

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

IDENTIFICACIÓN DE RIBOSWITCHES QUE ACTÚAN IN CIS E IN TRANS EN GENOMAS BACTERIANOS

TESIS

QUE PARA OPTAR POR EL GRADO DE: Maestro en Ciencias

PRESENTA: Ing. EDGAR RODRÍGUEZ GARCÍA

TUTOR PRINCIPAL Dr. ENRIQUE MERINO PEREZ Instituto de Biotecnología, UNAM campus Morelos

MIEMBROS DEL COMITÉ TUTOR Dr. José Luis Reyes Taboada Instituto de Biotecnología, UNAM campus Morelos

Dr. Cei Abreu-Goodger Laboratorio Nacional de Genómica para la Biodiversidad, CINVESTAV campus Irapuato

Cuernavaca, Morelos. Septiembre, 2019



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. El presente trabajo fue realizado en el Laboratorio 6 del grupo del Dr. Enrique Merino Pérez en el grupo de Genómica Computacional, adscrito al Departamento de Microbiología Molecular del Instituto de Biotecnología de la Universidad Nacional Autónoma de México, bajo el programa de Maestría y Doctorado en Ciencias Bioquímicas.

Este proyecto se realizó con fondos de PAPIIT al proyecto "Identificación in silico y caracterización molecular de la regulación genética basada en riboswitches que actúan en trans (trans-acting riboswitches) o que se transcriben en sentido opuesto al de su gen blanco (antiseneacting riboswitches)" con número IN202120. Además, se realiza un amplio reconocimiento al Programa de Apoyo a los Estudiantes del Posgrado (PAEP) por todos los apoyos otorgados.

Este trabajo se realizó gracias a la beca de CONACyT para estudios de posgrado nivel maestría, con número de becario 482375.

Dedicatoria

A mi familia

En especial a mis padres y hermanos, porque son mi motor y mi mayor motivación de cada día. Por su plena confianza en mí y en todo lo que puedo lograr, impulsándome a siempre dar lo mejor de mí. Por el ejemplo y todas las valiosas lecciones con que me han guiado, todo el sacrificio, apoyo y comprensión que con todo cariño me han brindado durante toda su vida.

Agradecimientos

Al Dr. Enrique Merino Pérez, por brindarme sus conocimientos, sabiduría y grandes observaciones, pero sobre todo por la confianza y el apoyo incondicional brindado.

A los miembros de mi comité tutor el Dr. Jose Luis Reyes Taboada y el Dr. Cei Abreu Goodger por enriquecer constantemente este proyecto mediante excelentes observaciones y apoyarnos para llevar por buen camino el proyecto.

A los miembros del jurado evaluador de este manuscrito por su disposición y cuyas aportaciones fueron claves para concluir de la mejor manera esta tesis: Dr. José Luis Puente García, Dr. Guillermo Gosset Lagarda, Dr. Damien Jean-Rene Formey de Saint Louvent, Dr. José Alberto Hernández Eligio y el Dr. Diego Cortez Quezada.

A la Dra. Rosa María Gutierrez por sus atinados consejos y al M. C. Antonio Loza Roman por la disposición de ayudar en todo momento.

A la M. B. María Luisa Tabche Barrera y el Dr. Raúl Noguéz, por sus buenas charlas y consejos de vida.

Al M. en C. José Ricardo Ciria Merce por su apoyo técnico y metodológico, siendo parte fundamental de este proyecto.

Al Lic. J. Antonio Bolanos Guillen y a Gloria Villa Herrera de la Unidad de Docencia y Formación de Recursos Humanos por facilitar cada uno de los múltiples trámites durante la maestría, asi como los de titulación.

A mis compañeros y amigos de laboratorio: Walter, Lizeth, María del Carmen, Natali, Nori, Diana, Joselyn, Mariela, Maricela, y a todos los miembros del laboratorio Espin-Merino por fomentar un gran ambiente de trabajo.

Por su incomparable amistad, porque a pesar del pasar de los años y la distancia que nos aqueja, a mis mejores amigos Erika, Osiris, Jacob y Tony, por su apoyo incondicional en momentos donde más lo necesitaba.

A EEPAJA (ustedes saben quiénes son), Thalía, Axel, Marieli, Jaime, Ale, Ivette y Miguel, por su invaluable amistad. Pero sobre todo porque con ustedes encontré a un grupo de personas admirables, hicieron que mi estancia durante esta travesía fuera maravillosa. Y lo más importante, se convirtieron en mi familia. Siempre estaré agradecido por haberlos conocido y recuerden que siempre estarán en mi corazón.

A todos los amigos y familiares mi más sincera gratitud por acompañarme en esta travesía.

Glosario

Ácido ribonucleico pequeño, (Small ribonucleic acid) sRNA SAM S-adenosil metiotina mRNA Ácido ribonucleico mensajero, (Messenger ribonucleic acid) CMs Modelos de covarianza, (covariance models) KEGG Base de datos de genes y genomas, (Kyoto Encyclopedia of Genes and Genomes TPP Tiamina pirofosfato mfe Energía libre mínima (minimum free energy), análogo de energía libre de Gibbs (DeltaG) UTR Región sin traducir, (untranslated region) ROSE Elemento termorregulador, (repression of heat shock gene expression element) COG Cluster de grupos ortólogos, (Clusters of orthologous groups) BLAST Herramienta utilizada para comparar y alinear secuencias, (Basic Local Alignment Search Tool)

Índice de figuras

Figura 1. Representación del mecanismo mediante el cual actúan los sRNAs en bacterias	2
Figura 2. Ejemplo de control de la transcripción por parte de un riboswitch	5
Figura 3. Transcrito de mRNA de una bacteria típica que es controlado por un elemento riboregulador,	
un riboswitch	5
Figura 4. Mecanismo de acción de un riboswitch	6
Figura 5. Representación del mecanismo de acción in trans del SAM-I riboswitch en Listeria	
monocytogenes	13
Figura 6. Riboswitches encontrados computacionalmente mediante el software CMsearch usando los	17
Eigura 7. Distribución de ribeswitches del año 2007	17 10
Figura 7. Distribución de la captidad de ribeswitches totales, con respecto a la captidad de ribeswitches	10
que en su plataforma de expresión también forman parte de terminadores transcripcionales Rho-	
independientes	21
Figura 9. Distribución de los porcentajes de tipos de atenuación por parte de cada tipo de	
riboswitch	22
Figura 10. Alineamiento de nucleótidos entre la secuencia de sreA y la secuencia del SAM riboswitch	
presente en el gen denominado Imo-Imo2419 de Listeria monocytogenes	23
Figura 11. Valores de mfe obtenidos del análisis entre el riboswitch SreA y la región UTR 5' del gen prfA	
utilizando el programa RNAhybrid	24
Figura 12. Representación de la función del termorregulador para el gen <i>agsA</i> en Salmonella enterica Figura 13. Diferencias entre la interacción de un sRNA con su gen blanco y la interacción entre la porción	25
de un riboswitch y un gen con un elemento termorregulador en su secuencia UTR 5'	26
Figura 14. Comportamiento de dos SAM riboswitches utilizando para el análisis riboswitches aleatorizados en su secuencia (protocolo controles negativo, líneas rojas) y los riboswitches sin modificar	
(protocolo predicciones, líneas azules)	28
Figura 15. Ecuación para obtener el valor tinificado 7 de cada mfe	-0 29
Figura 16. Ecuación para realizar el cálculo de la distribución hipergeométrica	29
Figura 17. Protocolo computacional aplicado con SAM riboswitches (sin modificar y con los	20
aleatorizados) a organismos representativos de la familia Listeraceae	30
Figura 18. Protocolo computacional usando ventana de 20 bases aplicado con 1-box (sin modificar y con	าา
Tios aleatorizados) a organismos representativos del orden Baciliales	33
Figura 19. Protocolo computacional aplicado con 1-box leader (sin modificar y con los aleatorizados) de	25
Bacillus subtilis en un genoma que no contiene dichos riboswitches como es el caso de Escherichia coli	35
Figura 20. Metabolismo de la metionina y cisteina.	30
rigura 21. Protocolo computacional aplicado con SAIVI riboswitches a organismos representativos por	77
Eigura 22. Protocolo computacional anlicado con riboswitch T boy a organismos representativos por	57
rigura 22. Protocolo computacional aplicado con riboswitch i-box a organismos representativos por	20
Eigura 22. Drotocolo computacional aplicado a gener blance con grandes regiones LITP's versida la T	JŎ
rigura 25. Frotocolo computacional aplicado a genes planco con grandes regiones OTK S, usando la 1-	<u>/</u> 7
Eigure 24. Metabolismo de la lisino	+2 12
rigula 24. Ivietavolisilio de la lisilia	+3

Figura 25. Mecanismo de regulación doble (in cis e in trans) por parte de un mismo tipo de riboswitch	
sobre un gen blanco	44
Figura S1. Ejemplos de algunas de las hibridaciones de los genes más representativas obtenidos con el	
SAM riboswitches y la T-box leader	47
Figura S2. Contexto genómico del gen STH703	47
Figura S3. Metabolismo de la metionina y cisteína	48
Figura S4. Metabolismo de la tiamina	48
Figura S5. Gen <i>axl-AXY_06500</i> con función de transportador de aminoácidos de tipo ABC	49
Figura S5.1. Gen gth-Geoth_0088_06500 con función de cisteína sintasa	49
Figura S6. Mecanismo de regulación por parte de una T-box	50

Índice de tablas

Tabla 1. Lista de clases representativas de riboswitches	9
Tabla 2. Programas usados para la predicción de secuencias blanco de miRNA	12
Tabla 3. Valores de corte (scores) obtenidos del análisis manual de los genes obtenidos como posibles	
candidatos a presentar riboswitches contra la base de datos y los posibles genes que está regulando	20
Tabla 4. Genes obtenidos con las mfe más representativas entre la interacción del SAM riboswitch con	
los genes blancos en Listeria monocytogenes	25
Tabla 5. COGs obtenidos en el análisis a organismos representativos de la familia Listeraceae usando los	
SAM riboswitches	30
Tabla 6. Tabla 6. COGs y su función obtenidos en el análisis a organismos representativos del orden	
Bacillales usando la T-box	34
Tabla 7 COGs obtenidos en el análisis para el SAM riboswitch a organismos representativos por género	
de la base de datos	37
Tabla 8. COGs obtenidos en el análisis para la T-box a organismos representativos por género de la base	
de datos	39
Tabla 9. COGs y su respectiva función obtenidos en el análisis a genes blanco con grandes regiones	
UTR's, usando T-box, SAM y TPP	42
Tabla S1. Genes regulados por un SAM Riboswitches in trans y el COG al que pertenecen	52

Resumen

Uno de los mecanismos de regulación de la expresión genética, dentro del amplio catálogo de mecanismos que poseen las células, son los riboswitches, elementos de RNA ubicados normalmente en el extremo 5'de los mRNAs capaces de reconocer a sus moléculas blanco con alta especificidad y afinidad, y en respuesta a ello, cambiar la conformación de su estructura secundaria para modular la transcripción o traducción de los genes (Winkler, W. C., Cohen-Chalamish, S., & Breaker, 2002). Hasta hace algunos años, todos los riboswitches conocidos regulaban la expresión de sus correspondientes genes blanco actuando en forma *cis*. Sin embargo, el grupo de Loh y colaboradores, en el año 2009, demostraron experimentalmente que el producto de la transcripción prematura de un SAM riboswitch (SreA), podía actuar como un sRNA y regular en forma *trans* la expresión de un gen localizado a distancia.

El objetivo del presente proyecto es el analizar exhaustivamente las secuencias genómicas bacterianas disponibles públicamente (en la base de datos KEGG –a la fecha de Junio de 2018-) para la identificación de riboswitches que presenten un mecanismo de regulación similar al previamente reportado (SreA), que pudieran actuar tanto *in cis* como *in trans*.

En el presente trabajo se desarrolló un protocolo computacional que tiene la capacidad de 1.predecir riboswitches en toda la escala de los genomas secuenciados, 2.- identificar potenciales terminadores transcripcionales Rho-independientes que pudieran ser parte de la plataforma de expresión de los riboswitches identificados, y 3.- analizar las potenciales interacciones de la plataforma de expresión de riboswitches transcripcionales, con la región 5' UTR de cada uno de los genes de los genomas en estudio con el fin de identificar posible regulación traduccional *in trans*.

El protocolo computacional desarrollado obtiene genes potencialmente regulados *in trans* por el riboswitch usado para el análisis, mediante estrategias estadísticas seleccionamos únicamente aquellos genes que realmente fueron representativos. Finalmente, los genes seleccionados fueron agrupados de acuerdo a sus relaciones de ortología en base a la clasificación de genes ortólogos COGs.

Abstract

One of the mechanisms of regulation of genetic expression, within the wide range of mechanisms that affect cells, are riboswitches, RNA elements normally determined at the 5 'end of mRNAs can be recognized by their target molecules with high specificity and affinity, in response to this, change the conformation of its secondary structure to modulate the transcription or translation of genes (Winkler, WC, Cohen-Chalamish, S. and Breaker, 2002). Until a few years ago, all known riboswitches regulate the expression of their corresponding target genes by acting *in cis* form. However, Loh and collaborators' group, in 2009, experimentally demonstrated that the product of premature transcription of a SAM riboswitch (SreA), acted as an sRNA and regulates *in trans* the expression of a gene located remotely.

The objective of this project is to analyze thoroughly the publicly available bacterial genomic sequences (in the KEGG database - as of June 2018 -) for the identification of riboswitches similar to the one reported (SreA), which could act both *in cis* as *in trans*.

In the present work, a computational protocol was developed that has the ability to predict riboswitches on the entire scale of sequenced genomes, identify potential Rho-independent transcriptional terminators that could be part of the expression platform of the identified riboswitches, and analyze the putative interactions of the expression platform of transcriptional riboswitches, with the 5 'UTR region of each of the genome genes under study in order to identify possible translational regulation *in trans*.

The developed computational protocol obtains genes potentially regulated *in trans* by the riboswitch used for the analysis, by means of statistical strategies specifically selecting those genes that were really representative. Finally, the selected genes were grouped according to their orthology relationships based on the classification of COG orthologous genes.

Índice

D	edicator	ia	II
A	gradecir	nientos	III
G	losario .		IV
Ín	dice de	figuras	V
Ín	dice de	tablas	VII
R	esumen		VIII
A	bstract		IX
1	Intro	ducción	1
	1.1	Regulación genómica por RNA (mundo del RNA)	1
	1.2	Regulación por RNA's	1
	1.2.1	Metabolismo car: metabolismo del carbono, metabolismo de amino ácidos, regulación de metales	3
	1.2.2	Formación de biopelícula y percepción del quórum (quorum sensing).	3
	1.2.3	sRNAs y patogénesis	4
2	Ante	cedentes	4
	2.1	Mecanismo de acción de riboswitches.	4
	2.1.1	Tipos de ligando (enzimas, precursores de nucleótidos, aminoácidos, metales ionicos)	6
	2.2	Familia de riboswitches SAM.	9
	2.2.1	SAM-I	9
	2.2.2	SAM-II	10
	2.2.3	SAM-III.	10
	2.2.4	SAM-IV.	10
	2.2.5	SAM-V.	10
	2.2.6	SAM-VI.	10
	2.3 blancos	Herramientas bioinformáticas empleadas para la identificación <i>de novo</i> de riboswitches y sus correspondies de regulación	entes 10
	2.4	Herramientas bioinformáticas empleadas para la identificación de blancos de miRNA	11
3	Ante	cedente directo.	12
	3.1	Primer y único ejemplo de un riboswitch actuando tanto in cis, como in trans.	12
4	Justi	ficación	13
5	Hipó	tesis	14
6	Obje	tivos	14
	6.1	Objetivo general	14
	6.2	Objetivos específicos	14
7	Meto	odología	15
	7.1 de los r	Utilizar la base de datos Rfam para obtener el conjunto de modelos de covarianza (CMs) reportado para ca nás de 40 riboswitches actualmente caracterizados	ada uno 15
	7.2 lmo241	Identificar, mediante comparación de secuencias usando BLAST, los genes ortólogos a los del operón <i>lmo</i> 8-lmo2417 de Listeria monocytogenes.)2 <i>419-</i> 15
	7.3 identifi	Utilizar el software Infernal (Nawrocki & Eddy, 2013) que usa los modelos de covarianza de Rfam para car los riboswitches en secuencias genómicas	15

	7.4 platafor	Identificar potenciales terminadores transcripcionales Rho-independientes que pudieran ser parte de la ma de expresión de los riboswitches identificados	16
	7.5 transcrij	Uso del programa RNAhybrid para analizar interacciones de la plataforma de expresión de riboswitches pcionales con la región 5' UTR de cada gen del genoma en estudio con el fin de identificar posible regulación	16
0	Traducci	ional in trans	16
0	o 1	Identificación de ribes vitebes en conomes estuelmente seguenciados	17
	0.1	Diberwitches identificades hase más de 10 eñes y surgeión menuel de les región encontrades	1/
	8.2	Riboswitches identificados nace mas de 10 anos y curación manual de los recien encontrados	. 18
	8.5	Identificación de riboswitches con atenuadores transcripcionales en su plataforma de expresión	20
	8.4	Identificación del riboswitch SAM <i>sreA</i> en el genoma de <i>Listeria monocytogenes</i> .	22
	8.5	Normalización de la MFE (minimum free energy) entre el riboswitch y la región 5' de cada gen	23
	8.6 blancos	Diferencias entre hibridaciones canonicas de sRNA, y genes con secuencias termorreguladoras en sus genes .24	
	8.7	Validaciones del método y estrategia para ubicar los hits más representativos	27
	8.7.1	Análisis de la potencial regulación in trans de otros SAM riboswitches	27
	8.7.2	Método estadístico para diferenciar genes representativos.	28
	8.8	Aplicación del protocolo computacional obtenido, usando pequeños volúmenes de datos	30
	8.9	Análisis de la potencial regulación in trans del SAM riboswitch.	36
	8.10	Análisis de la potencial regulación in trans del riboswitch de la T-box	38
	8.11	Análisis de la potencial regulación in trans del TPP riboswitch.	40
	8.12 regione	Análisis de regulación in trans de riboswitches considerando como potenciales blancos genes con grandes s UTR´s	41
	8.13	Ejemplos de doble regulación (<i>cis</i> e <i>in trans</i>) por parte de diferentes riboswitches	43
9	Conc	lusiones	45
1() Persp	ectivas.	46
A	péndice	A	46
	Hibrida	ciones particulares de algunos riboswitches y metabolismo afectado de los genes blanco.	46
A	nexos		52
	Anexo	1. Organismos representativos por género de la familia listeriaceas	52
	Anexo	2. Organismos representativos por género que pertenezcan al orden Bacillales.	52
B	ibliograf	ía	53

Identificación de riboswitches que actúan *in cis* e *in trans* en genomas bacterianos

1 Introducción.

La capacidad de los organismos para responder a ciertos estímulos, así como de adaptarse a diversas condiciones, implica procesos celulares cuya mayoría de respuestas necesitan ser reguladas por cambios en la expresión genética, es decir, genes que se encuentran apagados tendrán que ser expresados y genes expresados tendrán que ser apagados. Para llevar a cabo la regulación genética se necesitan elementos de regulación que puedan actuar a nivel de la transcripción, traducción o postranscripcional. De manera general, la mayoría de los reguladores son proteínas que se unen a sitios específicos al DNA o RNA, dependiendo de si la regulación es a nivel transcripcional, postranscripcional o traduccional, aunque existen casos específicos en que la regulación puede ser mediada por moléculas la actividad catalítica del RNA (ribozimas) o por interferir en el splicing alternativo de mRNAs, en casos excepcionales en donde los riboswitches se encuentran en organismos eucariontes.

1.1 Regulación genómica por RNA (mundo del RNA).

Durante muchos años se mantuvo la creencia que el papel de RNA se limitaba como molécula que participaba en el dogma central de la biología molecular como intermediario en la producción de una proteína. Fue hasta 1986 cuando Walter Gilbert postula la hipótesis del "Mundo del RNA" (Gilbert, 1986), la cual describe como hace millones de años el RNA era la principal molécula de los seres vivos ya que tiene la capacidad de autorreplicarse, y por lo tanto, pudo haber almacenado y transmitido información genética de forma independiente entre generaciones (función que actualmente desempeña el DNA), y funciones catalíticas que actualmente han sido reemplazadas por moléculas más versátiles, las proteínas. En la actualidad se sabe que el RNA es una molécula sumamente dinámica, con características y propiedades que le permiten participar y modular diversos procesos moleculares y celulares (Mattick, 2004).

1.2 Regulación por RNA's.

En los años 60s, durante los posteriores análisis del operón de biosíntesis de triptófano, se encontró que éste era regulado por un atenuador, un RNA capaz de formar diversas estructuras secundarias, las cuales tienen la capacidad de funcionar como terminador responsable de la atenuación transcripcional y de formar un antiterminador, ya que la formación de un antiterminador evitara que se forme el terminador. La segunda estructura es capaz de permitirle a la RNA polimerasa continuar con la transcripción. La formación de cualquiera de las estructuras está determinada por la

concentración de triptófano, mayores concentraciones de triptófano inducirán la formación del terminador y con bajas concentraciones se formará el antiterminador (Oxender, Zurawskit, & Yanofsky, 1979; Yanofsky, 1960). La clave de la regulación por atenuación, en el caso del operón de biosíntesis de triptófano fue explicada al ubicar dos codones *trp* en un pequeño péptido líder, que de acuerdo a la abundancia del tRNA cargado para dicho aminoácido, hacía que el ribosoma pausara o continuara su traducción. La posición de estos ribosomas traduciendo al péptido líder, respecto a la ubicación de las estructuras secundarias alternativas de antiterminación-terminación, hacían que el transcrito terminara o no, prematuramente.

Durante los últimos años el descubrimiento reguladores de sRNA aumento considerablemente, en gran parte debido a las nuevas tecnologías de secuenciación masiva, matrices de mosaico, y a nuevos métodos bioinformáticos que permiten predecir y encontrar nuevas moléculas de RNA capaces de funcionar como RNA reguladores (Dutta & Srivastava, 2018). Los sRNA bacterianos identificados se conocen por ser un grupo heterogéneo de moléculas con la capacidad de actuar por diferentes mecanismos con la finalidad de modular una amplia gama de respuestas fisiológicas. Algunas características de los sRNA son su longitud, llegando a tener longitudes entre 40-500 nucleótidos y localizarse en regiones intergénicas. La figura 1 muestra el mecanismo general por el cual actúan los sRNAs, interaccionan con sus mRNAs diana ya sea de forma in cis o in trans mediante mecanismos antisentido, generalmente cerca de los sitios donde inicia la traducción, con el fin de modular a nivel postranscripcional la expresión génica por medio de: a) la inhibición de la traducción; b) alterando la estabilidad de los RNAs. (Ahmed, Hafeez, & Mahmood, 2018; Michaux, Verneuil, Hartke, & Giard, 2014; Saberi, Kamali, Najafi, Yazdanparast, & Moghaddam, 2016).



Figura 1. Representación del mecanismo mediante el cual actúan los sRNAs en bacterias. Mecanismo de acción (A) de los sRNA codificados *in trans* y (B) los sRNA codificados *in cis*. Tomado de (Ahmed, Hafeez, & Mahmood, 2018).

1.2.1 Metabolismo car: metabolismo del carbono, metabolismo de amino ácidos, regulación de metales.

Los constantes cambios durante el crecimiento de los microorganismos pueden ser muy variables y para contender con dichos cambios los programas de expresión genética de las bacterias es regulado. Uno de los elementos de regulación con los que cuentan las bacterias para atender a dichos cambios que pueden originar diferentes tipos de estrés son los llamados RNAs pequeños (sRNAs). Algunos sRNA, como CsrB y CsrC de Escherichia coli, actúan sobre sistemas como el Csr (regulador de almacenaje de Carbono) siendo antagonistas de la actividad de CsrA (regulador de la inanición de carbono y biosíntesis de glucogeno) durante la fase de crecimiento exponencial y reprimiendo las vías metabólicas relacionadas a la fase estacionaria (Babitzke & Romeo, 2007). CsrA es una proteína de unión a RNA que reconoce secuencias en la región 5' no traducidas del mRNA blanco, dichas secuencias pueden sobrelapar la región Shine-Dalgarno de algunos mRNAs. Generalmente la unión de CsrA a su mRNA inhibe el inicio de la traducción de sus mRNAs. Los sRNAs CsrB y CsrC contienen también sitios de unión de CsrA, de tal manera que la presencia de estos sRNAs, titulan la actividad reguladora de CsrA. Otro sRNA involucrado en la regulación del metabolismo celular es el GcvB (206 nt) de Salmonella enterica, el cual reprime el transporte de péptidos inducidos por altos niveles de glicina, regulando los mRNAs de oppA y dppA, los cuales codifican para proteínas de unión al periplasma encargadas de sistemas de transporte para oligopéptidos y dipéptidos específicamente. GcvB también aumenta la supervivencia bacteriana bajo pH ácido a través de la regulación del gen que codifica para el factor transcripcional RpoS. Otro ejemplo de un regulador clave en el metabolismo celular de Escherichia coli es el sRNA RyhB, molécula fundamental en la adaptación a las condiciones que limitan algunos metales, específicamente el hierro. Su mecanismo de acción es utilizar a la chaperona Hfg para mantener su actividad y estabilidad, aparte recluta a la RNasa E para facilitar la degradación de sus blancos (Massé et al., 2005), los efectos posteriores a la inhibición de sus blancos muestran su importancia para redirigir el uso del hierro celular al evitar la nueva síntesis de enzimas no esenciales que contienen hierro (proceso conocido "ahorro de hierro") y mejora la capacidad de las células para sintetizar sidróforos depuradores de hierro (Richards & Vanderpool, 2011)

1.2.2 Formación de biopelícula y percepción del quórum (quorum sensing).

Las implicaciones de las capacidades de los sRNA para modular ciertos procesos no se detienen en el metabolismo celular, se ha demostrado una gran cantidad de moléculas de RNA capaces de regular la formación de biopelícula. Recientemente en un estudio usando RNA-seq se identificaron un conjunto de 29 transcritos de ncRNA en la archea *Sulfolobus acidocaldarius*. El ncRNA más abundante de este conjunto fue el ncRrrR que demostró tener influencia sobre la formación del biopelículade la archea actuando como un RNA antisentido contra el mRNA de una proteína hipotética de membrana (Orell et al., 2018). La formación de biopelícula no es el único proceso en el cual en los últimos años se ha encontrado un efecto regulador por parte de moléculas pequeñas de RNA. En el mecanismo de regulación *"quorum sensing"* se han encontrado muchos sRNA con capacidades de regular la expresión de genes específicos para detectar y responder a otros

microorganismos, ejemplo de dicho proceso es el sRNA ReaL, molécula implicada en unir proteínas del sistema PQS (señal de quinolona en *Pseudomonas*) (Carloni, Macchi, Sattin, Ferrara, & Bertoni, 2017)

1.2.3 sRNAs y patogénesis.

Son diversas las respuestas fisiológicas en función de sRNA y que desempeñan un papel fundamental en la adaptación de bacterias a estrés químico y condiciones de cultivo variables. Se ha demostrado que los sRNA tienen un papel fundamental en respuesta al estrés (nutricional, osmótico, etc.) y a la patogénesis microbiana ya que las células tienen que hacer frente a las tensiones durante el proceso de infección (fluidos corporales, fagocitosis y respuesta inmune). Durante una infección, las bacterias patógenas deben adaptarse rápidamente a las condiciones ambientales tan cambiantes a las que se enfrentan, incluidas la toxicidad sistémica, las estrategias de supervivencia en nichos específicos y evitar la exposición al sistema inmune. Entre las múltiples estrategias que han desarrollado los microorganismos, los sRNA son considerados los mejores transductores de señales entre el ambiente y el interior de la célula ya que varios sRNA han demostrado regular la síntesis de factores de virulencia y otros rasgos patógenos mediante la adaptación del metabolismo bacteriano en respuesta al huésped (Michaux et al., 2014)

2 Antecedentes.

Se conoce que las moléculas de RNA tienen un papel fundamental en la regulación de la expresión genética modulando metabolismo celular, formación de biopelícula y *quorum sensing*, respondiendo a situaciones de estrés y teniendo un papel fundamental en la patogénesis bacteriana. Estos ncRNAs que pertenecen a la misma clase comparten secuencias precisas y características estructurales, que se han conservado a lo largo de varios procesos evolutivos. El grado de conservación de la secuencia es menor que el observado para los genes que codifican proteínas, pero es crucial para explicar la heterogeneidad funcional de los ncRNA (Qu & Adelson, 2012). Uno de los ejemplos más significativos de RNA funcionales conservados son los riboswitches (Maumita Mandal & Breaker, 2004b; Miranda-Rios, Navarro, & Soberon, 2001).

2.1 Mecanismo de acción de riboswitches.

Los riboswitches son elementos de RNA capaces de reconocer con gran afinidad y alta especificidad moléculas blancos en ausencia de proteína alguna. Constan de dos dominios o plataformas: a) La plataforma de reconocimiento o aptámero, que está formado por una estructura tridimensional específica que reconoce a su ligando selectivamente, y b) la plataforma de expresión, que es una secuencia de nucleótidos que sufre cambios alostéricos en respuesta a la unión del metabolito por el aptámero. La plataforma de expresión comúnmente está constituida por dos tipos

de atenuadores: transcripcional, en cuyo caso la formación de un terminador Rho-independiente determina si la transcripción terminará o no prematuramente; o bien, un atenuador traduccional, en la que la secuencia Shine-Dalgarno (SD) es secuestrada por una estructura secundaria de RNA con lo que se impide la traducción del mRNA (Montange & Batey, 2006).



Figura 2. Ejemplo de control de la transcripción por parte de un riboswitch. Cuando los metabolitos no están unidos (-M), la plataforma de expresión incorpora la secuencia interruptora dentro del antiterminador (AT) y procede la transcripción a través de la región codificante del mRNA. Durante la unión del metabolito (+M), la secuencia interruptora está incorporada dentro del dominio del aptamero, y la plataforma de expresión se pliega dentro del terminador (T), causando el aborto de la transcripción. Dominio del aptamero (rojo), secuencia interruptura (morado) y plataforma de expresión (azul). Tomado de (Edwards, B. A. L., & Batey, R. T., 2010).

La naturaleza del riboswitch le permite formar diversas estructuras, la más básica es una hélice bicatenaria semejante al DNA, sin embargo, debido a que los RNAs no necesitan mantener un emparejamiento perfecto de Watson-Crick pueden formar otros tipos de estructuras. Por ejemplo, pueden plegarse y formar una horquilla compuesta de una hélice coronada y un bucle, formando lo que se conoce como estructura secundaria. Los riboswitches están localizados con mayor frecuencia en la región 5' no traducida (5' UTR) en donde regulan el paso de las señales para la atenuación de la transcripción o el inicio de la traducción (Edwards & Batey, 2010).



Figura 3. Transcrito de mRNA de una bacteria típica que es controlado por un elemento riboregulador, un riboswitch, que está compuesto de tres secciones: la región 5' sin traducir (UTR), la región codificante de la proteína empieza con el codon de inicio (AUG) y termina con el codon de terminación (UAA) en la región 3' sin traducir. Tomado de (Edwards, B. A. L., & Batey, R. T., 2010).

En genes regulados mediante atenuación de la transcripción, la región lider incluye un terminador transcripcional intrínseco. En contraste, aquellos genes regulados a nivel traduccional son siempre transcritos, pero mantienen un elemento de RNA con la capacidad de secuestrar el sitio de unión al ribosoma, usualmente por el apareamiento de la secuencia SD con la secuencia complementaria anti-Shine-Dalgarno (ASD); la formación de la estructura en hélice de ASD-SD previene la unión de la subunidad ribosomal 30s, y previene la formación de dicha hélice por el secuestro de la secuencia ASD en una estructura alternativa requerida para la expresión (Figura 4) (Gelaye, Rondon, Araya, & A, 2016).



Figura 4. Mecanismo de acción de un riboswitch. a) Regulación por atenuación de la transcripción. b) Regulación en el inicio de la traducción. Tomado de (Gelaye, Rondon, Araya, & A, 2016).

2.1.1 Tipos de ligando (enzimas, precursores de nucleótidos, aminoácidos, metales ionicos).

A continuación, se consideran los tipos de ligandos que son reconocidos por los riboswitch y constituyen una señal de regulación de la expresión génica. Es importante considerar el acoplamiento entre el ligando y el aptámero así como sus conformeros competentes para la unión pero que no formen el aptámero completo ya que las interacciones del ligando con estos conformeros puede tener un profundo efecto sobre la ruta de plegamiento del RNA (Aboul-ela, Huang, Abd Elrahman,

Boyapati, & Li, 2015). En la Tabla 1. se muestran para cada riboswitch los tipos de ligando. La familia o clase, el año en el cual se descubrió y su código en Rfam.

Tipo de ligando	Nombre del ligando	Referencia	Acceso en Rfam
Enzyme cofactor	S-Adenosylhomocisteine	(Wang, Lee, Morales, Lim, & Breaker, 2008)	RF01057
	Tetrahydrofolate (THF)	(Varilly & Chandler, 2012)	RF01831
	Tungsten cofactor	(Regulski et al., 2008)	
	Molybdenum cofactor	(Regulski et al., 2008)	RF01055
	S-adenosyl methionine (SAM) riboswitch	(Z. Weinberg et al., 2008)	RF00634
	SAM-V riboswitch	(Meyer et al., 2009; Poiata, Meyer, Ames, & Breaker, 2009)	RF01826
	SAM-I/IV variant riboswitch	(Zasha Weinberg et al., 2010)	RF01725
	SAM riboswitch (S box leader)	(Zasha Weinberg et al., 2010)	RF00162
	SAM-VI riboswitch	(Mirihana, G., Sherlock, Weinberg, & Breaker, 2018)	RF02885
	ykoK leader	(J. E. Barrick et al., 2004; Ramesh, Wakeman, & Winkler, 2011)	RF00380
	AdoCbl variant RNA	(Zasha Weinberg et al., 2010)	RF01689*
	TPP riboswitch (THI element)	(Rodionov, Vitreschak, Mironov, & Gelfand, 2002)	RF00059
	ydaO/yuaA leader	(J. E. Barrick et al., 2004)	RF00379
	ykkC-yxkD leader	(Corbino et al., 2005)	RF00442
	YybP-ykoY leader	(Saito, Kakeshita, & Nakamura, 2009)	RF00080
	NiCo riboswitch	(Medsker, Forno, Simhan, Juan, & Sciences, 2016)	RF02683
	Cyclic di-GMP-II riboswitch	(Nedelman, Luke, Freedman, & Moossavi, 2010)	RF01786
	SAM/SAH riboswitch	(Zasha Weinberg et al., 2010)	RF01727
	SAM riboswitch (alpha-	(Corbino et al., 2005)	RF00521

	proteobacteria)		
	SMK box translational riboswitch	(Fuchs, Grundy, & Henkin, 2006)	RF01767
	Magnesium Sensor	(Phop, Cora, & Lps, 2006)	RF01056
	M. florum riboswitch	(Z. Weinberg et al., 2008)	RF01510
	Thiamin pyrophosphate (TPP)	(Miranda-Rios et al., 2001)	
	Aquacobalamin	(Yang, Sau, Lai, Cichon, & Li, 2015)	
	Adenosylcobalamin (AdoCbl)	(Nahvi et al., 2002)	RF01482
	Flavin mononucleotide (FMN)	(Winkler, W. C., Cohen-Chalamish, S., & Breaker, 2002)	RF00050
Nucleotide precursor	Adenine	(Maumita Mandal & Breaker, 2004a)	
	Guanine	(Maumita Mandal, Boese, Barrick, Winkler, & Breaker, 2003)	
	Prequeuosine-1	(Roth et al., 2007)	RF00522
	Prequeuosine 1-II	(Roth et al. <i>,</i> 2007)	RF01054
	Prequeuosine 1-III	(Meyer, Roth, Chervin, Garcia, & Breaker, 2008)	RF02680
	2'-Deoxyguanosine	(Kim, Roth, & Breaker, 2007)	
	Purine	(Maumita Mandal et al., 2003)	RF00167
Amino acid	Glutamine	(Ames & Breaker, 2011)	RF01739
	Glycine	(M. Mandal, 2004)	RF00504
	Lysine	(F. J. Grundy, Lehman, & Henkin, 2003)	RF00168
Metal ion	F ⁻ (Fluoride) (crcB)	(Baker et al., 2012)	RF01734
	Mg ²⁺	(Dann et al., 2007)	
	Mn ²⁺	(Shi, Zhao, & Kong, 2014)	
	Ni ²⁺ /Co ²⁺	(Medsker et al., 2016)	
Signaling	Cyclic di-GMP-I	(Zasha Weinberg et al., 2007)	RF01051
Molecules	Cyclic di-AMP	(Nelson et al., 2013)	

	ZTP (ZMP/ZTP)	(Zasha Weinberg et al., 2010)	RF01750
	Cyclic AMP-GMP	(Nelson et al., 2013)	
Other Metabolites	GlcN6P (glms)	(J. E. Barrick et al., 2004)	RF00234
	Azaaromatics	(Li, Hwang, Stav, & Breaker, 2016)	
	Guanidine	(Nelson, Atilho, Sherlock, Stockbridge, & Breaker, 2017)	
	T-box	(Sherlock & Breaker, 2017)	RF00230

Tabla 1. Lista de clases representativas de riboswitches. Una visión general de los riboswitches conocidos y caracterizados,incluidas las clases mejor estudiadas hasta la fecha.

Dada la distribución de los riboswitches por su ligando, también se agrupan por familias, cada familia de riboswitches están relacionadas por los ligandos que reconocen. Por ejemplo, la familia de los SAM riboswithces reconocen al metabolito S-adenosilmetionina; sin embargo, existen distintas clases de riboswitches que pueden pertenecer a una misma familia, cada clase se distingue por un patrón de secuencia común que generalmente define el sitio de unión al ligando, así como las características necesarias para plegar el RNA en una forma tridimensional adecuada. La característica anterior de los riboswithces, y en especial de los SAM riboswitches, es importante para el proyecto porque a la fecha es el único ejemplo del riboswitch reportado en la literatura actuando *in cis e in trans* simultáneamente, y pertenece a la familia de los SAM riboswitches (Loh et al., 2009).

2.2 Familia de riboswitches SAM.

A la fecha han sido descritas seis familias que responden a SAM los cuales tienen la capacidad de discriminar efectivamente entre S-adenosilmetionina y S-adenosilhomocisteina (SAM y SAH), el bioproducto de la utilización de SAM como donador de grupos metilo. Estos riboswitches difieren en estructura, mecanismo de acción y propiedades para unirse a su ligando, ilustrando la habilidad y la facilidad de los diversos RNAs para converger en un efecto regulatorio efectivo mientras conservan importantes diferencias funcionales (Gelaye et al., 2016). A continuación, se describen las seis familias descritas a la fecha que actúan por medio de SAM.

2.2.1 SAM-I.

La regulación ocurre a nivel de atenuar la transcripción, aunque existen modelos en donde por medio de predicciones se ha encontrado que puede actuar regulando el inicio de la traducción. Es encontrado principalmente in bacteria Gram-positivas, donde regula genes involucrados en la biosíntesis y captación de metionina o de SAM (Frank J. Grundy & Henkin, 1998).

2.2.2 SAM-II.

Encontrado principalmente en proteobacterias, con un tamaño considerablemente menor que el SAM-I y con una regulación mayormente a ocurrir a nivel del inicio de la traducción mediante la unión a los promotores de SAM, formando un pseudoknot (estructura secundaria de ácidos nucleicos) que resulta en el bloqueo de la secuencia SD (Corbino et al., 2005).

2.2.3 SAM-III.

Regula principalmente a nivel traduccional. En *Enterococcus faecalis* regula la expresión de genes rio arriba por medio de cambios conformacionales al momento de detectar SAM (Gong, Wang, Wang, Wang, & Zhang, 2016).

2.2.4 SAM-IV.

Se encuentra principalmente en Actinomycetales, y está localizado río arriba de los genes implicados en el metabolismo del azufre. Es capaz de plegarse en una estructura secundaria similar a la del SAM-I riboswitch, con la notable excepción de que SAM-IV carece de la hélice P4 del SAM-I riboswitch (Kaphingst, Persky, & Lachance, 2010).

2.2.5 SAM-V.

Esta clase de riboswitch se ha encontrado únicamente en α -proteobacterias y bacteoridetes marinos, exhibe una serie de características distintas como regular la expresión de genes relacionados con el metabolismo del sulfuro, aunque la secuencia y estructura consenso tiene similitudes con el sitio de unión de los aptámeros del SAM-II riboswitch (Poiata, Meyer, Ames, & Breaker, 2009).

2.2.6 SAM-VI.

Se ha encontrado predominantemente en especies de *Bifidobacterium*, los aptameros de su secuencia unen selectivamente el cofactor SAM y discriminan fuertemente a SAH. la secuencia y estructura consenso comparte algunas características con el modelo para la clase del SAM-III riboswitch (Gayan, Sherlock, Weinberg, & Breaker, 2018).

2.3 Herramientas bioinformáticas empleadas para la identificación *de novo* de riboswitches y sus correspondientes blancos de regulación.

Debido a la importancia que han mostrado los riboswitches en la regulación génica, es necesario el uso y la creación de nuevos métodos para la identificación *de novo* de tales elementos de regulación. A la fecha, la mayoría de los programas que se han desarrollado funcionan con el principio de alineamiento de secuencias para identificar secuencias conservadas en elementos de RNA no

traducidos, junto con algoritmos de plegamiento estructural de RNA. Los programas que están actualmente disponibles generalmente se limitan a la identificación de un subconjunto específico de tipos conocidos de riboswitches, y solo son capaces de buscar entradas de secuencias de cierto tamaño. Una vez hecha dicha identificación, es posible su representación mediante matrices de probabilidades de la conservación de sus secuencias y/o estructuras secundarias, como lo hace el programa de análisis CMsearch (Nawrocki and Eddy, 2013). Hasta la fecha, se han desarrollado pocos protocolos computacionales para la identificación de novo de riboswitches que puedan aprovechar la abundancia de secuencias de genomas totalmente secuenciados (Havill, Bhatiya, Johnson, Sheets, & Thompson, 2014). Uno de dichos protocolos fue desarrollado en nuestro grupo y se basa en la búsqueda iterativa de motivos conservados en secuencias intergénicas agrupadas con base al grupo de ortología COG de sus correspondientes genes (Abreu et al., 2004). En cualquiera caso, debido a la idea de que la regulación de los riboswitches se daba exclusivamente in cis, los blancos predicha de su regulación siempre se limitó a que fueran exclusivamente a los genes contiguos ubicados inmediatamente río debajo de ellos. Para abordar la necesidad de predicción de riboswitches en toda la escala de los genomas secuenciados considerando no solo su conocida regulación in cis, sino contemplando la posibilidad excepcional de también tener un efecto de regulación a distancia, en la presente tesis se muestra el desarrollo y uso de nuestra tubería de cómputo para la identificación de riboswitches con potencial capacidad de regular la expresión genética tanto in cis como in trans en secuencias genómicas de DNA.

2.4 Herramientas bioinformáticas empleadas para la identificación de blancos de miRNA.

El desarrollo de herramientas y algoritmos que permitan la identificación de los blancos por parte de miRNA ha sido una labor constante debido a la importancia que presentan los miRNA en la regulación de la expresión de genes, y dada la complejidad que presentan las hibridaciones miRNA:mRNA, a continuación se muestran algunas características utilizadas por los programas que predicen blancos de miRNAs (Min & Yoon, 2010).

- 1. Los patrones del emparejamiento entre bases.
- 2. La estabilidad termodinámica generada por la hibridación miRNA:mRNA.
- 3. Buscar conservación en las secuencias mediante análisis comparativo de la secuencia
- 4. Presencia de múltiples sitios blanco

Sin embargo, los algoritmos actuales usan alguna de las características antes mencionadas para la búsqueda de sitios blancos y por lo cual no existe un algoritmo que sea el mejor de todos; cada uno tiene sus respectivas ventajas y desventajas, la forma en la que se determina el uso de cada uno será en función de las necesidades específicas de cada usuario. A continuación, se muestra en la Tabla 2 una lista de servidores web dedicadas a la predicción de sitios blancos de miRNA con sus respectivas características.

Nombre	Base del algoritmo	Especie blanco	Bibliografía
RNAhybrid	Estabilidad termodinámica y modelos estadísticos.	Cualquiera	(Rehmsmeier, Steffen, Höchsmann, Giegerich, & Ho, 2004)
PicTar	Estabilidad termodinámica	Vertebrados	(Chen & Rajewsky, 2006; Krek et al., 2005)
TargetScan	Complementariedad de hebras	Vertebrados	(Agarwal, Bell, Nam, & Bartel, 2015)
ΡΙΤΑ	Complementariedad de hebras	Vertebrados	(Kertesz, Iovino, Unnerstall, Gaul, & Segal, 2007)
MBSTAR	Estabilidad termodinámica usando aprendizaje de maquina	Vertebrados	(Bandyopadhyay, Ghosh, Mitra, & Zhao, 2015)
StarScan	Complementariedad de hebras	Vertebrados	(Liu et al., 2015)

Tabla 2. Programas usados para la predicción de secuencias blanco de miRNA. La base del algoritmo se refiere a la característica en la cual está basado el algoritmo para su mecanismo de búsqueda, la especie blanco son aquellos organismos en los cuales ha sido usado y publicado el algoritmo para realizar alguna predicción.

Dada la gran variedad de algoritmos que existen para la predicción de sitios blancos de miRNA, la elección del programa a utilizar es de acuerdo al usuario y al problema que pretenda resolver. Durante el presente trabajo el programa a usar es RNAhybrid y cuyas características se presentan más a detalle en la sección de Metodología.

3 Antecedente directo.

3.1 Primer y único ejemplo de un riboswitch actuando tanto in cis, como in trans.

Una de las formas en las que el RNA puede modular la expresión genética es por medio de riboswitches, elementos con capacidad de reconocer a sus moléculas blanco con gran afinidad y alta especificidad y que, en función de dicho reconocimiento, cambiar la conformación de su estructura secundaria y modular la transcripción o traducción de los genes (Winkler, W. C., Cohen-Chalamish, S., & Breaker, 2002). Hasta hace un par de años la evidencia mostraba que los riboswitches únicamente actuaban en forma *cis*, es decir, a genes cercanos ubicados inmediatamente río abajo de los mismos. Sin embargo, en un trabajo se logró identificar, por primera y única vez, riboswitches capaces de regular tanto en *cis* como en *trans* (Loh et al., 2009). En *Listeria monocytogenes* un riboswitch de tipo SAM-I, llamado SreA, y que detecta a la S-adenosylmetionina, se identificó que el RNA producido por el término de su transcripción, podría actuar como un sRNA al interaccionar con la región SD de los mensajeros del gen que codifica para el regulador PrfA, inhibiendo su traducción. El riboswitch SreA

se localiza río arriba de los genes *lmo2419-lmo2418-lmo2417* que codifican para un complejo de transporte de tipo ABC (Loh et al., 2009).



Figura 5. Representación del mecanismo de acción *in trans* del SAM-I riboswitch en *Listeria monocytogenes*. El RNA producido por el término de la transcripción del SAM-I riboswitch, podría actuar como un sRNA al interaccionar con la región SD de los mensajeros del gen que codifica para el regulador PrfA, inhibiendo su traducción.

4 Justificación.

Con excepción del estudio del riboswitch SreA de *Listeria monocytogenes*, el estudio de los riboswitches bacterianos se ha circunscrito a su capacidad de regulación *in cis* de sus correspondientes genes río abajo. En la actualidad, con el desarrollo de las nuevas tecnologías de secuenciación masiva se han obtenido las secuencias completas de decenas de miles de genomas. En el presente trabajo se pretende analizar exhaustivamente los genomas bacterianos para la identificación de riboswitches similares al reportado (SreA), que pudieran actuar tanto *in cis* como *in trans*.

5 Hipótesis.

Dada la ventaja funcional de regular *in cis* e *in trans* la expresión génica por el Sadenosilmetionina de *Listeria monocytogenes*, es factible que otros tipos de riboswitches en otros organismos, también presenten este tipo de regulación dual.

6 Objetivos.

6.1 Objetivo general.

Identificar en los genomas bacterianos secuenciados los riboswitches que actúen simultáneamente *in cis* e *in trans.*

6.2 Objetivos específicos.

- I. Desarrollo de un protocolo computacional que permita identificar al SAM riboswitches SreA encontrado en el genoma de *Listeria monocytogenes*. Este análisis será un primer control positivo de la eficacia del método a desarrollar.
- II. Uso del protocolo computacional desarrollado para el análisis e identificación de SAM riboswitches homólogos a SreA en los demás organismos de la familia Listeriaceae.
- III. Uso del protocolo computacional desarrollado para el análisis e identificación de SAM riboswitches homólogos a SreA en todos los genomas actualmente secuenciados.
- IV. Uso del protocolo computacional desarrollado para el análisis e identificación en todos los genomas actualmente secuenciados de todos los riboswitches que tengan la capacidad teórica de regular simultáneamente *in cis* e *in trans* en todos los genomas actualmente secuenciados.
- V. Presentar los resultados obtenidos de una forma que pueda ser consultada fácilmente por la comunidad científica.

7 Metodología.

7.1 Utilizar la base de datos Rfam para obtener el conjunto de modelos de covarianza (CMs) reportado para cada uno de los más de 40 riboswitches actualmente caracterizados.

La base de datos de Rfam es una colección de modelos probabilísticos de covarianza que representa la conservación de secuencia primaria de nucleótidos y la conservación de posiciones de nucleótidos apareados en la estructura secundaria consenso de los miembros de una familia de RNA que son observadas en alineamientos múltiples de la secuencia y alineamientos múltiples de estructura secundaria, respectivamente. Adicionalmente, cada familia de RNA en la base de datos Rfam, está documentada con información recopilada de datos publicados en artículos científicos (Griffiths-Jones et al., 2003). De dicha base de datos lo que obtenemos son los modelos estadísticos de covarianza (CMs) obtenidos a partir de los alineamientos de secuencia y estructuras secundarias de RNA que capturan la información de la posición específica sobre cómo se conserva cada columna del alineamiento y qué residuos son los que covarían entre columnas que representan la conservación de interacciónes del tipo Watson-Crick entre nucleótidos. La señal añadida de la covariación puede ser significativa al usar CMs para búsquedas de homología en grandes bases de datos, así que en muchos casos, es más capaz de identificar homólogos del RNA que conservan su estructura secundaria más que su secuencia primaria (Kalvari, Argasinska, et al., 2018; Kalvari, Nawrocki, et al., 2018; Yoon, 2009). Esta conservación, de secuencia y de covarianza es representada por diferentes estados en Modelos Ocultos de Markov (HMM) en la base de datos Rfam que son utilizadas por el programa de análisis CMsearch de la suit de software llamada Infernal (Nawrocki & Eddy, 2013).

7.2 Identificar, mediante comparación de secuencias usando BLAST, los genes ortólogos a los del operón *lmo2419-lmo2418-lmo2417* de *Listeria monocytogenes.*

Utilizando la secuencia de los SAM riboswitches encontrados en *Listeria monocytogenes,* se utilizó la herramienta de BLAST para identificar los genes ortólogos a los del operón *Imo2419-Imo2418-Imo2417* de *Listeria monocytogenes*.

7.3 Utilizar el software Infernal (Nawrocki & Eddy, 2013) que usa los modelos de covarianza de Rfam para identificar los riboswitches en secuencias genómicas

Dentro del software Infernal, se utilizó su dependencia llamada CMsearch para la busqueda de los riboswitches en las secuencias de la base de datos de las regiones intergénicas de los genomas. Esta búsqueda se realizó para los riboswitches descritos en la Tabla 1, para los cuales se tienen sus correspondientes modelos de covarianza. Como resultado se identificaron las secuencias intergénicas con los hits más significativos respecto al modelo de covarianza.

7.4 Identificar potenciales terminadores transcripcionales Rho-independientes que pudieran ser parte de la plataforma de expresión de los riboswitches identificados.

En este análisis, utilizaremos el algoritmo desarrollado en nuestro grupo para la identificación de atenuadores transcripcionales basados en la energía de estabilidad de estructuras de RNA, la existencia de residuos consecutivos de uridina y la distancia entre estos residuos y la estructura secundaria en el RNA (Merino & Yanofsky, 2005)

7.5 Uso del programa RNAhybrid para analizar interacciones de la plataforma de expresión de riboswitches transcripcionales con la región 5' UTR de cada gen del genoma en estudio con el fin de identificar posible regulación traduccional in trans.

El algoritmo de RNAhybrid (Starmer, Stomp, Vouk, & Bitzer, 2006) se basa en una variación de la clásica predicción de la estructura secundaria del RNA, en lugar de una sola secuencia que se pliega de nuevo sobre sí misma de la manera energéticamente más favorable. RNAhybrid determina el sitio de hibridación más favorable entre dos moléculas de RNA. El algoritmo no considera las bifurcaciones (también llamadas multi-loops) entre las interacciones de miRNA con su objetivo, aumentando así considerablemente la velocidad del algoritmo. Otra característica importante es que no utiliza ningún código para estructurar o alinear las secuencias de RNA, pero implementa un algoritmo que fue diseñado específicamente para la hibridación de RNA.

8 Resultados y discusión.

8.1 Identificación de riboswitches en genomas actualmente secuenciados.

Con base en el conjunto de modelos de covarianza (CMs) obtenidos de Rfam para los riboswitches presentes en su base de datos, se analizaron las secuencias de 400 nucleótidos rio arriba de cada gen en todos los organismos depositados en la base de datos del grupo de genómica computacional para identificar todos los posibles riboswitches presentes en los organismos anotados en la base de datos KEGG: Kyoto Encyclopedia of Genes and Genomes - a la fecha de Junio de 2018-(Figura 6), con un total de 5264 organismos, de los cuales únicamente se usaron 4826 que corresponden a genomas de arqueas y bacterias.



Figura 6. Riboswitches encontrados computacionalmente mediante el software CMsearch usando los CMs de Rfam y la base de datos del grupo de genómica computacional. El tamaño de los círculos y la escala de colores se refieren a la abundancia y la cantidad de cada uno de los riboswitches respecto a todos los organismos que están agrupados en sus respectivos filos.

8.2 Riboswitches identificados hace más de 10 años y curación manual de los recién encontrados.

Haciendo una comparación con la cantidad de riboswitches descritos por (Barrick & Breaker, 2007) podemos observar el aumento tan grande en los últimos diez años tanto en la cantidad como en las nuevas clases de riboswitches (Figura 7).



Figura 7. Distribución de riboswitches del año 2007. Las dimensiones de cada esquina son proporcionales a la frecuencia con la cual aparece el riboswitch en el correspondiente grupo taxonómico. Los gama de colores refiere cada uno de los riboswitches expuestos, y el tamaño del cuadro es la frecuencia de cada uno de los riboswitches. Tomado de (Barrick & Breaker, 2007).

Sin embargo, los resultados obtenidos fueron ajustados mediante la estandarizaron de valores de corte (scores) obtenidos manualmente mediante la comparación de las secuencias obtenidas contra la base de datos del grupo (Gene Context Tool NG) (Abdala et al., 2019) y analizando conforme se asignaba un score diferente la cantidad de posibles riboswitches en función de los genes contiguos que están regulando, es decir, en función de su contexto biológico (Tabla 3).

Riboswitch		Dorcontaio	ID	Trusted
		Porcentaje	Riboswitch	cutoff
FMN (RFN element)	47	98.77	RF00050	74.2
ТРР	36	99.89	RF00059	40
yybP-ykoY leader	35	94.50	RF00080	45
SAM riboswitch (S box leader)	41	97.45	RF00162	44
Purine	50	96.71	RF00167	37
Lysine	38	99.11	RF00168	30.1
Cobalamin	35	94.22	RF00174	30
T-box leader	31	99.85	RF00230	93
glmS glucosamine-6-phosphate activated ribozyme	50	94.03	RF00234	38.5
ydaO/yuaA leader	36	99.57	RF00379	38
ykoK leader	45	97.21	RF00380	60.9
ykkC-yxkD leader	36	97.61	RF00442	72.2
Glycine	45	87.75	RF00504	32
SAM riboswitch (alpha-proteobacteria)	34	93.94	RF00521	36
PreQ1	34	95.02	RF00522	39.3
S-adenosyl methionine (SAM) riboswitch (SAM-IV)	60	93.75	RF00634	38.6
Cyclic di-GMP-I	45	91.22	RF01051	35
preQ1-II (pre queuosine)	40	95.92	RF01054	43
Moco (molybdenum cofactor)	41	92.05	RF01055	36
Magnesium Sensor	90	94.69	RF01056	41.4
S-adenosyl-L-homocysteine	43	94.38	RF01057	37.1
AdoCbl	39	85.48	RF01482	50
Mesoplasma florum (MFR)	36	89.86	RF01510	70.7
AdoCbl variant RNA	36	95.24	RF01689	35
SAM-I/IV variant	49	75.86	RF01725	38
SAM/SAH	52	94.44	RF01727	43.1
Fluoride (crcB)	41	82.84	RF01734	31
Glutamine (glnA)	57	91.58	RF01739	36.1
ZMP/ZTP (pfl)	46	89.47	RF01750	39
SMK box translational riboswitch	45	91.13	RF01767	48.3
Cyclic di-GMP-II	47	85.71	RF01786	35.1
SAM-V	60	75.00	RF01826	41

THF	50	93.97	RF01831	36.7
PreQ1-III	115	100.00	RF02680	65.3
NiCo	42	91.38	RF02683	31.1
SAM-VI	65	100	RF02885	62.8

Tabla 3. Valores de corte (scores) obtenidos del análisis manual de los genes obtenidos como posibles candidatos a presentar riboswitches contra la base de datos y los posibles genes que está regulando.

8.3 Identificación de riboswitches con atenuadores transcripcionales en su plataforma de expresión.

La forma de buscar atenuadores transcripcionales está basada en verificar cada una de las varias condiciones que debe de contener una secuencia para ser considerada como dicho elemento regulador *(Merino & Yanofsky, 2005).* Entre las condiciones más importantes y que se tomaron en cuenta en este proyecto se encuentran las siguientes:

- a) La energía libre de las estructuras secundarias que la forman (valor de corte mfe = -10 Kcal/mol).
- b) Que la estructura secundaria que puede ser formada previo al terminador transcripcional y el terminador transcripcional, sean mutuamente excluyentes.
- c) Que se encuentren en la región 5' de regulación y cercanas al gen estructural inmediatamente río arriba del mismo.
- d) Adicionalmente a las estructuras de antiterminación y terminación, el algoritmo busca la existencia de potenciales estructuras de anti-antiterminación, es decir secuencias en el mRNA que al transcribirse compitan con la formación del antiterminador favoreciendo con ello la estructura de terminación transcripcional (se utilizó la condición de tener en una secuencia mínima de 7 bases, por lo menos 5 bases uracilo contiguas, y que 6 de esas 7 bases deben ser uracilos donde la distancia entre los 5 uracilos contiguos y el otro uracilo sea máximo de una base).

Una vez obtenidas aquellas secuencias cuyo contexto biológico asegurara que efectivamente pertenece a un riboswitch, el siguiente paso fue analizar la plataforma de expresión de cada secuencia encontrada por medio del algoritmo desarrollado en el grupo de genómica computacional para identificar potenciales terminadores transcripcionales Rho-independientes que pudieran ser parte de la plataforma de expresión de los riboswitches identificados. La figura 7 muestra los resultados obtenidos en donde se observa que la abundancia de las familias de riboswitches varía significativamente, siendo la familia del riboswitch de la T-box (RF00230), la más numerosa, seguida de la familia del TPP riboswitch (RF00059).

De acuerdo con nuestro protocolo de identificación de terminadores transcripcionales Rhoindependientes, la frecuencia en que una familia de riboswitches presenta en su plataforma de expresión terminadores transcripcionales Rho-independientes varía significativamente, tal y como se muestra en la figura 8 y sus respectivos porcentajes en la figura 9. Los riboswitches con mayor frecuencia de este elemento de regulación son el riboswitch de la T-box (RF00230), el riboswitch SAM (RF00162) y el riboswitch que detecta a lisina (RF00168). A partir de este momento, únicamente aquellos riboswitches con terminadores transcripcionales se utilizaron para los análisis posteriores.



Figura 8. Distribución de la cantidad de riboswitches totales (barras rojas) con respecto a la cantidad de riboswitches que en su plataforma de expresión también forman parte de terminadores transcripcionales Rho-independientes (barras azules).



Figura 9. Distribución de los porcentajes de tipos de atenuación por parte de cada tipo de riboswitch. Atenuadores traduccionales (barras rojas) con respecto a la cantidad de riboswitches que en su plataforma de expresión también forman parte de terminadores transcripcionales Rho-independientes (barras azules).

8.4 Identificación del riboswitch SAM *sreA* en el genoma de *Listeria monocytogenes*.

Dado que *Listeria monocytogenes* tiene presentes en su genoma 7 SAM-riboswitches, se utilizó la herramienta de BLASTn (Johnson et al, 2008) para hacer el alineamiento de la secuencia presentada por como riboswitch *sreA* contra las secuencias de los SAM riboswitches presentes en *Listeria monocytogenes* (Loh et al., 2009), los resultados obtenidos se observan en la figura 10.

Los resultados obtenidos del alineamiento muestran que de los 7 SAM riboswitches presentes en *Listeria monocytogenes,* el riboswitch *sreA* corresponde a la secuencia río arriba del gen *lmo-lmo2419*, a partir de este momento se utilizara el id del gen (*lmo-lmo2419*) para futuras referencias en lugar de *sreA*.



Figura 10. Alineamiento de nucleótidos entre la secuencia de *sreA* y la secuencia del SAM riboswitch presente en el gen denominado *lmo-lmo2419* de *Listeria monocytogenes*. La figura A muestra el resultado gráfico de la porción UR río arriba del gen *lmo-lmo2419* en donde híbrido a la perfección la secuencia del riboswitch *sreA*. En la figura B se observa el alineamiento entre bases de la secuencia del riboswitch *sreA* y la porción UR del *lmo-lmo2200*.

8.5 Normalización de la MFE (minimum free energy) entre el riboswitch y la región 5' de cada gen.

Debido a las características antes mencionadas para el programa RNAhybrid sobre las demás herramientas para la predicción de sitios blancos de sRNA, como la velocidad de procesamiento, análisis termodinámicos, modelos estadísticos y que existan trabajos publicados en donde los organismos estudiados pertenecen al dominio de los procariotes, para la predicción de sitios blanco de miRNA así como para poder repetir el análisis in silico de la interacción del sRNA-mRNA hecho por (Loh et al., 2009) y validar nuestro método, se utilizó el programa de RNAhybrid para los análisis de búsqueda de sitios blanco. El análisis inicial contempló una búsqueda de parámetros para RNAhybrid y el tamaño de la región a analizar de los sitios blancos (región UTR 5' de cada gen) hasta obtener una réplica computacional de la interacción obtenida y validada experimentalmente por el grupo sueco de Loh entre el SAM riboswitch (Imo-Imo2419) y su sitio blanco (región UTR 5' del gen Imo-Imo0200 o también llamado prfA) (Loh et al., 2009). En la figura11 se muestran los resultados obtenidos en donde se observa una réplica idéntica de la interacción entre el SAM riboswitch con su gen blanco (Imo-Imo0200) al reportado en la literatura. El análisis se realizó a todas las regiones UTR 5' de cada gen presente en Listeria monocytogenes, arrojando a la interacción entre el SAM riboswitch con la región río arriba del gen prfA con la energía libre de Gibbs (DeltaG o mfe) más significativa. Con este resultado se demuestra que el protocolo computacional implementado es capaz de identificar el gen con el hit más representativo, es decir, aquellos posibles genes que pudieran ser blanco de una
regulación *in trans* por parte de un riboswitch. Adicionalmente, más adelante se mostrarán una serie de controles que le ofrecieron una mayor validez al método desarrollado.

Α	<pre>sreA:prfA-UTR -38.1 kcal/mol prfA-UTR 5 -UGU-14nts-A</pre>				c	:	GC	с		чс	cc	: АА				A-35nts- <u>AUG</u> -3'
			GCGUGACUU	U	cuuu	AACA	۰u	AA	AAUUGUUGUU	U	G	U	U	guuuu	UAGGGUAUUUU	I
					:.::	::::	:	::		:	:	:	:	:.::.		
			CGUACUGAA	G	GGAA	UUGL	JA	υu	UUAAUAACAA	Α	С	Α	Α	CGAAG	AUCCCGUAAAA	ì
	sreA	3 -UUU-85nts-A	. u	сu	ιı	I	GU	AC		ι	J C/	A CG	cc	: L	J	G-68nts-AAC-5'

sreA:prfA-UTR -38.1 kcal/mol

В

prfA-UTR 5' A C GC C AC CC AA A-28nts-AUG-3' GCGUGACUU U CUUU AACA UAA AAUUGUUGUU U G U U GUUUU UAGGGUAUUUU CGUACUGAA G GGAA UUGU AUU UUAAUAACAA A C A A CGAAG AUCCCGUAAAG SAM sreA 3' A UC U U GU AC U CA CG CC U G 5'

Figura 11. Valores de mfe obtenidos del análisis entre el riboswitch SreA y la región UTR 5' del gen *prfA* utilizando el programa RNAhybrid. En la figura A se muestra la representación del análisis experimental realizado por el grupo sueco, la figura B representa el análisis utilizando nuestro protocolo computacional.

8.6 Diferencias entre hibridaciones canonicas de sRNA, y genes con secuencias termorreguladoras en sus genes blancos.

La Tabla 4 muestra, en orden ascendente, los genes con los valores de mfe más representativas cambiando el tamaño de la longitud de la secuencia blanco tomada para el análisis con el SAM riboswitch. Se encontró que con una amplitud de por lo menos 100 nucleótidos en la región UTR 5' del gen blanco, se obtiene al hit con el gen *prfA* (*Imo-Imo0200*) como el gen con la mfe más significativa, lo cual es notoriamente más grande que la longitud de interacción reportada para la mayoría de los sRNAs y sus blancos, que es de 15 a 30 nucleótidos (Gottesman, 2005).

ID gen	Tamañ	o ventan	as en 5'	Función		
blanco	100 120 140		140			
lmo-lmo0200	х	х	х	Regulador positivo de listeriolisina		
lmo-lmo1190	х	х	х	Proteína hipotética		
lmo-lmo1559	х	х	х	Treonil-tRNA sintetasa		
lmo-lmo0641	х	х	х	Proteína hipotética		
lmo-lmo2174	Х	Х		Proteína hipotética		

Tabla 4. Genes obtenidos con las mfe más representativas entre la interacción del SAM riboswitch con los genes blancos en *Listeria monocytogenes*. El simbolo de la "X" representa si el gen en cuestión apareció como uno de los 5 genes más representativos obtenidos al variar la amplitud de la secuencia blanco a analizar.

La diferencia entre el tamaño con el que hibridan los sRNAs y nuestro riboswitch se puede explicar a que la región UTR 5' del gen blanco (*Imo-Imo0200*) es mucho más grande que la que existe en la mayoría de los mRNAs que existe un elemento termorregulador de tipo ROSE (por sus siglas en inglés, repression of heat shock gene expression element). ROSE se encuentra presente en ciertos genes cuya traducción se regula negativamente a 30 °C por medio de la formación de una estructura secundaria que bloquea la unión del ribosoma con la cadena de RNA, sin embargo, a partir de 37 °C la estructura secundaria de RNA se desestabiliza modificando su conformación y permitiendo que la secuencia anti-Shine-Dalgarno del 16S ribosomal pueda interacuar con la secuencia Shine-Dalgarno y con ello permitir el inicio de la traducción, como se oberva en la figura 12, (Meyer, Carlson, & Lucks, 2017; Narberhaus, Waldminghaus, & Chowdhury, 2006).



Reprime la traducción

Permite la traducción

Figura 12. Representación de la función del termorregulador para el gen *agsA* en *Salmonella enterica*. A 30 °C se forma una estructura que no permite la unión del ribosoma para que se lleva a cabo la traducción, a partir de 37 °C se desestabiliza la estructura secuendaria original y la nueva conformación permite el ensamble de la maquinaria ribosomal con la secuencia del gen a traducir. Tomado y modificado de (Meyer, Carlson, & Lucks, 2017).

Reportado desde el 2002 por (Mandin et al., 2002), la región UTR 5' del gen *prfA* (*Imo-Imo0200*) en *Listeria monocytogenes* contiene un termorregulador. Esta característica ocasiona que el tamaño de la secuencia blanco del SAM riboswitch (el gen *Imo-Imo0200*) tenga que ser tan amplia, llegando a ser hasta de 100 nucleótidos, ya que la secuencia proveniente del riboswitch tiene que competir con el termorregulador presente en ese gen para también ser capaz de poder regular su expresión.

La figura 13A muestra la interacción entre el sRNA *RyhB* con dos de sus blancos; se aprecia que el tamaño de la region UTR 5' en donde interacciona el sRNA es de máximo 30 nucleótidos y una mfe de máximo -20 kcal/mol.

La figura 13B representa la interacción del SAM riboswitch *lmo-lmo2419* con la región UTR 5' del gen *prfA*; se observa que el tamaño de la secuencia del gen *prfA* que interacciona con el SAM

riboswitch es mucho mayor a la presentada en la figura 13A, llegando a ser de hasta 100 nucleótidos de longitud y con una mfe (minimum free energy) de -38.1 kcal/mol.

De acuerdo a las características presentadas anteriormente, a partir de este momento para aquellos genes que presenten elementos termoreguladores en sus regiones UTR 5' el análisis de su secuencia blanco se llevó a cabo utilizando una longitud de 100 nucleótidos, y basándonos en trabajos como el de (Gottesman, 2005) para el resto de genes, la longitud de la secuencias UTR 5' considerada para el análisis de hibridación *in silico* fue de 20 nucleótidos.



Figura 13. Diferencias entre la interacción de un sRNA con su gen blanco y la interacción entre la porción de un riboswitch y un gen con un elemento termorregulador en su secuencia UTR 5'. La figura A) muestra la interacción entre el sRNA *RyhB* con dos de sus target, destacando el tamaño de la region UTR 5' en donde interacciona el sRNA las mfe obtenidas de esa interacción, tomado y modificado de (Gottesman, 2005). La figura B) representa la interacción del SAM riboswitch (sreA) con la región UTR 5' del gen *prfA*, destacan de igual forma el tamaño de la secuencia del gen *prfA* que interacciona con el SAM riboswitch y el valor de la mfe obtenida de dicha interacción, tomada y modificada de (Loh et al., 2009).

8.7 Validaciones del método y estrategia para ubicar los hits más representativos.

Con la finalidad de validar estadísticamente nuestras predicciones, repetimos nuestros análisis de hibridización *in silico* usando como sonda las secuencias permutadas de los riboswitches. La comparación de los resultados de hibridación con las secuencias originales de los riboswitches (protocolo que llamaremos de *predicciones*, de ahora en adelante) y los resultados de hibridación con las secuencias permutadas de los riboswitches (protocolo que llamaremos de *predicciones*, de ahora en adelante) y los resultados de hibridación con las secuencias permutadas de los riboswitches (protocolo que llamaremos *control negativo*, de ahora en adelante) se realizó considerando como potencial blanco de hibridización la región comprendida en los primeros 20 nucleótidos río arriba del inicio de la traducción (sitio ubicado por el codón ATG) de cada gen del organismo en donde se identificó el riboswitch, tal como establece el protocolo mencionado en la sección de Metodología de esta tesis.

8.7.1 Análisis de la potencial regulación *in trans* de otros SAM riboswitches.

Los resultados obtenidos se muestran en la figura 14. Las distribuciones de las energías obtenidas para ambos SAM riboswitches (*sje-AAV35_009465* y *spsy-AZE41_03035*) frente a las regiones UTR 5' de cada gen en el organismo de estudio. Dado que para las predicciones se espera que si el riboswitch es capaz de regular ciertos genes, la proporción debe ser mínima, es decir, menos del 5 % del total de genes tendrían que sobresalir del total, el resto de los genes debería tener un comportamiento casi idéntico al esperado para los controles negativos ya que en ambos casos ninguna de las secuencias sería capaz de estar regulando algún gen. En la figura 14A se observa que tanto las predicciones como los controles negativos tienen un comportamiento muy similar, sin embargo, en las muestras de las predicciones sobresalen un grupo de genes (encerrados en el círculo rojo) del resto de la curva con respecto a la muestra de los controles negativos, ese comportamiento es el esperado para aquellos genes candidatos a que el SAM riboswitch (*sje-AAV35_009465*) y cualquier otro riboswitch puedan estar regulando. Para la figura 14B usando otro SAM riboswitch (*spsy-AZE41_03035*) se observa un comportamiento casi idéntico entre las predicciones y los controles negativos, sin ningún grupo de genes que sobresalgan por parte de ningún grupo, indicando que el SAM riboswitch del análisis (*spsy-AZE41_03035*) no estaría regulando a ningún gen.



Figura 14. Comportamiento de dos SAM riboswitches utilizando para el análisis riboswitches aleatorizados en su secuencia (protocolo controles negativo, líneas rojas) y los riboswitches sin modificar (protocolo predicciones, líneas azules). La figura A representa un SAM riboswitch y el comportamiento esperado para el total de genes blanco y para aquellos posibles candidatos que puedan estarse viendo regulados por dicho riboswitch. La figura B representa otro SAM riboswitch cuyo comportamiento es muy similar tanto en el protocolo con los controles negativos y las predicciones, mostrando que ningún gene o grupo de genes mantiene un comportamiento sobresaliente que indicara una afectación por parte del riboswitch.

8.7.2 Método estadístico para diferenciar genes representativos.

Para los análisis posteriores es necesario tener un método que del total de genes analizados, únicamente nos quedemos con aquellos genes cuyo comportamiento es parecido al mostrado en la figura 14A. Dicho método no puede estar basado exclusivamente en un valor de corte de la energía libre de hibridación, sino que tiene que considerar la distribución del conjunto de valores de hibridación obtenidos para el universo completo de mRNAs, tal y como se muestra en la figura 14. Si tomamos los genes cuya energía es menor a -27 kcal/mol, para el análisis posterior de enriquecimiento de grupos ortólogos con el SAM riboswitch (*sje-AAV35_009465*), los genes que pasaron el valor de corte si son los genes representativos; sin embargo, con el mismo valor de corte para los genes obtenidos en el análisis con el SAM riboswitch (*spsy-AZE41_03035*), los genes que se obtienen con ese valor de corte son genes que no están regulados *in trans* por dicho riboswitch. Por lo anterior, se calculó el valor tipificado Z de cada observación (cada energía libre obtenida) con el fin de poder comparar los datos en distintas distribuciones, basándonos en puntuar cada dato (energía libre obtenida) en unidades de desviación estándar de acuerdo a la ecuación de la figura 15 (Ritzema, 1994).

$$z = \frac{valor mfe-media}{std}$$

Figura 15. Ecuación para obtener el valor tipificado Z de cada mfe. Se obtiene mediante el cociente de la resta de cada valor menos la media de la distribución entre su desviación estándar.

Mediante la obtención del valor Z para cada valor de mfe, lo que estamos haciendo es quedarnos con aquellos valores cuyo comportamiento se encuentre lo más alejado y diferente del total de los valores obtenidos.

Una vez teniendo el método computacional junto con el método estadístico que nos asegura quedarnos únicamente con aquellos genes cuya potencial interacción con el riboswitch tuviera valores de mfe realmente significativos. El siguiente paso fue corroborar que el método fuera adecuado al ser aplicado a una mayor cantidad de datos y con otros riboswitches. Por lo cual, de acuerdo a los objetivos del proyecto, se tomaron al grupo de las Listerias representativas (organismos con la mayor cantidad de genes codificantes para proteínas) de la familia Listeraceae, en total fueron 6 las *Listerias* que su usaron (ver anexo 1), los SAM riboswitches sin modificar y los SAM riboswitches aleatorizados para hacer el análisis con el método computacional mencionado anteriormente.

La figura 17 muestra los resultados obtenidos del análisis con los riboswitches sin modificar y los riboswitches aleatorizados para las listerias representativas de la familia Listeraceae con una ventana de 20 nucleótidos. Los genes encontrados con mayor significancia se agruparon en sus respectivos Clusters of Orthologous Groups of proteins (COGs), y se realizó un análisis de enriquecimiento de las señales para obtener el significado estadístico del mismo a través del valor p-value de dicho enriquecimiento y considerando una distribución hipergeométrica calculada usando una librería de R (figura 16) para cada grupo COGs con señales de regulación obtenidos.

$$P(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Figura 16. Ecuación para realizar el cálculo de la distribución hipergeométrica. Se utilizó una librería de R, donde *N* = genes totales, *n* = hits totales por COG, *K* = suma de los hits obtenidos por COG y *x* = hits obtenidos por COG



8.8 Aplicación del protocolo computacional obtenido, usando pequeños volúmenes de datos.

Figura 17. Protocolo computacional aplicado con SAM riboswitches (sin modificar y con los aleatorizados) a organismos representativos de la familia Listeraceae. La figura A, representa los resultados de los genes obtenidos utilizando un valor de corte de $Z \ge 1$ y agrupados en sus respectivos COGs. La figura B, representa los resultados de los genes obtenidos utilizando un valor de corte de $Z \ge 1.5$ y agrupados en sus respectivos COGs. La figura C, representa los resultados del análisis usando los riboswitches aleatorizados y se muestran los genes obtenidos utilizando un valor de corte de $Z \ge 1$ y agrupados en sus respectivos COGs.

En la figura 17C se muestra el resultado del protocolo computacional, utilizando una ventana de 20 nucleótidos, las secuencias del riboswitch aleatorizado y un valor de Z = 1 que corresponde a un p-value de 0.1587. Para dicha figura se aprecian dos COGs, el COG4496 y el COG3272 cuya función de los genes pertenecientes es de "proteínas sin caracterizar". De manera interesante, si se considera un valor de corte de Z \ge 1.5 para obtener genes significativos, que correspondería a un p-value de 0.0668, un valor mucho más significativo y representativo que el obtenido con el valor de corte de Z \ge 1, el

enriquecimiento de COGs es nulo, es decir, ya no se obtiene ningún COG representativo en nuestro experimento control con la secuencia del riboswitch permutado aleatoriamente.

Las figuras 17A y 17B, corresponden al análisis con los SAM riboswitches sin modificar de los organismos representativos de la familia Listeraceae y ventana de 20 nucleótidos de la región UTR 5' de los genes blanco. La diferencia entre ambas imágenes es el valor de corte de Z que se tomó para cada una. En la figura 17A el valor usado fue de $Z \ge 1$, que equivale a un p-value de 0.1587, y en la figura 17B el valor de $Z \ge 1.5$, que es un p-value de 0.0668. Los genes obtenidos en el análisis, se agruparon dentro de COGS referentes a hidrolasas de la súper familia alfa/beta, sistema de transporte de tipo ABC y predicciones de glutamina amidotransferasas. Es importante recordar que los SAM riboswitches regulan la expresión de genes relacionados a proteínas transportadoras que pertenecen al tipo de transporte ABC y que por lo tanto, el resultado que obtuvimos es biologicamente significativo.

Lo importante del análisis es la obtención de un COG cuando el valor de Z obtenido en el análisis de enriquecimiento es mayor a 1.5, ya que en el análisis con los controles negativos (secuencias aleatorizadas de los riboswitches) no se obtuvo ni un solo COG con ese valor de corte. Adicionalmente a este criterio estadístico, nuestro estudio también verificó que los grupos de ortología identificados como enriquecidos tuvieran un sentido biológico, es decir, que su función estuviera relacionada con el metabolito o señal reconocida por el riboswitch de estudio.

De igual modo, el protocolo fue aplicado, pero ahora usando otro riboswitch, el riboswitch Tbox que tiene la característica de detectar el nivel aminoacilación de los tRNAs. Existen variantes del riboswitch T-box que son específicas para cada tipo particular de tRNA. En caso de haber una concentración elevada de tRNA descargado, el tRNA correspondiente interactúa específicamente con la T-box estabilizando un anti-terminador transcripcional permitiendo que la transcripción continúe, en caso contrario, se favorece la formación de un terminador transcripcional (ver figura suplementaria 6). El estudio realizado con este riboswitch contempló a los 36 organismos representativos por género que pertenezcan al orden Bacillales. La lista de dichos organismos se incluye en el Anexo 2 de esta tesis. Con las condiciones mencionadas se utilizó el protocolo computacional desarrollado para 671 riboswitches de la familia T-box y los resultados se muestran en la figura 18. Tal y como se esperaba, los COGs que aparecen en los controles negativos tuvieron un pvalue poco significativo y las funciones de dichos COGs no tuvieron un contexto genetico relacionado con los genes canónicamente regulados por este riboswitch que son la de aminoacilación de tRNAs, transporte o síntesis de aminoácidos. Con un valor de corte de Z ≥ 1.5, solamente se obtuvieron dos COGs: El COG3676, correspondiente a una transposas y el COG2072 correspondiente a una flavoproteínas. Cuando el valor de corte fue Z ≥ 1.5, no se obtuvo ni un solo COG en nuestro control negativo. De acuerdo a los resultados obtenidos podemos estar seguros que el protocolo funciona y podemos utilizar de forma general para futuros análisis un cutoff del valor de Z ≥2, o que el valor del – log de p-value de los COGs sea \geq 2.5, así nos aseguramos que al final solo aquellos genes presentes en los COGs obtenidos puedan ser candidatos a ser regulados por una acción *in trans* por parte de un Tbox.

Los COGs obtenidos cuyos p-value los hace más representativos tanto para valores de Z \ge 1.5 y de Z \ge 2, corresponden en su mayoría a funciones relacionadas con proteínas de membrana (ver figura 18), lo que tiene sentido biológico dado la función de regulación del riboswitch de la T-box. Cuando en la célula existe la suficiente cantidad de tRNA cargados con aminoácidos, el tRNA no se une a la T-box y se genera el terminador transcripcional de la plataforma de expresión para finalizar la transcripción del operón (Green et al., 2010). Los resultados sugieren que si la célula está en ambientes favorables donde la mayoría de sus tRNA están cargados, el riboswitch no solamente pudiera regular negativamente *in cis* a los genes antes mencionados involucrados en la síntesis de aminoácidos y aminoacilación de tRNAs, sino que también podría existir una regulación negativa *in trans*, cuando el riboswitch de la T-box prudujera un tipo de sRNA por la formación de la estructura del terrminador transcripcional que pudiera inhibir la traducción de mRNAs de los genes que sintetizan proteínas de membrana que pudieran corresponder a transportadores de aminoácidos



Figura 18. Protocolo computacional usando ventana de 20 bases aplicado con T-box (sin modificar y con los aleatorizados) a organismos representativos del orden Bacillales. Las muestras del Z control negativo, se refieren a los COGs obtenidos del análisis con la T-box aleatorizada, se aprecia que únicamente aparecen dos COGs y no se muestra a una Z \geq 1.5. Debido a que no se obtuvo ningún grupo de genes ortólogos COG representativo con un valor de corte de Z \geq 2, no se incluye en la figura el resultado del control negativo para este valor. Los COGs obtenidos con un valor de corte de Z \geq 2 y Z \geq Z 1.5 son casi los mismos Z, aunque el valor de p-value varía.

COGs	Función
COG2707	Proteína de membrana
COG3949	Proteína sin caracterizar de membrana
COG1836	Proteína de membrana
COG4826	Inhibidor de serin proteasas
COG2072	Flavoproteínas*
COG3676	Transposasas y derivados inactivos*

Tabla 6. COGs y su función obtenidos en el análisis a organismos representativos del orden Bacillales usando la T-box. Se usó la secuencia sin modificar y la secuencia aleatorizada, representados únicamente los COGs cuyo (-log p-value \geq 1.5). *Se refiere a la función de los COGs obtenidos para el análisis con las muestras aleatorizadas, el protocolo para controles negativos.

Sin embargo, a pesar de los controles mostrados hasta el momento, donde se han utilizado riboswitches que están presentes en el mismo organismo y su secuencia aleatorizada para el análisis, lo mostrado en la figura 19 corresponde a un análisis en donde se tomaron los T-box leader presentes en *Bacillus subtilis* así como su secuencia aleatorizada para hacer el mismo análisis pero ahora en el genoma de *Escherichia coli*, con el fin de poder analizar el comportamiento del programa utilizando un riboswitch que un organismo no tiene presente en su genoma (como es el caso de la T-box leader en *Escherichia coli*).





Figura 19. Protocolo computacional aplicado con T-box leader (sin modificar y con los aleatorizados) de *Bacillus subtilis* en un genoma que no contiene dichos riboswitches como es el caso de *Escherichia coli*. Las figuras A, B, C, D, E y F corresponden a las T-box leader de *Bacillus subtilis* denominados bsu-BSU38460, bsu-BSU28950, bsu-BSU28090, bsu-BSU23800, bsu-BSU00930 y bsu-BSU00130 respectivamente. Las líneas rojas corresponden a la secuencia aleatorizada del riboswitch en cuestión, y la línea azul al riboswitch sin modificaciones.

Los comportamientos obtenidos se ajustan a los esperados, ya que ni el riboswitch que pertenece a otro organismo ni su secuencia aleatorizada pueden ser capaces de poder regular genes en otro organismo, el comportamiento debe ser muy semejante entre ambos análisis, sin embargo, para corroborar los comportamientos obtenidos se analizaron los resultados buscando genes que pudieran ser representativos por medio del valor Z, sin obtener algún resultado con los valores de 0.5,

1, 2, 3 para Z lo cual corrobora la eficiencia del método para la predicción de posibles genes blancos que puedan estar siendo regulados *in trans* por algún riboswitch.

8.9 Análisis de la potencial regulación *in trans* del SAM riboswitch.

Debido a lo mostrado anteriormente, podemos tener le certeza de que el protocolo computacional es confiable y los resultados obtenidos tienen validez estadística y biológica. A continuación, se muestran los resultados obtenidos del protocolo usando al SAM riboswitch con los organismos representativos de la base de datos utilizando una ventana de 20 bases para los genes blanco. La figura 21 representa los COGs obtenidos del análisis. En la tabla 7 se muestra que la función del COG0527 corresponde a una aspartato quinasa, enzima participante en el metabolismo de la cisteína y metionina, ambos aminoácidos regulados por el SAM riboswitch. La función del COG0411 es muy importante ya que transporta del exterior de la célula al interior, aminoácidos. Esto, sugiere un sentido biológico a una posible regulación *in cis e in trans* por parte del SAM riboswitch, de tal forma que si la célula tiene un exceso de aminoácidos, se lleva a cabo una regulación negativa *in cis e in trans* por parte del SAM riboswitch en el metabolismo y transporte de aminoácidos relacionados con la S-adenosilmetionina. En la figura 20 se observa la participación del COG0527 durante el metabolismo de la metionina y cisteína.



Figura 20. Metabolismo de la metionina y cisteína. Se resalta la aspartato sinasa de la vía metabólica. Imagen tomada del KEGG -a la fecha de Junio 2019-



Enriquecimiento de COGs en SAM riboswitch

Figura 21. Protocolo computacional aplicado con SAM riboswitches a organismos representativos por género de la base de datos. Se utilizaron un total de 1051 organismos, los genes obtenidos con un valor de $Z \ge 2$ se agruparon en sus respectivos COGs.

COGs	Función
COG5454	Proteína secretora
COG4463	Represor transcripcional clase 3 (genes estress)
COG1663	Tetraacildisacarido 1-P 4- quinasa
COG1294	Citocromo bd quinol oxidasa
COG0527	Asparto quinasa
COG0411	Transportador de aminoácidos, tipo ABC

Tabla 7. COGs obtenidos en el análisis para el SAM riboswitch a organismos representativos por género de la base de datos., así como la función respectiva de los COGs más representativos en todos los análisis. Datos obtenidos del KEGG -a la fechad de Junio 2019-.

Los COGs 5454, 4463, 1663 y 1294, a la fecha no se ha podido encontrar una relación entre sus funciones y la parte del metabolismo regulado por parte de los SAM riboswitches, sin embargo, no se

descarta la posibilidad de que próximas investigaciones encuentren alguna relación y le otorguen una mayor validez a los resultados mostrados en este trabajo.

8.10 Análisis de la potencial regulación *in trans* del riboswitch de la T-box.

De todos los riboswitches, la T-box es el riboswitch que tiene la mayor cantidad de secuencias presentes en los organismos, como se observa en la figura 6, con poco más de 15 mil secuencias. Dada la amplia distribución y la gran cantidad de secuencias, el protocolo se utilizó con la T-box, los 1050 organismos representativos por genero presentes en la base de datos, una ventana de 20 bases y un valor de Z \ge 2 para la obtención de los genes con los cuales se realizó el enriquecimiento de COGs. El tiempo de cómputo empleado en este análisis fue de una semana usando un cluster de computadoras con 5 nodos de 64 cores y 512 Gb de memoria. También se utilizó un cluster de 1 nodo de 64 cores y 512 Gb de memoria para distribuir y agilizar la carga en ambos servidores.





Figura 22. Protocolo computacional aplicado riboswitch T-box con organismos representativos por género de la base de datos. Se utilizaron un total de 1051 organismos, los genes obtenidos con un valor de $Z \ge 2$ se agruparon en sus respectivos COGs.

La figura 22 es una representación gráfica de los resultados obtenidos usando el protocolo con las T-box. El COG3935 y el COG1933, corresponden a funciones sin relación directa demostrada; sin embargo, no se descarta la idea que posteriormente los avances científicos encuentren relación entre los COGs mencionados y la T-box. Se aprecia una mayoría de funciones relacionadas con proteínas de membrana, de acuerdo a la tabla 8, tomando en cuenta la función de la T-box de encargarse de regular genes que sintetizan aminoacil-tRNA y la biosíntesis de aminoácidos (Green et al., 2010), mediante la unión de un tRNA cargado o descargado de su respectivo aminoácido al aptamero del riboswitch, puede que las funciones de proteínas obtenidas en el enriquecimiento de COGs correspondan a genes que intervienen en el transporte de aminoácidos, para así lograr un equilibrio en la célula y de esta manera regular de forma directa la cantidad de tRNAs que se encuentran cargados y que pueden actuar en la T-box.

COGs	Función
COG2707	Proteína de membrana
COG3949	Proteína de membrana
COG3935	Componentes del primosoma
COG1933	RNA polimerasa
COG3379	Proteína sin caracterizar
COG1300	Proteína de membrana

Tabla 8. COGs obtenidos en el análisis para la T-box a organismos representativos por género de la base de datos, así como la función respectiva de los COGs más representativos en todos los análisis (-log p-value \ge 2). Datos obtenidos del KEGG -a la fechad de Junio 2019-.

Algunos de los genes obtenidos más representativos que se obtuvieron con el protocolo usando el SAM riboswitch y la T-box leader (Figura S1) corresponden a genes cuya función mantiene relación biológica con las funciones de regulación por parte de los SAM riboswitches y la T-box leader. El primer caso de estudio (Figura S1A), corresponde a la interacción obtenida entre un gen regulado por un SAM riboswitch (sth-STH1685) y su respectivo gen blanco (sth-STH703) cuya función corresponde a la de un transportador de aminoácidos de tipo ABC (Figura S2), indicando que mientras la célula tenga suficientes nutrientes, el SAM riboswitch actuara tanto in cis como in trans tratando de reducir la entrada de aminoácidos para mantener un equilibrio nutricional dentro de la célula. En la figura S1B, se aprecia la interacción obtenida entre el SAM riboswitch (tmr-Tmar 1035) y su gen blanco obtenido (tmr-Tmar_0571) cuya función es la de una aspartato sinasa. La relevancia biológica de esta potencial regulación en que el producto codificado por el gen tmr-Tmar 0571 es fundamental en el metabolismo de la cisteína y metionina (Figura 20), aminoácidos precursores de metabolitos como el SAM que se encarga de donar grupos metilo para diversas reacciones, mostrando que el SAM riboswitch actuando in trans, puede regular negativamente la producción del metabolito que se une al riboswitch (S-Adenosilmetionina). La regulación de este proceso es razonable si se toma en cuenta que para que se lleve a cabo la regulación in trans es porque en la célula se encuentran altos niveles de S-Adenosilmetionina, así que una forma de disminuir esos niveles es regulando el metabolismo de aquellos aminoácidos que forman parte del metabolismo del SAM. Con el mismo protocolo se obtuvo la interacción (Figura S1C) entre un SAM riboswitch (*Imob-BN419_0699*) con una ventana de 100 sobre la región UTR 5' del gen blanco (Imob-BN419_1598), su función corresponde a la de una 1-deoxy-D-xilulosa-5-fosfato sintasa, enzima involucrada en la biosíntesis de tiamina, formando parte de reacciones reversibles involucradas en la reducción de SAM radicales para generar radicales 5-deoxyadenosil, es decir, en exceso de aminoácidos (específicamente tiamina) se van a producir especies radicales SAM, por lo cual es lógico pensar que el SAM ribsowitch inhiba la producción de la 1-deoxy-D-xilulosa-5-fosfato sintasa para que no se generen una mayor cantidad de especies SAM ya que la célula ya las contiene en altas concentraciones.

En las figuras S1D- al S1E, se presentan algunos genes potencialmente regulados in trans por el riboswitch de la T-box leader (usando una ventana de 20 bases, la interacción presentada en la figura S1D corresponde a un T-box leader (axl-AXY_0650) con su respectivo blanco (axl-AXY_23920). En este caso, cabe destacar que la función del potencial gen blanco corresponde a un transportador de Lcisteina, que podría no ser requerido, y por ende no sintetizado cuando la célula está en condiciones favorables y tiene la capacidad de tener una gran cantidad de de sus tRNA están cargados; de esta forma el riboswitch tiene la capacidad de regular negativamente in cis a los genes involucrados en la síntesis de aminoácidos y aminoacilación de tRNAs, y también podría existir una regulación negativa in trans, cuando el riboswitch de la T-box prudujera un tipo de sRNA por la formación de la estructura del terminador transcripcional que pudiera inhibir la traducción de mRNAs de los genes que sintetizan proteínas de membrana que pudieran corresponder a transportadores de aminoácidos, específicamente de L-cisteína (Figura S5). Con la figura S1E se pretende hacer énfasis en la relevancia de los procesos implicados durante una posible regulación in trans por parte de un T-box-leader, ya que el protocolo usado con T-box (*qth-Geoth 0822*) obtuvo al gen blanco (*qth-Geoth 0088*) con una función de transportador de aminoácidos, específicamente de L-cisteína (Figura S5.1), denotando la relevancia de regular tanto in cis como in trans por parte de la T-box aquellos procesos involucrados en el transporte y síntesis de aminoácidos.

8.11 Análisis de la potencial regulación *in trans* del TPP riboswitch.

El siguiente riboswitch que cuenta con la mayor cantidad de riboswitches presentes y distribuidos a lo largo de una gran cantidad de organismos, es el TPP riboswitch. Al utilizar el protocolo usando una ventana de 20 y de 100 bases con la familia de TPP riboswitch, sorprendentemente, no se obtuvo ni un solo COG al momento de hacer el análisis del enriquecimiento de señales dentro de los COGs, pese a que esta familia de riboswitches es la más numerosa, después del riboswitch de la T-box (Ver Figura 7).

8.12 Análisis de regulación in trans de riboswitches considerando como potenciales blancos genes con grandes regiones UTR´s.

Los análisis anteriormente descritos, consideraban como potencial blanco de regulación la región inmediatamente río arriba del codón de inicio de la traducción ATG interaccionando y secuestrando la región correspondiente al sitio de unión a ribosoma, también llamada secuencia Shine-Dalgarno. No obstante, como se comentó en la Introducción, el único antecedente de regulación *in trans* de riboswitches se da en un gen con un elemento ROSE en su región 5'UTR y para su identificación computacional fue necesario utilizar una ventana de interacción de 100 bases. El protocolo desarrollado utilizando una ventana de 100 bases con la T-box, el SAM y el TPP riboswitche se muestran en la figura 23 y los resultados obtenidos se muestran en la tabla 8. Es importante hacer notar que para este análisis, el número de potenciales genes blanco es muy inferior al estudio con la ventana de 20 nucleótidos en la que se analizaron la las UTR's 5' de todos los genes de los organismos. En términos generales, se encontraron de uno a dos genes regulados por el elemento ROSE por organismo, lo que equivale a menos del 0.1% del total de genes. Los controles utilizados para el protocolo usando una ventana de 100, fueron las mismas secuencias de los riboswitches aleatorizados (controles negativos), debido a que es mínima la cantidad de genes que presentan una secuencia termorreguladora en su región UTR 5', no la distribución y el uso del valor tipificado de Z quedo descartado. Sin embargo, finalizado el análisis ningún gen en los controles negativos se obtuvo, por tal razón los genes obtenidos con las secuencias correctas de los riboswitches se utilizaron directamente para hacer el análisis de enriquecimiento de señales en los grupos COGs.



COGs	Función
COG1154	Desoxixilulosa-5-fosfato sintasa
COG1544	Proteína asociada a ribosoma
COG0624*	Succinil-diaminopimelato desuccinilasa

Tabla 9. COGs y su respectiva función obtenidos en el análisis a genes blanco con grandes regiones UTR's, usando T-box, SAM y TPP. Datos obtenidos del KEGG -a la fecha de Junio 2019-.

* Corresponde al COG obtenido con la T-box

Figura 23. Protocolo computacional aplicado a genes blanco con grandes regiones UTR's, usando la T-box, SAM y TPP con organismos representativos por género de la base de datos. Se utilizaron un total de 1051 organismos. Los genes obtenidos se agruparon en sus respectivos COGs.

Se muestra una correlación entre los resultados obtenidos utilizando el protocolo independiente de la ventana de 20 y 100 bases. El COG1154 obtenido para el SAM riboswitch corresponde al gen presentado y explicado anteriormente durante el análisis a fondo de algunos de los genes obtenidos más representativos (Figura S1C).



00300 6/23/17 (c) Kanehisa Laboratorie:

Figura 24. Metabolismo de la Lisina. Se resalta la de la vía metabólica la succinil-diaminopimelato desuccinilasa en rojo y en verde el final de la síntesis de lisina. Imagen tomada del KEGG -a la fecha de Junio 2019-

Utilizando la T-box, el COG0624 corresponde a una succinil-diaminopimelato desuccinilasa, enzima encargada de los últimos pasos de la síntesis de la lisina, la figura 24 muestra en rojo el lugar donde actúa la enzima y en verde se observa el término de la síntesis de la lisina, teniendo una relación directa con los resultados obtenidos para la T-box al utilizar el protocolo con ventana de 20 bases. Dentro de los cuales se encontró que existe una posible relación entre el mecanismo y las funciones de regulación *in cis* de la T-box y aquellos genes que pudiera estar regulando a distancia, es decir *in trans,* que se relacionan con el transporte y metabolismo de aminoácidos, en este caso de la lisina.

Para el TPP riboswitch ocurrió algo similar a lo obtenido con el mismo protocolo pero usando una ventana de 20 bases, ningún gen apareció al usar una ventana de 100 bases en aquellos genes con secuencias ROSE en sus UTR 5'. El TPP riboswitch se encuentra distribuido en la mayor cantidad de organismos, llegando a presentarse incluso en algunos organismos eucariotas como hongos y plantas, quizá la divergencia evolutiva que le permitió aparecer en una gran cantidad de organismos, es también la razón por la cual de acuerdo a nuestros resultados, no tiene la capacidad de poder regular la expresión de genes *in cis* e *in trans*.

8.13 Ejemplos de doble regulación (*cis* e *in trans*) por parte de diferentes riboswitches.

Con el protocolo usando la T-box y ventana de 20 bases, se obtuvieron los primeros casos descritos hasta el momento donde un riboswitch puede actuar *in trans* y el gen blanco también se ve regulado por otro riboswitch que reconoce al mismo ligando.



Figura 25. Mecanismo de regulación doble (*in cis* e *in trans*) por parte de un mismo tipo de riboswitch sobre un gen blanco. En la figura se aprecia que un T-box, que reconoce a tRNA cargados con isoleucina, y que regula *in cis* un gen relacionado a una isoleucil tRNA sintetasa, implicada en la síntesis de isolecuina. Posteriormente el transcrito corto producto de la transcripción del mismo riboswitch puede regular *in trans* al mRNA del gen que codifica para una 2-isopropilmalato sintasa (enzima participante en la síntesis de isoleucina). Aparte el gen de la 2-isopropilmalato sintasa es regulado *in cis* por una T-box que reconoce a un tRNA cargado con isoleucina

En la figura 25 se aprecia un T-box, que reconoce a tRNA cargados con isoleucina, y que regula *in cis* un gen relacionado a una isoleucil tRNA sintetasa, implicada en la síntesis de isolecuina. Posteriormente el transcrito corto producto de la transcripción del mismo riboswitch puede regular *in trans* al mRNA del gen que codifica para una 2-isopropilmalato sintasa (enzima participante en la síntesis de isoleucina). Aparte el gen de la 2-isopropilmalato sintasa es regulado in cis por una T-box que reconoce a un tRNA cargado con isoleucina. El ejemplo anterior muestra una doble regulación a genes involucrados en el metabolismo de la isoleucina. Una doble regulación por parte del mismo riboswitch a procesos involucrados en el metabolismo de un aminoácido en específico, involucran un mayor control para la célula de sus recursos. En las tablas suplementarias se muestran los resultados obtenidos del análisis con el SAM-I y la T-box que involucran procesos semejantes al mostrado en la figura 25.

La tabla suplementaria 1 muestra todos los COGs obtenidos en el enriquecimiento de COGs usando el protocolo computacional con el SAM riboswitch, también se muestran los riboswitches que pueden estar regulando *in cis* (genes del lado izquierdo) a genes pertenecientes (genes del lado derecho, son los genes que se ven afectados por una regulación *in trans*) a los COGs mostrados, mostrando que existen más ejemplos de genes que se ven doblemente regulados por diferentes tipos de riboswitches y del mismo tipo.

9 Conclusiones.

Con los métodos computacionales basados en modelos de covariancia hemos logrado detectar a la mayoría de los riboswitches presentes en cada organismo secuenciados y determinar para cada uno de ellos el tipo de atenuación, transcripcional o traduccional, a los que están sujetos. En el caso de aquellos riboswitches regulados transcripcionalmente, hemos analizado si su terminación prematura de la transcripción pudiera generar sRNAs capaces de regular in trans la expresión de genes localizados a distancia. Para realizar lo anterior, hemos evaluado la ventana de análisis de la secuencia 5' UTR de cada gen para establecer aquella que permitiera discernir los candidatos positivos previamente publicados de los verdaderos negativos, en base a valores de mfe de interacción del sRNA con sus respectivos blancos. Durante los análisis exploratorios para identificar el riboswitch con la capacidad de actuar in cis e in trans que Loh y colaboradores demostraron experimentalmente en 2009, pero ahora de manera computacional usando nuestro método, encontramos estructuras de tipo ROSE que afectan la interacción entre el sRNA y su blanco, encontrando dos tamaños de ventana. Para genes que presenten elementos termoreguladores (ROSE) en sus regiones UTR 5' el análisis de su secuencia blanco se llevó a cabo utilizando una longitud de 100 nucleótidos, y basándonos en trabajos como el de (Gottesman, 2005) para el resto de genes la longitud de la secuencias UTR 5' fue de 20 nucleótidos.

Nuestro estudio inicial de la regulación *in trans* del riboswitch SAM en el género Listeria fue extendido a todos los géneros de bacterias y arqueobacterias. Para ello, se obtuvo una muestra representativa de organismos únicos por género, dando lugar a 1051 organismos de estudio. El protocolo computacional desarrollado obtiene genes potencialmente regulados *in trans* por el riboswitch usado para el análisis, sin embargo, la manera de seleccionar únicamente aquellos genes que realmente fueron representativos fue mediante el cálculo del valor tipificado Z de cada observación (cada energía libre obtenida) con el objetivo de comparar los datos en distintas distribuciones, con base en puntuar cada dato (energía libre obtenida) en unidades de desviación estándar. Con la obtención del valor Z para cada valor de mfe, se logró discriminar aquellos valores cuyo comportamiento se encuentre lo más alejado y diferente del total de los valores obtenidos. Finalmente, los genes seleccionados fueron agrupados de acuerdo a sus relaciones de ortología en base a la clasificación de genes ortólogos COGs.

10 Perspectivas.

Los controles usados tanto las secuencias aleatorizadas de los riboswitches y los riboswitches usados contra el genoma de un organismo que no contiene esa clase de secuencias, demuestran que el protocolo computacional es muy eficiente para la identificación de genes con relevancia biológica que puedan estarse viendo regulados por una actividad *in trans* de un riboswitch. Finalmente, un análisis estadístico del enriquecimiento de señales positivas de los grupos de ortología COG nos permitió identificar grupos de genes ortólogos con relevancia biológica.

El trabajo posterior se enfocará en incrementar la eficacia del método, modificando y agregando hiper parametros con el objetivo de mejorar la sensibilidad del método, puesto que muchas herramientas actuales emplean métodos de aprendizaje de maquina obteniendo resultados más específicos, se pretende crear un protocolo computacional que utilice herramientas de aprendizaje de maquina con el fin de mejorar los resultados obtenidos.

Los resultados presentes muestran COGs enriquecidos por genes representativos que pueden verse afectados *in trans* por un riboswitch, los genes encontrados son candidatos para estudios experimentales donde se pueda comprobar la relación entre los resultados obtenidos computacionalmente y los resultados que se puedan obtener experimentalmente.

En su conjunto, el trabajo muestra un protocolo computacional funcional para la identificación de riboswitches que puedan actuar *in trans*. De igual forma, es una poderosa herramienta que puede usarse con los datos de las nuevas versiones de base de datos de KEGG, así como sus métodos estadísticos en la obtención de candidatos representativos de cualquier mezcla homogenea y la obtención de secuencias candidatas con la capacidad de regular la expresión de genes localizados a distancia.

Apéndice A

Hibridaciones particulares de algunos riboswitches y metabolismo afectado de los genes blanco.





Figura S1. Ejemplos de algunas de las hibridaciones de los genes más representativas obtenidos con el SAM riboswitches y la T-box leader. Las figuras A, B y C corresponden al protocolo computacional usando un SAM riboswitch, las figuras A y B corresponden a una ventana de 20 bases y la figura C a una ventana de 100 bases. Las figuras D y E se obtuvieron usando el T-box leader y una ventana de 20 bases.



Figura S2. Contexto genomico del gen *STH703*. Que corresponde a un transportador de aminoácidos de tipo ABC, el gen se encuentra encerrado en rojo. Imagen tomada y modificada de la base de datos de la página (Gene Context Tool NG) (Abdala et al., 2019).



Figura S3. Metabolismo de la metionina y cisteína. Se resalta la aspartato sinasa de la vía metabólica. Imagen tomada del KEGG -a la fecha de Junio 2019-.



Figura S4. Metabolismo de la Tiamina. Se resalta la 1-deoxy-D-xilulosa-5-fosfato sintasa en verde y los pasos generadores de SAM en rojo de la vía metabólica. Imagen tomada y modificada de (Julie A. Maupin-Furlow, 2018).



Phosphate and amino acid transporters

Figura S5. Gen *axI-AXY_06500* con función de transportador de aminoácidos de tipo ABC, específicamente del sistema de transporte de L-cisteina. Datos tomados del KEGG –a fecha de Junio de 2019-.



Figura S5.1. Gen *gth-Geoth_0088_06500* con función de cisteína sintasa. Datos tomados del KEGG –a fecha de Junio de 2019-.



Figura S6. Mecanismo de regulación por parte de una T-box. Durante la transcripción de una región por parte de la RNA polimerasa (óvalos rojos), el producto de RNA se pliega en una estructura competente para la unión del tRNA afín en su sitio de unión. La unión del tRNA no cargado (imagen superior) con el riboswitch, estabilizan el antiterminador (segmento verde) evitando la formación del terminador de la transcripción. Cuando el tRNA está cargado y se une al riboswitch (figura inferior), no se puede estabilizar el antiterminador y se forma la hélice del terminador (segmento de RNA rojo), y la transcripción se termina antes de que se pueda transcribir la región codificante aguas abajo. Tomado y modificado de (Gutierrez-Preciado & Merino E., 2009).

COG0007 Uroporphyrinogen-III methylase					
brl-BZG35_14625 brl-BZG35_07190					
COG0111 Phosphoglycerate dehydrogenase and related dehydrogenases					
hsw-Hsw_3707	hsw-Hsw_0588				
COG0123 Deacetylases, including yeast histone deacetylase and acetoin utilization protei					
deu-DBW_3287 deu-DBW_2618					
COG0183 Acetyl-CoA acetyltransferase					
tmr-Tmar_0571	tmr-Tmar_0391				
COG0436 Aspartate/tyrosine/aromatic aminotransferase					
brl-BZG35_14625	brl-BZG35_07960				
COG0454 Histone acetyltransferase HPA2 and related acetyltransferases					

deu-DBW_3287	deu-DBW_2618							
sth-STH1685	sth-STH3111							
COG0457 FOG: TPR repeat								
kur-ASO14_792	kur-ASO14_2446							
sth-STH1685	sth-STH1018							
COG0477 Permea	ses of the major facilitator superfamily							
cwo-Cwoe_3191	cwo-Cwoe_5556							
tmr-Tmar_0571	tmr-Tmar_1924							
tmr-Tmar_0571	tmr-Tmar_2038							
COG0526 Thiol-di	sulfide isomerase and thioredoxins							
lpil-LIP_3402	lpil-LIP_1416							
COG0527 Asparto	kinases							
tmr-Tmar_0571	tmr-Tmar_1035							
COG0531 Amino a	acid transporters							
hsw-Hsw_3707	hsw-Hsw_2340							
COG0600 AB	C-type nitrate/sulfonate/bicarbonate transport system, permease							
component								
lpil-LIP_2986	lpil-LIP_0881							
COG0601 ABC	-type dipeptide/oligopeptide/nickel transport systems, permease							
components								
sti-Sthe_1120	sti-Sthe_1512							
COG0607 Rhodan	ese-related sulfurtransferase							
hsw-Hsw_3707	hsw-Hsw_3484							
COG0665 Glycine,	/D-amino acid oxidases (deaminating)							
hsw-Hsw_3707	hsw-Hsw_4126							
COG0714 MoxR-li	ke ATPases							
tmr-Tmar_0571	tmr-Tmar_1527							
COG0715 ABC	C-type nitrate/sulfonate/bicarbonate transport systems, periplasmic							
components								
tmr-Tmar_0571	tmr-Tmar_0484							
COG0730 Predicte	ed permeases							
tmr-Tmar_1286	tmr-Tmar_0809							
COG0747 ABC-type dipeptide transport system, periplasmic component								
lpil-LIP_2986	lpil-LIP_2024							
sth-STH1685	sth-STH2315							
COG0834 ABC-typ	e amino acid transport/signal transduction systems							
tmr-Tmar_0571	tmr-Tmar_1438							
COG1012 NAD-de	pendent aldehyde dehydrogenases							
hsw-Hsw_3707	hsw-Hsw_3846							

COG1135 ABC-type metal ion transport system, ATPase component						
lpil-LIP_2986	lpil-LIP_3402					
COG1414 Transcriptional regulator						
cwo-Cwoe_3191	cwo-Cwoe_5126					
lpil-LIP_3402	lpil-LIP_2767					
tmr-Tmar_0571	tmr-Tmar_1341					
COG1629 Outer membrane receptor proteins, mostly Fe transport						
brl-BZG35_14625	brl-BZG35_06355					
COG1648 Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)						
brl-BZG35_14625 brl-BZG35_07190						
COG1653 ABC-type sugar transport system, periplasmic component						
lpil-LIP_2986	lpil-LIP_2466					
COG1940 Transcriptional regulator/sugar kinase						
lpil-LIP_3402	lpil-LIP_0410					
COG2606 Uncharacterized conserved protein						
sth-STH1685	sth-STH1325					

Tabla S1. Genes regulados por un SAM Riboswitches *in trans* y el COG al que pertenecen. Los COGs y sus funciones mostradas arriba de cada par de genes, corresponden al COG al que pertenece el gen regulado *in trans* (genes del lado derecho) por parte de un SAM riboswitch (genes del lado izquierdo).

Anexos

Anexo 1. Organismos representativos por género de la familia listeriaceas.

Listeria monocytogenes Listeria ivanovii Listeria innocua Listeria seeligeri Listeria welshimeri Listeria weihenstephanensis

Anexo 1. Organismos representativos por género en la familia listerias.

Anexo 2. Organismos representativos por género que pertenezcan al orden Bacillales.

Bacillus megaterium Staphylococcus sciuri Listeria monocytogenes Paenibacillus mucilaginosus Geobacillus Planococcus maritimus Virgibacillus

Anoxybacillus Exiguobacterium Sporosarcina psychrophila Solibacillus silvestris Lysinibacillus fusiformis **Brevibacillus brevis** Macrococcus Tumebacillus Parageobacillus thermoglucosidasius Fictibacillus phosphorivorans Kyrpidia Aneurinibacillus Halobacillus mangrovi Novibacillus thermophilus Salimicrobium jeotgali Aeribacillus pallidus Lentibacillus amyloliquefaciens Auricoccus indicus Terribacillus aidingensis Brochothrix thermosphacta Amphibacillus xylanus Kurthia Oceanobacillus iheyensis Thermobacillus composti Salinicoccus halodurans Gemella Jeotgalibacillus malaysiensis Rummeliibacillus stabekisii Alicyclobacillus acidocaldarius

Anexo2. Organismos representativos por género pertenecientes al orden Bacillales

Bibliografía.

- Aboul-ela, F., Huang, W., Abd Elrahman, M., Boyapati, V., & Li, P. (2015). Linking aptamer-ligand binding and expression platform folding in riboswitches: prospects for mechanistic modeling and design. *Wiley Interdisciplinary Reviews: RNA*, *6*(6), 631–650. https://doi.org/10.1002/wrna.1300
- Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*, 1–38. https://doi.org/10.7554/elife.05005

- Ahmed, W., Hafeez, M. A., & Mahmood, S. (2018). Identification and functional characterization of bacterial small non-coding RNAs and their target: A review. *Gene Reports*, 10(December 2017), 167–176. https://doi.org/10.1016/j.genrep.2018.01.001
- Ames, T. D., & Breaker, R. R. (2011). Bacterial aptamers that selectively bind glutamine. *RNA Biology*, 8(1), 82–89. https://doi.org/10.4161/rna.8.1.13864
- Babitzke, P., & Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology*, *10*(2), 156–163. https://doi.org/10.1016/j.mib.2007.03.007
- Baker, J. L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R. B., & Breaker, R. R. (2012).
 Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*, *335*(6065), 233–235. https://doi.org/10.1126/science.1215063
- Bandyopadhyay, S., Ghosh, D., Mitra, R., & Zhao, Z. (2015). MBSTAR: Multiple instance learning for predicting specific functional binding sites in microRNA targets. *Scientific Reports*, *5*, 1–12. https://doi.org/10.1038/srep08004
- Barrick, J. E., & Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolitebinding riboswitches. *Genome Biology*, *8*(11). https://doi.org/10.1186/gb-2007-8-11-r239
- Carloni, S., Macchi, R., Sattin, S., Ferrara, S., & Bertoni, G. (2017). The small RNA ReaL: a novel regulatory element embedded in the Pseudomonas aeruginosa quorum sensing networks. *Environmental Microbiology*, *19*(10), 4220–4237. https://doi.org/10.1111/1462-2920.13886
- Chen, K., & Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. *Nature Genetics*, *38*(12), 1452–1456. https://doi.org/10.1038/ng1910
- Corbino, K. A., Barrick, J. E., Lim, J., Welz, R., Tucker, B. J., Puskarz, I., ... Breaker, R. R. (2005). Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biology*, *6*(8), R70. https://doi.org/10.1186/gb-2005-6-8-r70
- D. A. Abdala, R. Ciria and E. Merino GeConT 3: gene context analysis for orthologous proteins, conserved domains and metabolic pathways
- Dann, C. E., Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., & Winkler, W. C. (2007). Structure and Mechanism of a Metal-Sensing Regulatory RNA. *Cell*, 130(5), 878–892. https://doi.org/10.1016/j.cell.2007.06.051
- Dutta, T., & Srivastava, S. (2018). Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene*, *656*(February), 60–72. https://doi.org/10.1016/j.gene.2018.02.068
- Edwards, B. A. L., & Batey, R. T. (2010). Riboswitches : A Common RNA Regulatory Element. *Nature Education*, 3(9), 1–9. https://doi.org/http://www.nature.com/scitable/topicpage/riboswitches-a-common-rna-regulatory-element-14262702

- Eric P. Nawrocki & Sean R. Eddy (2013) Computational identification of functional RNA homologs in metagenomic data, RNA Biology, 10:7, 1170-1179, DOI: 10.4161/rna.25038
- Gayan MA, Sherlock ME, Weinberg Z, Breaker RR RNA Biol. 2018; [Epub ahead of print] SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches.
- Gelaye, B., Rondon, M., Araya, P. R., & A, P. M. (2016). Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal, 3(10), 973–982. https://doi.org/10.1016/S2215-0366(16)30284-X.Epidemiology
- Gilbert, W. (1986). The RNA world. Nature. https://doi.org/10.1038/319618a0
- Gong, S., Wang, Y., Wang, Z., Wang, Y., & Zhang, W. (2016). Reversible-Switch Mechanism of the SAM-III Riboswitch. *The Journal of Physical Chemistry B*, *120*(48), 12305–12311. https://doi.org/10.1021/acs.jpcb.6b09698
- Gottesman Susan. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. Trends in Genetics, Volume 21, Issue 7. https://doi.org/10.1016/j.tig.2005.05.008.
- Green, N. J., Grundy, F. J., & Henkin, T. M. (2010). The T box mechanism: tRNA as a regulatory molecule. FEBS letters, 584(2), 318–324. doi:10.1016/j.febslet.2009.11.056
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., & Eddy, S. R. (2003). Rfam: an RNA family database. Nucleic acids research, 31(1), 439–441. doi:10.1093/nar/gkg006
- Grundy, F. J., & Henkin, T. M. (1998). The S box regulon: A new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria.
 Molecular Microbiology, 30(4), 737–749. https://doi.org/10.1046/j.1365-2958.1998.01105.x
- Grundy, F. J., Lehman, S. C., & Henkin, T. M. (2003). The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proceedings of the National Academy of Sciences*, 100(21), 12057–12062. https://doi.org/10.1073/pnas.2133705100
- Gutiérrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C., & Merino, E. (2009). Biochemical features and functional implications of the RNA-based T-box regulatory mechanism.
 Microbiology and molecular biology reviews : MMBR, 73(1), 36–61. doi:10.1128/MMBR.00026-08
- Havill, J. T., Bhatiya, C., Johnson, S. M., Sheets, J. D., & Thompson, J. S. (2014). A new approach for detecting riboswitches in DNA sequences. *Bioinformatics*, 30(21), 3012–3019. https://doi.org/10.1093/bioinformatics/btu479
- Julie A. Maupin-Furlow, J. A. (2018, 5 noviembre). Vitamin B1 (Thiamine) Metabolism and Regulation in Archaea | IntechOpen. Recuperado 1 julio, 2019, de https://www.intechopen.com/books/bgroup-vitamins-current-uses-and-perspectives/vitamin-b1-thiamine-metabolism-and-regulationin-archaea

- Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, & Madden T.L. (2008) "NCBI BLAST: a better web interface" Nucleic Acids Res. 36:W5-W9
- Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., ... Petrov, A. I. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, *46*(D1), D335–D342. https://doi.org/10.1093/nar/gkx1038
- Kaphingst, K. A., Persky, S., & Lachance, C. (2010). Recognition of S-adenosylmethionine by riboswitches, *14*(4), 384–399. https://doi.org/10.1080/10810730902873927.Testing
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, *39*(10), 1278–1284. https://doi.org/10.1038/ng2135
- Kim, J. N., Roth, A., & Breaker, R. R. (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proceedings of the National Academy of Sciences*, 104(41), 16092–16097. https://doi.org/10.1073/pnas.0705884104
- Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., ... Rajewsky, N. (2005).
 Combinatorial microRNA target predictions. *Nature Genetics*, *37*(5), 495–500.
 https://doi.org/10.1038/ng1536
- Liu, S., Li, J. H., Wu, J., Zhou, K. R., Zhou, H., Yang, J. H., & Qu, L. H. (2015). StarScan: A web server for scanning small RNA targets from degradome sequencing data. *Nucleic Acids Research*, 43(W1), W480–W486. https://doi.org/10.1093/nar/gkv524
- Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., ... Johansson, J. (2009). A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. *Cell*, 139(4), 770–779. https://doi.org/10.1016/j.cell.2009.08.046
- Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., Cossart, P., Bacteries-cellules, I., & Johansson, J. (2002). An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. *Cell*, 110(5), 551–561. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12230973
- Mandal, M. (2004). A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science*, *306*(5694), 275–279. https://doi.org/10.1126/science.1100829
- Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., & Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, *113*(5), 577–586. https://doi.org/10.1016/S0092-8674(03)00391-X
- Mandal, M., & Breaker, R. R. (2004a). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural and Molecular Biology*, *11*(1), 29–35. https://doi.org/10.1038/nsmb710

- Mandal, M., & Breaker, R. R. (2004b). Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology*, 5(6), 451–463. https://doi.org/10.1038/nrm1403
- Massé, E., Vanderpool, C. K., & Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in Escherichia coli. Journal of bacteriology, 187(20), 6962–6971. doi:10.1128/JB.187.20.6962-6971.2005
- Martinez-Guerrero, C. E., Ciria, R., Abreu-Goodger, C., Moreno-Hagelsieb, G., & Merino, E. (2008). GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. Nucleic acids research, 36(Web Server issue), W176–W180. doi:10.1093/nar/gkn330
- Mattick, J. S. (2004). RNA regulation: a new genetics? *Nature Reviews. Genetics*, 5(4), 316–323. https://doi.org/10.1038/nrg1321
- Meyer, S., Carlson, P. D., & Lucks, J. B. (2017). Characterizing the Structure-Function Relationship of a Naturally Occurring RNA Thermometer. *Biochemistry*, 56(51), 6629–6638. https://doi.org/10.1021/acs.biochem.7b01170
- Michaux, C., Verneuil, N., Hartke, A., & Giard, J.-C. (2014). Physiological roles of small RNA molecules. *Microbiology*, *160*(Pt_6), 1007–1019. https://doi.org/10.1099/mic.0.076208-0
- Min, H., & Yoon, S. (2010). Got target?: Computational methods for microRNA target prediction and their extension. *Experimental and Molecular Medicine*, 42(4), 233–244. https://doi.org/10.3858/emm.2010.42.4.032
- Miranda-Rios, J., Navarro, M., & Soberon, M. (2001). A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proceedings of the National Academy of Sciences*, *98*(17), 9736–9741. https://doi.org/10.1073/pnas.161168098
- Mirihana Arachchilage, G., Sherlock, M. E., Weinberg, Z., & Breaker, R. R. (). SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches. RNA biology, 15(3), 371–378. doi:10.1080/15476286.2017.1399232
- Montange, R. K., & Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097), 1172–1175. https://doi.org/10.1038/nature04819
- Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., & Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. *Chemistry and Biology*, *9*(9), 1043–1049. https://doi.org/10.1016/S1074-5521(02)00224-7
- Narberhaus, F., Waldminghaus, T., & Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiology Reviews*, *30*(1), 3–16. https://doi.org/10.1111/j.1574-6976.2005.004.x
- Nawrocki E.P. and S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches, Bioinformatics 29:2933-2935 (2013)
- Ontiveros-Palacios N, et al. Molecular basis of gene regulation by the THI-box riboswitch. Mol Microbiol. 2008;67(4):793–803.

- Orell, A., Tripp, V., Aliaga-Tobar, V., Albers, S.-V., Maracaja-Coutinho, V., & Randau, L. (2018). A regulatory RNA is involved in RNA duplex formation and biofilm regulation in Sulfolobus acidocaldarius. *Nucleic Acids Research*, *46*(9), 4794–4806. https://doi.org/10.1093/nar/gky144
- Oxender, D. L., Zurawskit, G., & Yanofsky, C. (1979). Attenuation in the Escherichia coli tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(11), 5524–5528. https://doi.org/10.1073/pnas.76.11.5524
- Poiata, E., Meyer, M. M., Ames, T. D., & Breaker, R. R. (2009). A variant riboswitch aptamer class for Sadenosylmethionine common in marine bacteria. *Rna*, 15(11), 2046–2056. https://doi.org/10.1261/rna.1824209
- Qu, Z., & Adelson, D. L. (2012). Evolutionary conservation and functional roles of ncRNA. *Frontiers in Genetics*, *3*(OCT), 1–11. https://doi.org/10.3389/fgene.2012.00205
- Regulski, E. E., Moy, R. H., Weinberg, Z., Barrick, J. E., Yao, Z., Ruzzo, W. L., & Breaker, R. R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Molecular Microbiology*, 68(4), 918–932. https://doi.org/10.1111/j.1365-2958.2008.06208.x
- Rehmsmeier, M., Steffen, P., Höchsmann, M., Giegerich, R., & Ho, M. (2004). Fast and effective prediction of microRNA / target duplexes. *Spring*, (2003), 1507–1517. https://doi.org/10.1261/rna.5248604.and
- Richards, G. R., & Vanderpool, C. K. (2011). Molecular call and response: The physiology of bacterial small RNAs. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1809(10), 525–531. https://doi.org/10.1016/j.bbagrm.2011.07.013
- Ritzema (ed.), H.P. (1994). Frequency and Regression Analysis. Chapter 6 in: Drainage Principles and Applications, Publication 16, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. pp. 175-224. ISBN 90-70754-33-9.
- Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W. K., Lim, J., Jona, I., ... Breaker, R. R. (2007). A riboswitch selective for the queuosine precursor preQ1contains an unusually small aptamer domain. *Nature Structural and Molecular Biology*, 14(4), 308–317. https://doi.org/10.1038/nsmb1224
- Saberi, F., Kamali, M., Najafi, A., Yazdanparast, A., & Moghaddam, M. M. (2016). Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications. *Cellular & Molecular Biology Letters*, *21*(1), 6. https://doi.org/10.1186/s11658-016-0007-z
- Salvatore F, Izzo P, Colonna A, Trabón C, Cimino F. Transfer RNA methyltransferases: properties and role in the maturation of tRNA. In Transmethylation, E Usdin, RT Borchardt, CR Creveling eds. North-Holland, New York, Amsterdam and Oxford: Elsevier;1979. p.449-56.

- Varilly, P., & Chandler, D. (2012). A Eubacterial Riboswitch Class that Senses the Coenzyme Tetrahydrofolate, v(2), 265–275. https://doi.org/10.1007/s10955-011-0269-9.Quantifying
- Winkler, W. C., Cohen-Chalamish, S., & Breaker, R. R. (2002). An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences*, 99(25), 15908– 15913. https://doi.org/10.1073/pnas.212628899
- Yang, G., Sau, C., Lai, W., Cichon, J., & Li, W. (2015). B12 cofactors directly stabilize an mRNA regulatory switch, *344*(6188), 1173–1178. https://doi.org/10.1126/science.1249098.Sleep

Yanofsky, C. (1960). The tryptophan synthetase system. *Bacteriological Reviews*, *24*(2), 221–245. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgiartid=440991&tool=pmcentrez&rendertype=ab stract %5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/13846454%5Cnhttp://www.pubmedcentral.nih.go v/articlerender.fcgi?artid=PMC440991

Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6), 402–415. https://doi.org/10.2174/138920209789177575

Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature. 2002;419(6910):952–956.