*, **36**, W176-W180.
34. Masse,E. and Gottesman,S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*, **99**, 4620-4625.
35. Masuda,E.S., Anaguchi,H., Yamada,K., and Kobayashi,Y. (1988). Two developmental genes encoding sigma factor homologs are arranged in tandem in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A*, **85**, 7637-7641.
36. Mattick,J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930-939.
37. Maxwell,K.L., Mittermaier,A.K., Forman-Kay,J.D., and Davidson,A.R. (1999). A simple in vivo assay for increased protein solubility. *Protein Sci.*, **8**, 1908-1911.
38. Morita,M.T., Tanaka,Y., Kodama,T.S., Kyogoku,Y., Yanagi,H., and Yura,T. (1999). Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor. *Genes Dev.*, **13**, 655-665.
39. Narberhaus,F. (2002). mRNA-mediated detection of environmental conditions. *Arch. Microbiol.*, **178**, 404-410.
40. Narberhaus,F., Waldminghaus,T., and Chowdhury,S. (2006). RNA thermometers. *FEMS Microbiol. Rev.*, **30**, 3-16.
41. Nawrocki,E.P. and Eddy,S.R. (2007). Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS Comput. Biol.*, **3**, e56.
42. Newman,J.C. and Weiner,A. (2002). Measuring the immeasurable. *Mol. Cell*, **10**, 437-439.
43. Nocker,A., Hausherr,T., Balsiger,S., Krstulovic,N.P., Hennecke,H., and Narberhaus,F. (2001). A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res.*, **29**, 4800-4807.
44. Nudler,E. and Mironov,A.S. (2004). The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.*, **29**, 11-17.
45. Pallen,M. (1999). RpoN-dependent transcription of *rpoH*? *Mol. Microbiol.*, **31**, 393.

46. Phadtare,S. (2004). Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.*, **6**, 125-136.
47. Pressler,U., Staudenmaier,H., Zimmermann,L., and Braun,V. (1988). Genetics of the iron dicitrate transport system of Escherichia coli. *J. Bacteriol.*, **170**, 2716-2724.
48. Redford,P. and Welch,R.A. (2006). Role of sigma E-regulated genes in Escherichia coli uropathogenesis. *Infect. Immun.*, **74**, 4030-4038.
49. Salgado,H., Gama-Castro,S., Martinez-Antonio,A., Diaz-Peredo,E., Sanchez-Solano,F., Peralta-Gil,M., Garcia-Alonso,D., Jimenez-Jacinto,V., Santos-Zavaleta,A., Bonavides-Martinez,C., and Collado-Vides,J. (2004). RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res.*, **32**, D303-D306.
50. Schumann,W. (2003). The Bacillus subtilis heat shock stimulon. *Cell Stress. Chaperones.*, **8**, 207-217.
51. Seitz,T., Bocola,M., Claren,J., and Sterner,R. (2007). Stabilisation of a (betaalpha)8-barrel protein designed from identical half barrels. *J. Mol. Biol.*, **372**, 114-129.
52. Serganov,A. and Patel,D.J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.*, **8**, 776-790.
53. Smith,A.H., Blevins,J.S., Bachlani,G.N., Yang,X.F., and Norgard,M.V. (2007). Evidence that RpoS (sigmaS) in Borrelia burgdorferi is controlled directly by RpoN (sigma54/sigmaN). *J. Bacteriol.*, **189**, 2139-2144.
54. Storz,G. (1999). An RNA thermometer. *Genes Dev.*, **13**, 633-636.
55. Strauch,M.A. (1995). In vitro binding affinity of the Bacillus subtilis AbrB protein to six different DNA target regions. *J. Bacteriol.*, **177**, 4532-4536.
56. Sun,D.X., Cabrera-Martinez,R.M., and Setlow,P. (1991). Control of transcription of the Bacillus subtilis spoIIIG gene, which codes for the forespore-specific transcription factor sigma G. *J. Bacteriol.*, **173**, 2977-2984.
57. Vigh,L., Maresca,B., and Harwood,J.L. (1998). Does the membrane's physical state control the expression of heat shock and other genes? *Trends Biochem. Sci.*, **23**, 369-374.
58. Waldminghaus,T., Fippinger,A., Alfsmann,J., and Narberhaus,F. (2005). RNA thermometers are common in alpha- and gamma-proteobacteria. *Biol. Chem.*, **386**, 1279-1286.
59. Waldminghaus,T., Gaubig,L.C., and Narberhaus,F. (2007a). Genome-wide bioinformatic prediction and experimental evaluation of potential RNA thermometers. *Mol. Genet. Genomics*, **278**, 555-564.
60. Waldminghaus,T., Heidrich,N., Brantl,S., and Narberhaus,F. (2007b). FourU: a novel type of RNA thermometer in Salmonella. *Mol. Microbiol.*, **65**, 413-424.
61. Wosten,M.M. (1998). Eubacterial sigma-factors. *FEMS Microbiol. Rev.*, **22**, 127-150.

62. Xia,B., Ke,H., Jiang,W., and Inouye,M. (2002). The Cold Box stem-loop proximal to the 5'-end of the Escherichia coli cspA gene stabilizes its mRNA at low temperature. *J. Biol. Chem.*, **277**, 6005-6011.
63. Yamanaka,K. (1999). Cold shock response in Escherichia coli. *J. Mol. Microbiol. Biotechnol.*, **1**, 193-202.
64. Yuan,G. and Wong,S.L. (1995). Regulation of groE expression in Bacillus subtilis: the involvement of the sigma A-like promoter and the roles of the inverted repeat sequence (CIRCE). *J. Bacteriol.*, **177**, 5427-5433.
65. Yura,T., Nagai,H., and Mori,H. (1993). Regulation of the heat-shock response in bacteria. *Annu. Rev. Microbiol.*, **47**, 321-350.
66. Yura,T. and Nakahigashi,K. (1999). Regulation of the heat-shock response. *Curr. Opin. Microbiol.*, **2**, 153-158.
67. Zuker,M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.
68. Rodríguez-Escamilla Z. (2007) Diseño y construcción de riboswitches sintéticos. Tesis de licenciatura. Facultad de Ciencias, UNAM.
69. J. Sambrook,T. Maniatis, E.F. Fritsch. (1989) Molecular Cloning: A laboratory manual. CSHL. Second Edition.
70. Briones Martín del Campo M.C. (2008) Construcción de un vector para clonar productos de PCR. Tesis de Maestría. Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Apéndices

1. Apéndice I: Scripts de los programas bioinformáticos elaborados para esta tesis (Perl)

```
#Programa: 0_extrae_iMUR_COG_filogenia.pl

$min_elem = 2;
$minima_longitud = 40;
$longitud_analisis = 120;
#print "?Que filogenia :? ";
$filogenia = "Firmicutes"; # $ARGV[0];
chomp($filogenia);
#$bichos = `cat /home/DATABASES/phylogeny | cut -f1`;      #unico
cambio para todos los bichos
$bichos_NR = `cat noredundfinal.txt`;
chomp ($bichos_NR);

@organismos_NR = split(/\n/, $bichos_NR);
foreach $organ_NR (@organismos_NR)
{
    $or_NR{$organ_NR} = 1;
}
# Lista de grupos filogenéticos y numero de representantes para
cada uno de ellos
#1 Acidobacteria
#25 Actinobacteria
#1 Alveolata
#1 Aquificae
#6 Bacteroidetes
#11 Chlamydiae
#3 Chlorobi
#2 Chloroflexi
#5 Crenarchaeota
#19 Cyanobacteria
#4 Deinococcus-Thermus
#22 Euryarchaeota
#83 Firmicutes
#11 Fungi
#1 Fusobacteria
#2 Metazoa
#1 Nanoarchaeota
#1 Planctomycetes
#190 Proteobacteria
#7 Spirochaetes
#1 Thermotogae
#1 Viridiplantae

$bichos = `grep -i $filogenia /home/DATABASES/phylogeny | cut -f1`;
```

```

@organismos_redundantes = split(/\n/, $bichos);

chomp ($bichos);
foreach $o (@organismos_redundantes)
{
    if ($or_NR{$o}) {push (@organismos, $o);}
}

#foreach $organ (@organismos)
#    {
#        #print "$org \n";
#        #$org = "/home/DATABASES/iMUR_Genome/" . $organ . ".list";
#        #print ".....$org,,,,,,,,,\n";
#    }

$minima_longitud = 10;
#$gis_all = `cut -f4 cog0071.txt`;
#print "?Que COG :? \n";
#$cog = <STDIN>;
chomp($cog);
$cog = "COG1595";

foreach $o (@organismos)
{
    $archivo = "/home/DATABASES/SUPERLISTS_Genome/" . $o . ".list";
    $gis_all = $gis_all.`grep $cog $archivo | cut -f4`;
}

@gis = split(/\n/, $gis_all);
$n = $#gis+1;
print " $n elementos\n";
if ($n > $min_elem)
{
    $salida =
    "iMUR_" . $cog . "_" . $filogenia . "_" . $longitud_analisis . ".faa";
    open(S, ">$salida");
    foreach $gi (@gis)
    {
        $gi_cog{$gi}=1;
        #print "...>>>>>>>...$gi.....\n";
    }

    #@organismos = </home/DATABASES/iMUR_Genome/*>;

    foreach $organ (@organismos)
    {
        #print ".....$org \n";
        $org = "/home/DATABASES/iMUR_Genome/" . $organ . ".imur";

        open (BICH, $org);

        while(<BICH>)
        {
            chomp;

```

```

($gi_temp,$sec) = split (/\\t/,$_);
($gi)=split(/\\|/, $gi_temp);
#print ".....>>$gi<<.....$gi_cog{$gi}\\n";

if ($gi_cog{$gi} == 1)
{
#print "$gi.....$gi_cog{$gi}\\n";
$longitud = length ($sec);
#print S "$longitud...";
if ($longitud > $minima_longitud)
{
$sp = $longitud - $longitud_analisis;

if ($sp < 0 ) { $sp = 0; }
$largo = $longitud - $sp ;
#print "$sp $longitud $largo\\n";

$sub_sec = substr ($sec,$sp,$largo);
print S ">$gi $organ\\n$sub_sec\\n";
}
else
{
print "longitud chica $longitud $sec\\n";
}
}

}

close(BICH);
}
close (SALIDA);
}
else
{
print "Solo existen $n elementos del COG ....$cog...\\n";
}
}
close(S);

```

```

#Programa: programa_piloto_corre_meme.pl

@memes = (</home/viri/11_octubre/re analisis_nuevo_coli/*.meme>);

#@memes = (</home/viri/11_octubre/COG2032/*.meme>);

foreach $meme (@memes)
{
    print "$meme\n";
    $n_dir = $meme;
    $n_dir =~ s/\.+COG//g;

    $n_dir =~ s/\.meme//g;

    $name = "mast/COG$n_dir";
    system "mkdir -p $name" unless (-d $name);

    @db = (</home/DATABASES/iMUR_Genome_FAA/*.faa>);

    foreach $datab(@db) {
        $bicho = $datab;
        $bicho =~ s/\.faa//g;
        $frecuenc = "$bicho.freq";
        #system "cp $frecuenc ./$name";
        $bicho =~ s/\.+FAA\\//g;
        $out = "/home/viri/todos/$name/$bicho.mast";
        $out_short = $out. "_short";
        `mast $meme -d $datab -norc -brief -bfile $frecuenc -text -
stdout -ev 0.1 > $out`;
        system "grep $bicho $out > $out_short";
        system "rm $out";
        #system "rm *.freq";

    }
    #system "rm *.freq";

    #exit;

}
system "rm *.freq";
exit;

```