UNIVERSIDAD DE TALCA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN BIOINFORMÁTICA

ESTUDIO DE LA RELACIÓN ENTRE EL USO CODOGÉNICO Y LA REGULACIÓN POR sRNA EN *E. COLI* K-12 MG1655

FELIPE ESTEBAN ÁVALOS DÍAZ

Profesor Tutor: MARIO TELLO REYES Profesor Co-Tutor:OMAR ORELLANA

Proyecto de Memoria de Título de
Ingeniero en Bioinformática

Talca – Chile
Marzo, 2018

UNIVERSIDAD DE TALCA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN BIOINFORMÁTICA

ESTUDIO DE LA RELACIÓN ENTRE EL USO CODOGÉNICO Y LA REGULACIÓN POR sRNA EN *E. COLI* K-12 MG1655

FELIPE ESTEBAN ÁVALOS DÍAZ

Profesor Tutor: MARIO TELLO REYES

Profesor Co-Tutor: OMAR ORELLANA

Profesor Informante: GONZALO RIADI MAHIAS

Proyecto de Memoria de Título de
Ingeniero en Bioinformática

Talca – Chile
Marzo, 2018

INDICE DE CONTENIDOS

página

Indice de Contenidos I

Indice de Figuras IV

Indice de Tablas VI

Resumen VIII

Abstract X

1. MARCO TEÓRICO 1

1.1. Introducción . . . 1

1.1.1. Presentación del problema . . . 1

1.2. Marco teórico . . . 2

1.2.1. RNAs pequeños . . . 2

1.2.2. RNA mensajero (mRNA) . . . 3

1.2.3. Redes de Interacción . . . 6

1.2.4. Cytoscape . . . 6

1.2.5. Justificación del Problema . . . 9

2. HIPÓTESIS 10

2.1. Hipótesis . . . 10

3. OBJETIVOS 11

3.1. Objetivo . . . 11

3.2. Objetivos específicos . . . 11

4. MATERIALES Y MÉTODOS 12

4.1. Materiales . . . 12

4.1.1. Software . . . 12

4.1.2. Bases de Datos . . . 13

4.1.3. Hardware . . . 13

4.1.4. Scripts . . . 13

4.2. Metodología . . . 13

4.2.1. Identificación mediante búsqueda bibliográfica de las parejas de sRNAs y mRNAs que interaccionan entre sí en *Escherichia coli* . . 13
4.2.2. Predicción e identificación de las regiones de interacción entre mRNA y sRNA 14
4.2.3. Caracterización de la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción 15
4.2.4. Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo 17

5. Resultados **18**
5.1. OE1: Identificación mediante búsqueda en las bases de datos de las parejas de sRNAs y mRNA en *Escherichia coli* K-12 MG1655. 18
5.2. OE2: Identificación y predicción de las regiones de interacción entre mRNA y sRNA. 23
5.2.1. Predicción de los blancos de los sRNAs sin interacción identificada. 23
5.2.2. Identificación de las regiones de interacción sRNA-mRNA blanco. . 24
5.2.3. 5.2.3. Funciones biológicas reguladas por sRNA. 27
5.2.4. 5.2.4. Localización de sitios de interacción sRNA-mRNA en los sRNAs 29
5.3. OE3: Caracterización de la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción. 38
5.3.1. Análisis estadístico de los CAI y tAI de los mensajeros regulados por sRNAs. 38
5.3.2. Regiones conservadas en mRNAs reconocidos por el mismo sRNA. . 40
5.4. OE4: Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo y de la región de interacción. 46
5.4.1. Funciones Biológicas de los genes que interactúan con los módulos de los sRNAs. 54

6. Discusión **59**
6.1. Discusión 59

7. Conclusión **63**
7.1. Conclusión 63

Bibliografía **64**

Anexos

A: Resumen Congreso Somich Noviembre 2016 **69**

B: Poster Presentado Somich 2016 **71**

C: Gráficos, Estructuras y Representación sRNAs **72**

C.0.1. OxyS 72
C.0.2. MicA 74
C.0.3. RybB 76
C.0.4. FnrS 78
C.0.5. DsrA 80
C.0.6. SgrS 82
C.0.7. MicC 84
C.0.8. MicF 86
C.0.9. RprA 88
C.0.10. CyaR 90
C.0.11. RygD 92

D: Modelo Base de Datos **94**

INDICE DE FIGURAS

página

1.1. Ejemplos de sRNA en *Escherichia coli* y *Salmonella typhimurium* 2
1.2. Flujo de la información genética desde el DNA a proteínas 4
1.3. Esquema de la arquitectura y núcleo de Cytoscape 7
1.4. Representaciones de Cytoscape 3 8
1.5. Ventana de Cytoscape 9

5.1. Resumen de datos BSRD y SRNATarBase 18
5.2. Distribución de Interacciones no caracterizadas versus Interacciones Caracterizadas 19
5.3. Porcentaje de interacciones predichas y empíricas en *E. coli.* 25
5.4. Funciones biológicas de los mRNAs regulados por sRNA 28
5.5. Funciones biológicas de los mRNAs regulados por sRNA e identificados por métodos bioinformáticos 29
5.6. Análisis matricial de OmrA. 32
5.7. Análisis matricial de RyhB. 33
5.8. Análisis matricial de OmrB. 35
5.9. Análisis matricial de GcvB. 36
5.10. Análisis matricial de Spot42. 37
5.11. Modelo de red de interacción sRNA-mRNA en *E. coli* 48
5.12. Modelo de red de interacción de OmrA 49
5.13. Modelo red Interacción RyhB 50
5.14. Modelo de red de interacción OmrB 51
5.15. Modelo de red de interacción GcvB 52
5.16. Modelo de red de interacción Spot42 53
5.17. Procesos biológicos asociados a los genes regulados por sRNA 55
5.18. Funciones biológicas asociadas a los módulos. 57
5.19. Sesgo en el uso de codones en los módulos de los sRNA 58

B.1. Poster Somich 2016 71

C.1. Gráfico OxyS 72
C.2. Estructiura Secundaria OxyS 72
C.3. Red Interacción OxyS 73
C.4. Gráfico MicA 74

C.5. Estructura secundaria MicA 74
C.6. Red de Interacción MicA 75
C.7. Gráfico RybB 76
C.8. Estructura secundaria RybB 76
C.9. Red de Interacción RybB 77
C.10. Gráfico FnrS 78
C.11. Estructura secundaria FnrS 78
C.12. Red de Interacción FnrS 79
C.13. Gráfico DsrA 80
C.14. Estructura secundaria DsrA 80
C.15. Red de Interacción DsrA 81
C.16. Gráfico SgrS 82
C.17. Estructura secundaria SgrS 82
C.18. Red de Interacción SgrS 83
C.19. Gráfico MicC 84
C.20. Estructura secundaria MicC 84
C.21. Red de Interacción MicC 85
C.22. Gráfico MicF 86
C.23. Estructura secundaria MicF 86
C.24. Red de Interacción MicF 87
C.25. Gráfico RprA 88
C.26. Estructura secundaria RprA 88
C.27. Red de Interacción RprA 89
C.28. Gráfico CyaR 90
C.29. Estructura secundaria CyaR 90
C.30. Red de Interacción CyaR 91
C.31. Gráfico RygD 92
C.32. Estructura secundaria RygD 92
C.33. Red de Interacción RygD 93

D.1. Modelo Base de Datos 94

INDICE DE TABLAS

página

5.1. sRNAs y sus blancos presentes en la bases de datos BSRD (Bacterial Small Regulatory RNA Database) o en la base de datos sRNATarBase. 23
5.2. La tabla muestra los resultados de predicción de los blancos de los sRNAs huérfanos, junto con la predicción de su energía de interacción sRNA-mRNA. 24
5.3. Distribución de las interacciones por zonas del mensajero 26
5.4. Densidad de interacción en las diferentes regiones del mRNA. 27
5.5. Organización de las interacciones en los sRNAs con más de 4 mRNA blancos. 30
5.6. Resumen de la presencia de horquilla GC estables donde no ocurre interaccion con mensajeros en las estructuras secundarias de los sRNAs que interaccionan con 4 o más mensajeros, con la finalidad de identificar si en la estructura secundarias de los sRNAs se forma una zona en la cual no ocurre interacción sRNA-mRNA 37
5.7. Tabla comparativa CAI y tAI de los mensajeros que interaccionan con sRNA v/s los mensajeros que no interaccionan con sRNA. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney no pareado. El asterisco sugiere significancia estadística 39
5.8. Comparación de los valores de CAI y tAI entre la región de interacción y los valores presentes en las regiones codificantes. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney no pareado. El asterisco sugiere significancia estadística. 39
5.9. Comparación de los valores de CAI y tAI entre la región de interacción y los valores presentes en las regiones codificantes. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney pareado. El asterisco sugiere significancia estadística. 40
5.10. Secuencias consenso para OmrA, junto con las zonas que fueron localizadas en éste. 41
5.11. Secuencias consenso para RyhB, junto con las zonas que fueron identificadas para este sRNA. 43
5.12. Secuencias consenso para OmrB, junto con las zonas que fueron identificadas para este sRNA. 44
5.13. Secuencias consenso para GcvB, junto con las zonas que fueron identificadas para este sRNA 45

5.14. Secuencias consenso para Spot42, junto con las zonas que fueron identificadas para este sRNA 46

RESUMEN

La traducción de la información genética contenida en los RNA mensajeros (mRNA) es un proceso esencial para la célula, el cual se encuentra altamente regulado tanto en eucariontes como en procariontes. En bacterias, la tasa de traducción del mRNA depende de la disponibilidad del repertorio de aminoacil-tRNAs que reconocen a sus codones, de la estructura que adquiere y de la interacción con otras moléculas como proteínas o RNAs. Los RNA pequeños (sRNA o small RNA) son un grupo de RNA no codificantes, de tamaño pequeño (50 a 400 nucleótidos), que modifican la tasa de traducción de un mRNA mediante interacción con la región codificante o con los extremos UTR. Las razones por la cual sólo algunos genes son regulados por sRNAs no son claras, sin embargo, que algunas interacciones sRNA-mRNA ocurran en regiones codificantes, hace factible pensar que la regulación por sRNA y la adaptación a la disponibilidad del repertorio de tRNAs están estrechamente relacionadas, pues ambas dependen de los codones presentes en el mRNA. En esta tesis proponemos estudiar las interacciones sRNA-mRNA presentes en *E. coli* K12 para identificar si los genes regulados por sRNA presentan en su región codificante o en la región de interacción con el sRNA un conjunto de codones con una particular disponibilidad de tRNA. Además, se propone estudiar como esta regulación se integra con el metabolismo de *E. coli* formando una red de regulación. Para desarrollar esta tesis se propone como Hipótesis: En *Escherichia coli* K12 existe una red de regulación mediada por sRNA que integra los diferentes procesos metabólicos, en la cual los genes regulados por sRNA presentan un patrón de codones y adaptación al pool de tRNA diferente al presente en los genes no regulados por sRNAs. Para desarrollar esta hipótesis se propone como objetivos específicos 1.-Identificar mediante búsqueda en las bases de datos las parejas de sRNAs y mRNA en *Escherichia coli* K-12 MG1655. 2.- Identificar y predecir las regiones de interacción entre mRNA y sRNA. 3.- Caracterizar la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción. 4.- Generar un mapa de interacciones sRNA-mRNA que incluya la red metabólica a la cual pertenece el mRNA, y la adaptación al uso codogénico y repertorio de tRNA, tanto del gen completo como de la región de interacción. Se Identificó para *Escherichia coli* K12 varias redes de interacción que integró la regulación mediada por sRNA tanto en interacciones sRNA-mRNA determinadas empíricamente como establecidas por métodos predictivos y la eficiencia con la cual los genes son traducidos. Conclusión: Los resultados obtenidos en esta memoria de título permiten sugerir que en *E. coli*, el mecanismo de regulación por sRNA, ocurre en zonas del área codificante que contienen codones de alto uso y baja

disponibilidad de tRNA, lo que implica que la regulación ocurre en regiones en donde la velocidad de avance del ribosoma es menor.

ABSTRACT

The translation of the genetic information contained in a messenger RNA (mRNA) is an essential process for the cell, which is highly regulated, both in eukaryotes and prokaryoic cells. In bacteria, the rate of translation of mRNA depends on the availability of the repertoire of aminoacyl - tRNAs that recognize each of their codons, the mRNA structure and interaction with other molecules such as proteins or RNAs. Small RNAs (sRNA or small RNA) are a group of small (50 to 400 nucleotides), noncoding RNAs, which modify the rate of translation of mRNA through interaction with the coding region or with the UTR ends. The reasons why some genes are regulated by sRNAs are not clear, however, that some interactions sRNA-mRNA occur in coding regions, makes it feasible to propose that regulation by sRNA and the properties which depends of codon composition such the adaptation to the tRNA pool are closely related. This thesis proposes to study in *E. coli* K12 the sRNA-mRNA interactions present to identify whether genes regulated by sRNAs present in their coding region or in the region of interaction with the sRNA a set of codons with particular tRNA availability. Furthermore, we propose to study how this control is integrated with the metabolism of *E. coli* forming a regulatory network. To develop this thesis we suggest as hypothesis: "In *Escherichia coli* K12 there is a network of regulation mediated sRNA that integrates the various metabolic processes, in which the regulated sRNA genes exhibit a pattern of codons and adaptation pool of different tRNA to present in genes not regulated by sRNAs". To test this hypothesis we propose specific objectives 1. To identify by searching databases couples of sRNAs and mRNA in *Escherichia coli* K-12 MG1655. 2.- To identify and to predict the interaction regions between mRNA and sRNA. 3.- To characterize the adaptation to the tRNA repertoire of the genes regulated by sRNA and its interaction regions. 4.- To generate a map of sRNA-mRNA interactions that includes the metabolic network to which the mRNA belongs, and the adaptation to codogenic use and repertoire of tRNA, both of the complete gene and of the interaction region.

We identified in *Escherichia coli* K12 several sRNA-mRNA networks using interactions determined by empirically and predictive methods. This networks also includes the efficiency with which the genes are translated. Our results suggest that in Escherichia coli K12, the regulatory mechanism by small RNA, occurs in regions with codons highly represented by low availability of tRNA. This implies that the regulation by sRNA occurs in regions where the rate and advancement of the ribosome is lower.

1. MARCO TEÓRICO

1.1. Introducción

1.1.1. Presentación del problema

La traducción del mensaje genético es un complejo proceso biológico encargado de transformar la información genética contenida en el mRNA en proteína, el cual depende directamente de la disponibilidad de los tRNA aminoacilados, la cantidad de ribosomas disponibles y el nivel de expresión de los mRNA. Durante este proceso, la traducción de los mRNAs puede ser regulada por RNAs pequeños no codificantes (small RNA o sRNA), modificando su eficiencia traduccional. Se define eficiencia de traducción como la cantidad de proteínas producidas por mRNA en un periodo de tiempo determinado[21]. Actualmente se han descrito sRNAs en 783 especies bacterianas diferentes[15]. En *Escherichia coli* cepa K-12 MG1655 se han identificado 108 sRNAs[15] los cuales regulan diversos procesos metabólicos[25]. A pesar de que el mecanismo por el cual los sRNA regulan la expresión génica es bien conocido, no es claro por qué algunos mRNA interactúan con los sRNA y otros no. Muchos sRNA interactúan en regiones codificantes de los mRNA por lo cual, propiedades que dependen de la composición nucleotídica de las regiones codificantes como el contenido GC o el sesgo en el uso de codones podrían estar relacionados con la regulación por sRNAs. Como el sesgo en el uso de codones también está relacionado con la adaptación al pool de tRNAs disponibles en la célula, es factible suponer que ambos tipos de regulación de la eficiencia traduccional pudiesen estar relacionados entre sí. En esta memoria de título se propuso integrar la información relacionada a la regulación de la traducción de los sRNA sobre los mensajeros en *Escherichia coli* K-12 MG1655 por medio de la identificación y visualización de la red de interacción sRNA-mRNA, junto con evaluaciones estadísticas de la adaptación de codones al pool de los tRNAs presentes en la célula. De esta manera se logró observar que en *E. coli* K-12 MG1655, los genes regulados por sRNA, no existe una preferencia de uso codogénico diferente a los genes no regulados por sRNA. Sin embargo, en los sitios de

interacción sRNA-mRNA que se encuentran en la región codificante del mRNA, se observó un uso preferencial de codones de alta frecuencia, pero con baja disponibilidad de tRNAs.

1.2. Marco teórico

1.2.1. RNAs pequeños

Los RNA pequeños (small RNA o sRNA), son un tipo de ARN que no codifica para proteínas (ncRNA), altamente estructurados, y de un tamaño que varía entre 40 a 400 nucleótidos[13]. Los sRNA son producidos por bacterias, entre las que encontramos, especies de *Salmonella typhimurium*[36], *Escherichia coli*[37], *Pseudomonas aeruginosa*[17], *Pseudomonas fluorescens*[30], *Staphylococcus aureus*[6], entre otras. En *Escherichia coli* K-12 MG1655 se han identificado 108 sRNAs de los cuales 32 forman interacciones con mRNA que han sido caracterizadas por métodos empíricos[15]. En contraste, existen otros 76 sRNAs en los cuales se desconoce sus blancos de interacción o la forma en la cual interactuan con los genes que regulan.

Los sRNA regulan la traducción de ciertos genes, marcan algunas proteínas para su degradación y modulan la actividad de la RNA polimerasa[39]. De acuerdo al mecanismo por el cual regulan la expresión génica, los sRNA se pueden dividir en dos grupos: los que se alinean con hebras de ARN (Unión completaría) con la finalidad de afectar la estabilidad del mRNA y su traducción, y los que se unen a proteínas afectando su función[37].

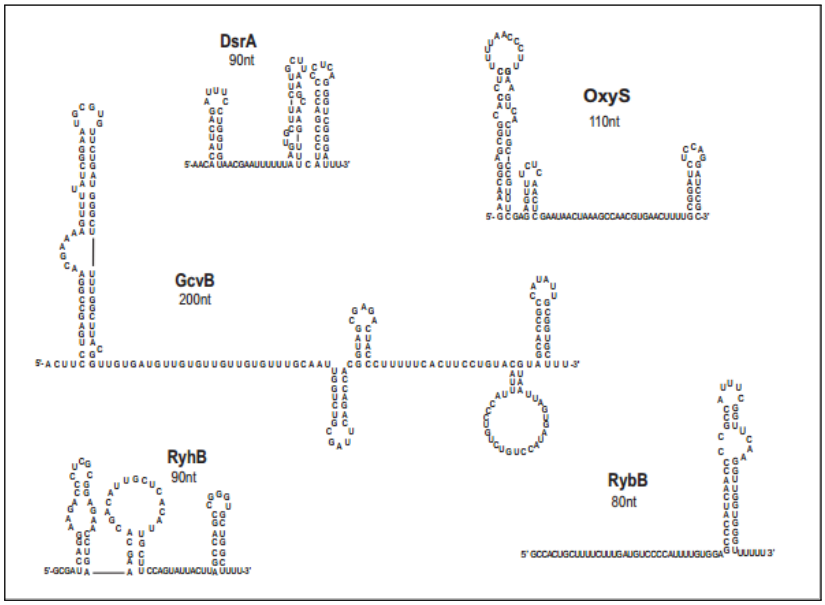


Figura 1.1: En la figura se muestra la estructura secundaria de los sRNA DsrA[14], OxyS[1], GcvB[27], RyhB[9], y RybB[3]. Las estructuras secundarias que se visualizan en la imagen están basadas en experimentos *in vitro*

Los sRNA bacterianos además de poseer diferentes tamaños, pueden adoptar diversas estructuras secundarias de tipo tallo-bucle, las cuales están separadas por regiones de hebra simple extendida (Figura 1.1).

Los sRNAs actúan como reguladores de la expresión génica en respuesta a condiciones de estrés como: los cambios de temperatura[28], bajas concentraciones de hierro[18], estrés oxidativo[1], variaciones en las concentración de glucosa[19, 23] y altos niveles de glucosa-6-fosfato [20].

La presencia de diversos sRNAs regulando varias rutas metabólicas sugiere la existencia de una red de interacción que permite la regulación coordinada de varios procesos metabólicos. En el año 2011 Gisela Stortz y cols.[31], identificaron por primera vez en *Escherichia coli* una red de interacción sRNA-mRNA utilizando los 20 sRNA descritos hasta esa fecha. En esta red se observan ocho grupos en los cuales se puede apreciar que el sRNA MicF coordina tanto el metabolismo de aminoácidos, como el trasporte de membrana y el metabolismo del hierro. El descubrimiento de nuevos sRNA y la caracterización de sus blanco, además de sus efectos sobre su expresión ha llevado a determinar que los sRNA establecen circuitos de regulación que, frente a cambios en las condiciones ambientales, permiten cambios más rápidos y estables en la expresión génica que la regulación mediada por factores transcripcionales [20]

Actualmente, de los 108 sRNAs determinados en *Escherichia coli*, 32 sRNAs tienen determinado mediante métodos empíricos, tanto el blanco mRNA blanco como el sitio de interacción. En 64 sRNAs se ha determinado el mRNA blanco, sin embargo se desconoce el sitio de interacción, mientras que en 12 sRNA no se conoce su blanco de interacción. Los 96 sRNAs con blancos conocidos forman un total de al menos 412 interacciones sRNA-mRNA.

1.2.2. RNA mensajero (mRNA)

1.2.2.1 mRNA en el proceso de traducción y la adaptación al uso de codones y al pool de tRNA.

El RNA mensajero (mRNA) es un tipo de ARN, que tiene por finalidad transportar la información genética desde el ADN a los sitios donde se sintetizan proteínas a partir de esta información (ribosomas) (Figura 1.2).

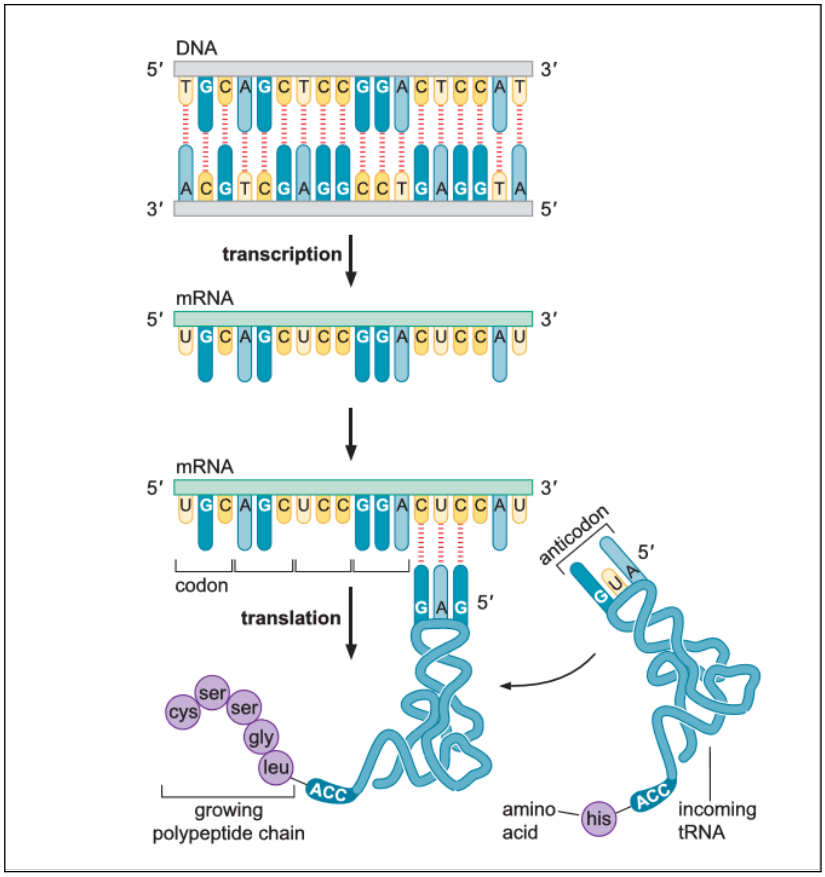


Figura 1.2: ***Flujo de la información genética desde el ADN a proteínas***. La figura esquematiza los pasos de síntesis de mRNA en el proceso de transcripción, y posterior proceso de traducción. Los tRNA actúan como moléculas adaptadoras que mediante complementariedad con los codones presentes en los mRNA van descodificando la información genética permitiendo la síntesis de proteínas en los ribosomas.

Del total de ARN celular solo un pequeño porcentaje (entre el 15 % y 20 %) es ARN mensajero, el cual puede ser leído por varios ribosomas simultáneamente. En el proceso de traducción, el mRNA es decodificado en los ribosomas por los tRNAs. Cada tRNA reconoce mediante complementariedad de bases un triplete de nucleótidos denominado codón, a través de su anti-codón. Existen 64 codones distintos, de los cuales 61 codifican para aminoácidos y tres corresponden a señales que permiten detener el proceso de traducción[7]. El código genético es conservado, sin embargo cada aminoácido puede estar codificado por uno o varios codones.

Los codones que codifican para el mismo aminoácido se denominan sinónimos y su distribución no es aleatoria, existiendo codones que se utilizan más frecuentemente que otros. Este sesgo es propio de cada organismo, y de cada gen[38]. Uno de los factores del cual depende la traducción de proteínas es la disponibilidad o abundancia de los tRNA

cargados con los respectivos aminoácidos (aminoacil-tRNA) y que tienen la función de decodificar los codones presentes en el mRNA. En muchos organismos los codones más frecuentemente utilizados también son aquellos decodificados por los tRNAs que presentan un mayor nivel de expresión o un mayor número de copias génicas[38]. Los genes que más se expresan también tienden a utilizar el conjunto de codones más abundantemente representados en el genoma y con mayor disponibilidad de tRNAs, de tal manera de hacer más eficiente su proceso de traducción[38]. El grado de adaptación de un gen al conjunto de codones más utilizado en un organismos puede ser determinada mediante el uso de indicadores, como el índice de adaptación codogénica o CAI por sus siglas en inglés[11]. Los valores de CAI de un gen pueden ser calculados a partir de su secuencia codificante y la frecuencia de uso de codones que existe en cada organismo. El uso codogénico para *Escherichia coli* y varios organismos se encuentran tabulados y son de libre disposición, así como los algoritmos que sirven para el cálculo de CAI. Por otra parte, la adaptación al pool de tRNA puede ser evaluada mediante el cálculo del índice de adaptación al repertorio de tRNA (tRNA Adaptation Index o tAI)[8]. Los valores de tAI también pueden ser determinados a partir de la secuencia codificante, pero necesitan una estimación de la expresión de los tRNAs presentes en el organismo. Esta estimación puede ser inferida a partir del número de copias de genes de tRNA[33] o a partir de la cuantificación de la expresión de los tRNA por RNAseq u otra técnica. Para *Escherichia coli* se cuenta con ambos tipos de información[10].

1.2.2.2 Relación entre sRNAs, adaptación al uso codogénico y repertorio de tRNA

Los sRNAs interactúan con el mRNA tanto en regiones codificantes como también en los extremos 5' UTR o 3' UTR. Hasta el momento se desconoce la relación existente entre la interacción sRNA-mRNA con la adaptación al uso de codones o al repertorio de tRNA presente en un organismo. A pesar de esto, existen varias razones que permiten suponer que tal relación existe:

1. En las regiones codificantes si la interacción del sRNA con el mRNA dependiera también de los codones presentes, mutaciones sinónimas podrían alterar la regulación por sRNAs (modificando su unión al mRNA) y la adaptación al repertorio de tRNA.

2. Una mayor adaptación al repertorio de tRNAs, permite una traducción más eficiente, por lo tanto, se requiere una menor cantidad de mRNA para producir una mayor cantidad de proteína. Dado que los genes con alta adaptación al repertorio de tRNAs corresponden a genes constitutivos, los cuales no son regulados en su expresión, es

factible pensar que los genes regulados por sRNA tengan valores de adaptación al repertorio de tRNA en un rango intermedio, o muy bajo

Las identificaciones de estas relaciones se pueden apreciar mediante la construcción de redes de interacción que combinen tanto la información de interacción sRNA-mRNA, adaptación al uso codogénico, y adaptación al repertorio de tRNAs. Si, además estas redes de interacción también incluyeran la información sobre la red metabólica que es regulada, permitiría apreciar de mejor manera la interacción entre sRNAs y el uso de codones presentes en un gen.

1.2.3. Redes de Interacción

Las redes de interacción representan relaciones entre 2 objetos de un sistema, comúnmente se utiliza para estudiar sistemas complejos en muchas disciplinas, como por ejemplo las ciencias de la computación, ciencias sociales y la biología. En biología los nodos pueden representar diferentes tipos de relaciones, como por ejemplo interacciones físicas entre las proteínas, genes o metabolitos o regulación de la expresión génica[32].

Las redes de interacción pueden ser generadas a partir de un conocimiento previo, como también por medio de la deducción a partir de los datos experimentales. Existen muchos repositorios en línea que contienen información para la identificación de redes de interacción, por ejemplo, KEGG (**Kyoto Encyclopedia of Genes and Genomes**), BIND (**Biomolecular Interaction Network Database**) [2], y HPRD (**Human Protein Reference Database**) [24], en las cuales están almacenas una gran cantidad de información, la cual puede ser integrada y representada como una red para posteriormente ser analizada. Para poder identificar y visualizar estas redes de interacción existen distintos softwares por ejemplo cytoscape[22] el cual tiene la capacidad de crear redes de interacción a partir de datos provistos por el usuario, los cuales pueden proceder de una base de datos como también datos ingresados manualmente por el usuario.

1.2.4. Cytoscape

Cytoscape es el programa de código abierto más popular que es utilizado para la exploración visual de las redes biológicas que pueden ser formadas por proteínas genes u otros tipos de interacciones. Escrito en su mayoría en el lenguaje de programación java, la primera versión apareció en 2003[22] y ofreció a los investigadores una interfaz de visualización que permite cargar, visualizar, buscar, filtrar y guardar redes, siendo una herramienta versátil e interactiva, útil para explorar interconexiones biológicas

complejas con el apoyo de diversas aplicaciones[32]. Esto facilita la realización de tareas de investigación, tales como la predicción de función de genes.

1.2.4.1 Arquitectura y funcionalidad del núcleo

El eje central del software Cytoscape son los gráficos de redes, en donde las especies moleculares están representadas como nodos y las interacciones moleculares se representan como enlaces, los cuales también son conocidos como las aristas entre los nodos. La funcionalidad básica que proporciona el núcleo del programa Cytoscape consta de la integración de datos arbitrarios en la gráfica, en la representación visual del gráfico y la interfaz de métodos externos al software que son implementados como paquetes de expansión[26].

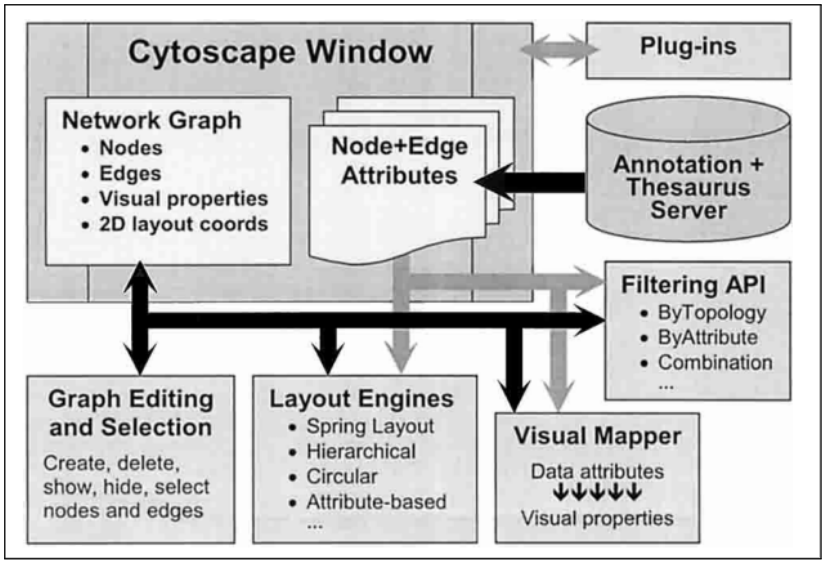


Figura 1.3: Esquema de la arquitectura y núcleo de Cytoscape

El eje principal de Cytoscape es la llamada ventana de Cytoscape (Cytoscape Windows), en la cual están contenidos los datos y sus atributos para representar los gráficos de red, además de operar los métodos básicos, como por ejemplo: la edición gráfica, el diseño gráfico, el mapeo de atributos visuales y el filtrado[26]. Cuando se requiere hacer una interpretación que no está soportada por Cytoscape se utilizan los paquetes de expansión, los cuales mediante una implementación de algoritmos mejoran la funcionalidad del programa. Estos algoritmos que expanden la funcionalidad del programa son descargables desde Cytoscape App Store (http://apps.cytoscape.org/)[26]. Cytoscape además nos da la posibilidad de almacenar todas las anotaciones, provenientes de un servidor independiente, las cuales son almacenadas en el módulo llamado Anotaciones

(Annotation + Thesaurus Server) (Figura 1.3).

Además de los módulos mencionados anteriormente, también existen otros más en los que encontramos por ejemplo el módulo llamado Maquinaria Gráfica (Layout Engines), el cual permite visualizar las redes en una disposición jerárquica (Figura 1.4.A), de resorte (Figura 1.4.C), o circular (Figura 1.4.B)[26]

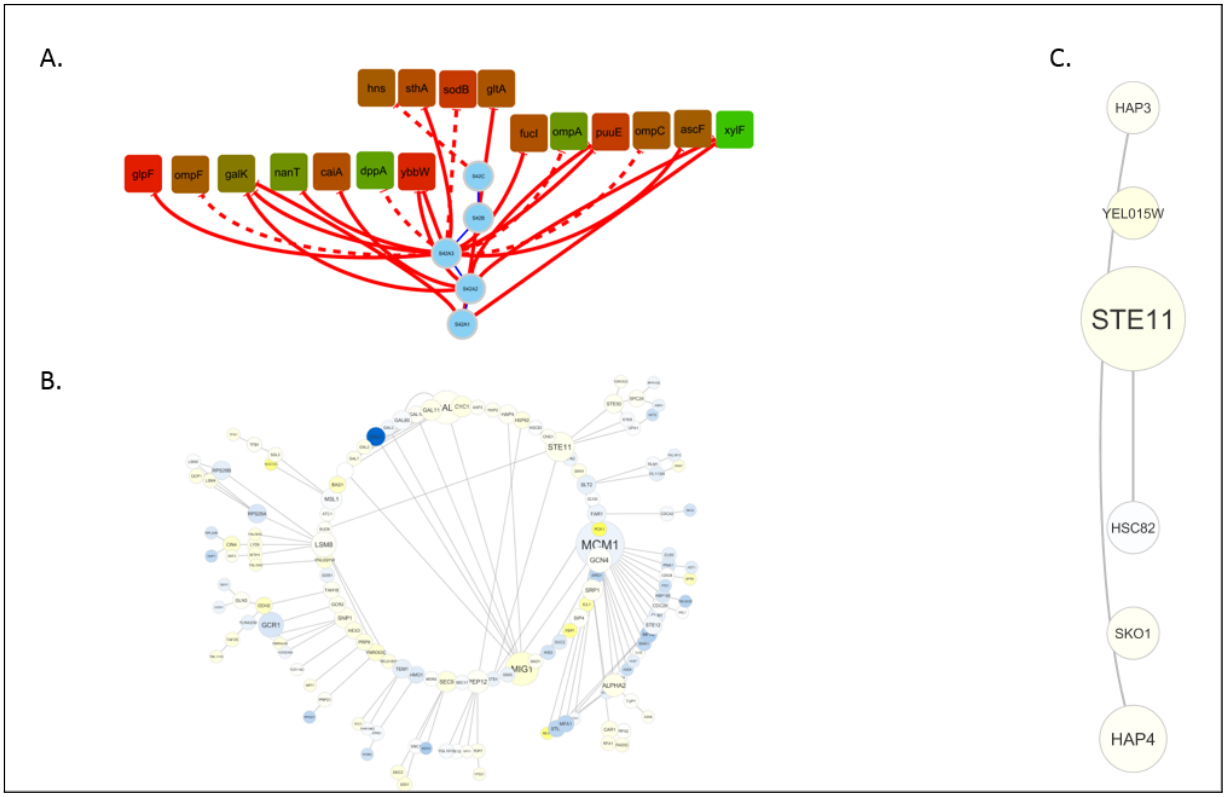


Figura 1.4: **Representaciones de Cytoscape 3**. La figura muestra la representación jerárquica (A) en donde hay un nodo que actúa como raíz y los nodos siguientes son considerados como nodos hijos, en donde estos nodos hijos también pueden actuar como nodo raíz o nodo padre, Disposición Circular (B) y la disposición en resorte (C).

También podemos encontrar un módulo llamado Mapeo Visual (Visual Mapper), el cual está encargado de la forma y el tamaño de los nodos como por ejemplo el color del borde, el grosor y el estilo, en la cual se le asigna estas características a un atributo de datos dependiendo si es un atributo continuo o discreto (Figura 1.5)[26].

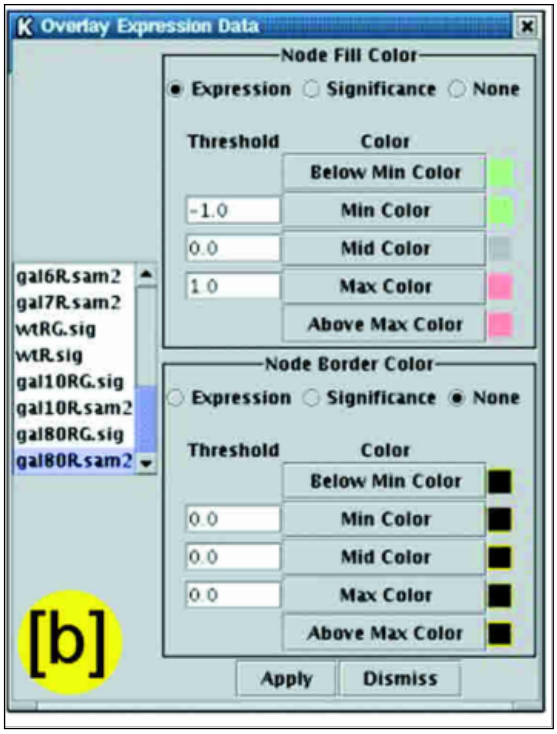


Figura 1.5: Ventana de Cytoscape donde se ajustan los detalles visuales de los atributos, como por ejemplo el color del nodo

1.2.5. Justificación del Problema

Dado que algunas interacciones sRNA-mRNA ocurran en regiones codificantes, es factible pensar que la regulación por sRNA y la adaptación a la disponibilidad del repertorio de tRNAs están estrechamente relacionadas. Hasta el momento se desconoce si ambos tipos de regulación de la expresión génica pueden actuar en forma independiente o están relacionados. La comprensión de la relación entre ambos mecanismos de regulación podría ser fundamental en el área de la ingeniería metabólica para el diseño racional de microorganismos genéticamente modificados, con aplicaciones biotecnológicas o biomédicas.

2. HIPÓTESIS

2.1. Hipótesis

En *Escherichia coli* K12 existe un conjunto de redes de regulación mediada por sRNA, en la cual los genes regulados por sRNA presentan un patrón de codones y adaptación al pool de tRNA diferente al presente en los genes no regulados por sRNAs.

3. OBJETIVOS

3.1. Objetivo

Caracterizar las redes de regulación sRNA-mRNA presente en *Escherichia coli* K-12 MG1655. Las redes se identificaron a partir de la información contenida en bases de datos de interacciones sRNA-blanco, junto con nuevas predicciones de interacciones sRNA-mRNA obtenidas a partir de la secuencia genética de este organismo bacteriano. Esta red además incluyó datos relevantes relacionados a los codones usados en cada gen, como la adaptación al repertorio de tRNA (tAI) y el valor de la adaptación al uso de codones (CAI).

3.2. Objetivos específicos

1. Identificación mediante búsqueda en las bases de datos las parejas de sRNAs y mRNA en *Escherichia coli* K-12 MG1655

2. Identificación y predicción de las regiones de interacción entre mRNA y sRNA.

3. Caracterización de la adaptación al repertorio de tRNA y uso codogénico de los genes regulados por sRNA y de sus regiones de interacción.

4. Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo y de la región de interacción.

4. MATERIALES Y MÉTODOS

4.1. Materiales

Los materiales utilizados para el desarrollo de esta tesis fueron:

4.1.1. Software

Los softwares que fueron utilizados para el desarrollo de esta memoria fueron los siguientes:

a. IntaRNA (http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp)

b. StarPicker (http://ccb.bmi.ac.cn/starpicker/prediction.php)

c. TargetRNA2 (http://cs.wellesley.edu/ btjaden/TargetRNA2/)

d. RNA Predator (http://nibiru.tbi.univie.ac.at/cgi-bin/RNApredator/target_search.cgi)

e. Cytoscape (Descargar desde: http://www.cytoscape.org/download.php)

f. MEME Suite (http://meme-suite.org/)

g. RNAfold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi)

h. MySQL Workbench (Descarga desde: https://dev.mysql.com/downloads/workbench/)

i. WampServer (Descarga desde: http://www.wampserver.com/en/)

j. JEMBOSS(Descarga desde: http://emboss.sourceforge.net/Jemboss/)

k. GraphPad Prism 6

l. Microsoft Excel

4.1.2. Bases de Datos

Las bases de datos que fueron utilizadas para el desarrollo de esta memoria fueron las siguientes:

a. sRNATarBase (http://ccb.bmi.ac.cn/srnatarbase/)

b. BSRD (http://www.bac-srna.org/BSRD/index.jsp)

4.1.3. Hardware

El Hardware que fue utilizado para el desarrollo de esta memoria fueron los siguientes:

a. Alienware 17R2: Memoria RAM: 16 GB DDR3; Procesador: Intel Core i7-4720HQ; Almacenamiento: 128 GB SSD y 1 TB HDD; GPU: Nvidia Gforce GTX980M 4GB GDDR5.

b. RedFox: Memoria RAM: 8GB DDR3; Procesador: Intel Core i5 CPU 650; Almacenamiento: 1TB HDD.

c. VAIO VPCL225FL: Memoria RAM 6GB DDR3; Procesador Intel Core i7-2630QM; Almacenamiento: 1 TB HDD; GPU: Nvidia Gforce GT 540M 4GB GDDR3

4.1.4. Scripts

a. Para la realización del cálculo de adaptación al pool de tRNA se utilizó el script PTAI.py desarrollado en lenguaje Python [31] y para el cálculo de la adaptación al uso codogénico se utilizó el script PCAI.py desarrollado en lenguaje Python [26].

b. Para dividir la secuencia codificante en 5'-CDS, CDS-M y CDS-3' se utilizó el script Contador_Carac.py desarrollado en lenguaje Python.

4.2. Metodología

4.2.1. Identificación mediante búsqueda bibliográfica de las parejas de sRNAs y mRNAs que interaccionan entre sí en *Escherichia coli*

4.3.1.1. Construcción de base de datos.

Para la construcción de la base de datos se utilizaron los datos ya existentes de las bases de datos sRNATarBase y BSRD, de las cuales se extrajo: ID, Nombre del sRNA, nombre del mRNA target, efecto (represión o inducción), secuencia del sRNA, secuencia del mRNA y el código DOI del artículo que hace referencia al estudio del sRNA.

4.2.2. Predicción e identificación de las regiones de interacción entre mRNA y sRNA

4.2.2.1. Identificar las regiones de interacción entre mRNA y sRNA

Para identificar y predecir las regiones de interacción de las interacciones no caracterizadas se utilizó el programa IntaRNA[1]. Se ajustaron los parámetros de entrada, de tal manera de utilizar una semilla equivalente a un tamaño de 7 bases[16]. Se almacenó en la base de datos anteriormente creada la energía de interacción, la ubicación de la interacción del mRNA con el sRNA, la posición de la semilla del sRNA y la posición de la semilla del mRNA.

4.2.2.2. Predecir regiones de interacción mRNA-sRNA

Para la predicción de las regiones de interacción se utilizó el programa RNA Predator. Se ingresó la especie bacteriana a la cual pertenece el sRNA al cual es objeto de la predicción de la interacción con los posibles mensajeros. Una vez que se obtuvieron las predicciones de cada uno de ellos se eligieron 2 interacciones con las energías más bajas. Posteriormente,? las interacciones se almacenaron en la base de datos.

4.2.2.3. Perfiles de interacción sRNA-mRNA.

Para la creación de zonas en los sRNAs lo primero que se hizo fue crear perfiles de interacción sRNA-mRNA. Para poder lograr esto se necesitó el largo del sRNA y la posición de la interacción del mensajero con el sRNA que se estaba perfilando. Teniendo estos datos se procedió a utilizarlos en el script llamado Matriz_suma.cpp, el cual crea la matriz en donde el largo de esta matriz corresponde al tamaño del sRNA y el ancho la cantidad de mensajeros que interaccionaban con él. Una vez obtenida la matriz se procedió a sumar los datos de interacción sRNA-mRNA, para posteriormente graficarlos dando como resultado un gráfico de perfiles en donde se identificaban las zonas de los sRNA donde ocurrían interacciones y las zonas en donde no ocurrían interacciones.

4.2.2.4. Cálculo de la densidad de interacción en las regiones de los mensajeros

Para calcular la densidad de las diferentes regiones de los mensajeros se dividió la razón de las interacciones (interacciones que ocurren en una zona específica / total de interacciones), por el tamaño de la zona que esta siendo analizada, y este valor fue multiplicado por 1000, de esta manera se obtuvo la densidad de las interacciones.

[1] http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp

4.2.3. Caracterización de la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción

4.2.3.1. Determinación de la Adaptación al repertorio de tRNA

La adaptación al repertorio de tRNA se determinó mediante el cálculo de tAI (tRNA Adaptation Index), para este cálculo se aplicó el script PTAI.py sobre el archivo que contiene las secuencias de los cDNAs en formato FASTA. Se registró el output del archivo (valor de tAI) asociado al identificador del gen. Posteriormente se agregó esta información a la base de datos. La adaptación del repertorio de tRNA sobre la región de interacción se determinó mediante el cálculo de tAI de la región de hibridación, para esto se completaron los codones que faltaban para hacer coincidir los aminoácidos con los de la secuencia codificante completa, a continuación, se aplicó el script PTAI.py sobre el archivo que contenía las secuencias de las regiones de interacción de los mensajeros en formato FASTA. Se registró el output del archivo (valor de tAI) asociado al identificador del gen. Posteriormente se agregó esta información a la base de datos.

La adaptación del repertorio de tRNA sobre las zonas del área codificante (5'-CDS, CDS-M, CDS-3') de los mensajeros que interaccionan con sRNAs se determinó mediante el cálculo de tAI de las zonas mencionadas anteriormente, para la realización de este cálculo se tomó la secuencia codificante, la cual mediante el uso de un script llamado Contador_Carac.py dividió la zona codificante en las tres zonas mencionadas anteriormente. Una vez dividida la secuencia codificante en las 3 zonas se aplicó el script PTAI.py sobre los archivos que contenían las secuencias de cada una de las zonas del área codificante en formato FASTA. Se registraron los outputs de los archivos (valores de tAI) asociados a los identificadores de los genes. Posteriormente se agregó esta información a la base de datos.

La adaptación al repertorio de tRNA en las zonas del área codificante (5'-CDS, CDS-M, CDS-3') para los mensajeros que no interaccionan con sRNAs se determinó mediante el cálculo de tAI de las zonas mencionadas anteriormente, para la realización de este cálculo se tomaron todos los mensajeros de *Escherichia coli* K-12 MG1655 se eliminaron los mensajeros que interaccionaban con sRNAs, una vez eliminados estos mensajeros se utilizó el script llamado Contador_Caract.py, el cual dividió la zona codificante en las 3 zonas mencionadas anteriormente. Una vez dividida la secuencia codificante en las 3 zonas se aplicó el script PTAI.py sobre los archivos que contenían las secuencias de cada una de las zonas del área codificante en formato FASTA. Se registraron los outputs de los archivos (valores de tAI) asociados a los identificadores de los genes.

4.2.3.2. Determinación de la adaptación al uso codogénico

La determinación de la adaptación al uso codogénico se realizó mediante el cálculo de CAI (Codon Adaptation Index), para este cálculo se utilizó el software llamado JEMBOSS y su módulo llamado CAI, este se configuró para calcular el CAI de *Escherichia coli* K-12 MG1655 (E. coli) sobre el archivo en el que estaban contenidos los cDNA en formato FASTA. Se registró el output de JEMBOSS (valor de CAI) asociado al identificador del gen. Posteriormente se agregaron los datos obtenidos a la base de datos.

La determinación de la adaptación al uso codogénico sobre la región de interacción se determinó mediante el cálculo de CAI de la zona de hibridación sRNA-mRNA, se realizó de manera similar al cálculo de tAI para esta región, pero utilizando el módulo CAI del software JEMBOSS, configurándolo para calcular el CAI de *Escherichia coli* K-12 MG1655 (E. coli) sobre el archivo en el que estaban contenidas las secuencias de las regiones de interacción de los mensajeros en formato FASTA.

La determinación de la adaptación al uso codogénico sobre las zonas del área codificante (5'-CDS, CDS-M, CDS-3') tanto de los mensajeros que interaccionan con sRNAs como aquellos que no, se determinó de forma similar al cálculo de tAI para estas regiones, pero esta vez usando el software JEMBOSS, configurado para calcular el valor de CAI de *Escherichia coli* K-12 MG1655 (Eecoli). Posteriormente, para cada conjunto de datos se registró el output de JEMBOSS (valor de CAI) asociado al identificador del gen.

4.2.3.3. Identificación de diferencias significativas entre los distintos resultados de CAI y tAI

Para la identificación de diferencias significativas se hicieron comparaciones entre pares de conjunto de valores por medio del software GraphPad Prism 6 el cual por medio de los analisis estadisticos Mann-Whitney no pareado y Mann-Whitney pareado. Se definío si el conjunto de valores era significativamente diferente al par estudiado. Una vez obtenidos estos resultados se procedió a tabularlos para realizar las comparaciones entre los valores de uso codogénico y adaptación al tRNA de los mensajeros que presentan interacción con sRNA con los valores de uso codogénico y adaptación al tRNA de los mensajeros que no presentan interacción con sRNAs.

4.2.3.4. Identificación de regiones significativas (motivos) entre los mensajeros que interaccionan con los sRNAs

Para la identificación de motivos comunes entre los mensajeros que interaccionan con un mismo sRNAs, se utilizó el software MEME [4]. Para esto se comparó exclusivamente

las regiones de interacción mRNA-sRNA localizadas en todos los RNA mensajero que interactuaban con un mismo sRNA. Se consideró como un motivo válido, aquel que presentara un valor $p \leq 0{,}01$

4.2.4. Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo

4.2.4.1. Generación de red de interacción

Para la generación de la red de interacción biológica se utilizó el software Cytoscape de manera local[2], por medio de la generación de un archivo en formato SIF, en el cual están contenidos los datos de la interacción sRNA-mRNA, junto con los datos de relevancia biológica en los que encontramos el valor de tAI, el valor de CAI, el efecto de la interacción (represión, represión predicha, inducción, inducción predicha).

[2]http://www.cytoscape.org/download.php

5. Resultados

5.1. OE1: Identificación mediante búsqueda en las bases de datos de las parejas de sRNAs y mRNA en *Escherichia coli* K-12 MG1655.

A partir de los datos obtenidos desde sRNATarBase y BSRD, se obtuvieron un total de 96 sRNAs que interaccionaban con 304 mensajeros formando un total de 412 interacciones sRNA-mRNA, en la figura 6 se muestra un resumen de los datos obtenidos de ambas bases de datos:

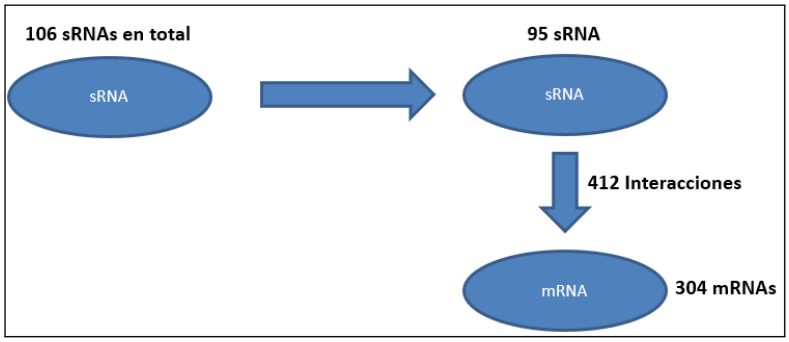


Figura 5.1: Resumen de datos obtenidos a partir de las bases de datos sRNATarBase y BSRD

De estas 412 interacciones, el 28,6 % (118) de ellas correspondió a interacciones donde se conocía tanto la identidad del sRNA como del mRNA, además del sitio de interacción sRNA-mRNA. El restante 71,3 % (294) correspondió a interacciones donde sólo se conocía la identidad del sRNA y mRNA que interactuaban. El primer conjunto de interacciones se denominó «interacciones caracterizadas» mientras el segundo se denominó interacciones «no caracterizadas».

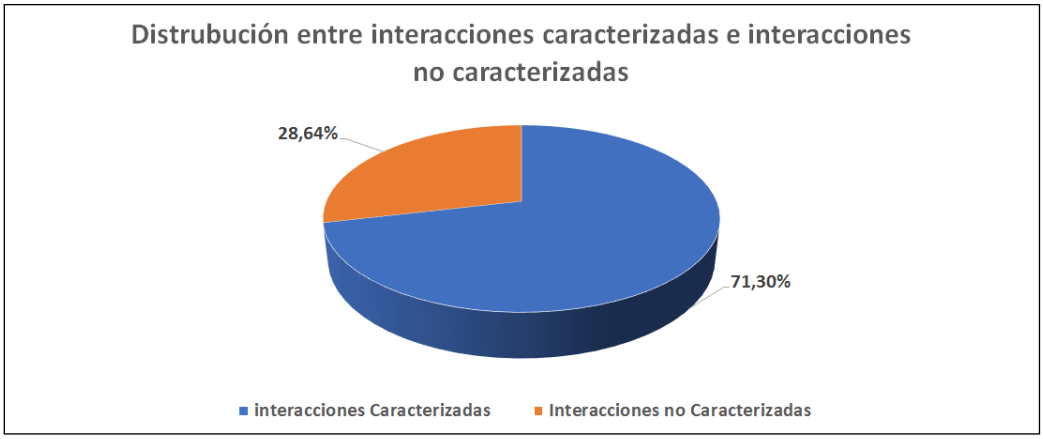


Figura 5.2: Distribución de Interacciones no caracterizadas versus Interacciones Caracterizadas, de color naranja encontramos las interacciones no caracterizadas, que corresponden al 71,30 % (294 interacciones) y de color azul se encuentran las interacciones caracterizadas con un 28,64 % (118 interacciones).

A continuación, se muestran las interacciones sRNA-mRNA extraídas desde las bases de datos utilizadas:

N°	sRNA	mRNA	Fuente
1	tpke11	ompX; proA	BSRD
2	sokC	hokC; mokC	BSRD
3	nc9	No posee Interacciones Caracterizadas	BSRD
4	sroA	gslM; maeA;	BSDR
5	sgrS	rpoS; ptsG; sodB; hns; ompA; dppA; ompF; ompC	BSRD
6	tp2	livJ;rplP	BSRD
7	tff	bssS; yihA	BSRD
8	c0067	No posee Interacciones Caracterizadas	BSRD
9	eyeA	minC; yibQ	BSRD
10	nc1	No posee Interacciones Caracterizadas	BSRD
11	sroB	agaV; chbC	BSRD
12	ipeX	proV; rzoQ	BSRD
13	sokE	cobB; menE	BSRD
14	sroC	carB; sodC	BSRD
15	rybA	sucA; yoeI	BSRD
16	rybB	ompC; ompS; pagC; yaeC; fadL; tsx; ompN; ompF; fiu; ompA; ompW	BSRD; sRNATarBase
17	psrD	proP; yahK	BSRD

18	c0293	uidC; yceJ	BSRD
19	dicF	ftsZ	BSRD; sRNATarBase
20	nc10	No posee Interacciones Caracterizadas	BSRD
21	c0299	No posee Interacciones Caracterizadas	BSRD
22	rdlA	frdC; hybC	BSRD
23	rdlB	atoC; ybjG	BSRD
24	rdlC	atoC; ybjG	BSRD
25	nc3	No posee Interacciones Caracterizadas	BSRD
26	rttR	mokC; xapR	BSRD
27	nc7	No posee Interacciones Caracterizadas	BSRD
28	mcaS	csgD; flhD	BSRD; sRNATarBase
29	fnrS	sodA; sodB; metE; folX; gpmA; folE; yobA; maeA; cydD; narK	BSRD; sRNATarBase
30	c0343	fliY; yjjP	BSRD
31	micC	ompC; sodB; hns; ptsG; ompA; dppA; ompF; galK	BSRD; sRNATarBase
32	rydC	yejA	BSRD; sRNATarBase
33	sokB	hokB; mokB	BSRD; sRNATarBase
34	c0362	creB; fliE	BSRD
35	mgrR	insA; ynjD	BSRD
36	dicF	ftsZ	BSRD
37	rydB	acrA; ppiC	BSRD
38	rprA	rpoS; sodB; ompA; ompC; ompF; hns	BSRD; sRNATarBase
39	sroD	ampC; ulaB	BSRD
40	ryeA	ydgI; ygeV	BSRD
41	ryeB	rpoS	BSRD
42	ryeF	psuT; rcsC	BSRD
43	c0465	argD; mscL	BSRD
44	isrB	rffC; ycJO	BSRD
45	dsrA	hns; rbsD; argR; ilvI; rpoS; ptsG; ompA; ompC; ompF; galK	BSRD; sRNATarBase
46	rseX	ompA; ompC	BSRD; sRNATarBase
47	isrC	tmcA; yqaE	BSRD
48	sibA	apaG; xylF	BSRD

49	ryeC	mtn; yaeR	BSRD
50	sibB	Fbp; xylF	BSRD
51	ryeD	rseP; yhhK	BSRD
52	cyaR	luxS; yqaE; nadE; ompX	BSRD; sRNATarBase
53	micF	ompF; hns; ptsG; ompA; ompC; dppA; galK	BSRD; sRNATarBase
54	tpke70	mnmA; nlpC	BSRD
55	sroE	fdhD; phoP	BSRD
56	c0614	gltX; rmA	BSRD
57	is128	Kup; trmL	BSRD
58	ryfA	ascG; napB	BSRD
59	glmY	glmS	BSRD; sRNATarBase
60	ryfB	yaeI; yeeJ	BSRD
61	ryfC	gltD; ycjZ	BSRD
62	ryfD	mprA; yfbS	BSRD
63	ssrA	smpB	BSRD
64	sraD	tsx; lamB; ompX; ompA; phoP	BSRD; sRNATarBase
65	c0664	sucC; ytfR	BSRD
66	sokX	Frr; ybaQ	BSRD
67	csrB	csrA	BSRD; sRNATarBase
68	gcvB	cycA; sstT; ilvE; ybdH; oppA; dppA; ndK; brnQ; argT; gdhA; thrL; livK; yaeC; gltI; sstT; iciA; serA; livJ; livC; lrp; sodB; dksA; hns; ptsG; ompA; ompC; ompF	BSRD; sRNATarBase
69	omrA	flhD; flhC; fecA; btuB; glmM; folP; fepA; cirA; ompT; gntP; fecD; csgD; lit; gmhB; deoR; nadA; glcD; uup; clpB; yccS; cydD; ybdC; hisM; folA; yadD; sufD; mipA; ybcS; ygjN; yhbE; csiE; xylH; hokD; fdoI; yeaZ; yfbT; rumA; yzgL; narH; yrfC; yadL; cheZ; malK; yaeP; ssuC; hokB; ydhT; ompA; fmF; csgB; osmB; ybeT;	BSRD; sRNATarBase

70	omrB	fecA; glmM; folP; gntP; csgD; ykfI; ypdD; yaeH; b2680; yaiY; yphD; trxC; ybjE; fldA; yjhl; srlB; mutM; ybhT; ykgJ; yeeE; yhdN; ompA; fimF; csgB; osmB; ybeT; fepA; csgD	BSRD; sRNATarBase
71	invR	ompC	BSRD
72	ssrS	sspA	BSRD
73	sibC	kduD; menH	BSRD
74	rygC	kduD; menH	BSRD
75	c0719	appB; rbbA	BSRD
76	sroG	pgI; ulaR	BSRD
77	rygD	rfbB; sufD; ybaS; yqjE	BSRD
78	sibB	Fbp; xylF	BSRD
79	rygE	yeaE; yjaZ; ypdI	BSRD
80	sibE	stpA	BSRD
81	psrN	fadH; tnaC	BSRD
82	nc4	No posee Interacciones Caracterizadas	BSRD
83	psrO	caiA; garL	BSRD
84	sraH	sdaC; STM3216; tpx	BSRD
85	ryhB	iscS; fur; sdhD; sodB; acnA; fumA; shiA; bfr; ftn; sucB; sucC; sucD; gltA; mdh; icd; ybjG; yagT; perM; yadS; kdpA; citG; yagJ; sugE; yheF; motA; pinH; ygeZ; ynfF; yiaM; yegK; ygiQ; ydaN; ykgE; yecD; metH; proA; metI; ygiT; hns; ompA; ompC; ompF; galK; ptsG	BSRD; sRNATarBase
86	arrS	hofC; msyB	BSRD
87	gadY	gadX	BSRD, sRNATarBase
88	nc8	No posee Interacciones Caracterizadas	BSRD
89	rdlD	ldrD	BSRD, sRNATarBase
90	sokA	corA; galK	BSRD
91	istR-1, istR2	tisB	BSRD, sRNATarBase
92	glmZ	glmS	BSRD, sRNATarBase
93	glmY	glmS	BSRD

94	esre	cynT; thiM	BSRD
95	spot42	nanC; glpF; xylF; caiA; atoD; galK; fucI; puuE; ybbW; paaK; gltA; nanT; fucP; srlAM; sthA; ascF; sucC; ompA; ompC; ompF; hns; dppA; sodB	BSRD; sRNATarBase
96	csrC	csrA	BSRD; sRNATarBase
97	oxyS	rpoS; yobF; fhlA; tolA; tolB; tolR; yeaK; yfdH; dppD; rpmG; gfcB; fabB; ybbB; yccE; yheN; pmbA; moaD; yeaC	BSRD; sRNATarBase
98	p26	abgA; rcnA	BSRD
99	sroH	gntX; pitB	BSRD
100	sraL	rlmA; ycjU	BSRD
101	ryjA	No posee Interacciones Caracterizadas	BSRD
102	nc5	No posee Interacciones Caracterizadas	BSRD
103	ryjB	cdgR; ydbK	BSRD
104	nc6	No posee Interacciones Caracterizadas	BSRD
105	symR	symE	BSRD; sRNATarBase
106	sraA	idyK; prpR	BSRD
107	ffs	bcsA; ymgD	BSRD
108	nc2	No posee Interacciones Caracterizadas	BSRD

Tabla 5.1: sRNAs y sus blancos presentes en la bases de datos BSRD (Bacterial Small Regulatory RNA Database) o en la base de datos sRNATarBase.

5.2. OE2: Identificación y predicción de las regiones de interacción entre mRNA y sRNA.

5.2.1. Predicción de los blancos de los sRNAs sin interacción identificada.

Los resultados anteriores mostraron que existían 12 sRNA cuyos blancos de interacción no estaban identificados (sRNA huérfanos). Para completar nuestra base de datos se decidió identificar los posibles blancos para estos sRNA huérfanos. Para esto se utilizó el software **RNAPredator**[1]. De los sitios de interacción obtenidos por el programa, para cada sRNA solo se consideraron los dos mRNA blancos que mostraron una interacción más favorable (menores valores de energía libre de Gibbs) (Tabla 5.2).

[1]http://nibiru.tbi.univie.ac.at/cgi-bin/RNApredator/target_search.cgi

N°	sRNA	Blanco
1	b1574	*dedA (-17,39 Kcal/mol), ymdF (-23.29 Kcal/mol)*
2	nc1	*intF (-145,85 Kcal/mol), livM (-19,15 Kcal/mol)*
3	nc2	*fepB (-26,12 Kcal/mol), ybdD (-22,93 Kcal/mol)*
4	nc3	*cysA (-19,54 Kcal/mol), narX (-40,05 Kcal/mol)*
5	nc4	*tdcB (-38,6 Kcal/mol), uhpT (-22.28 Kcal/mol)*
6	nc5	*flgG (-31,04 Kcal/mol), uhpT (-41,74 Kcal/mol)*
7	nc6	*dkgA (-18,82 Kcal/mol), yjiN (-37,95 Kcal/mol)*
8	nc7	*fabI (-43,68 Kcal/mol), melR (-25,86 Kcal/mol)*
9	nc8	*astE (-24,5 Kcal/mol), rnpA (-22,9 Kcal/mol)*
10	nc9	*chaA (-24,04 Kcal/mol), tdcC (-22,83 Kcal/mol)*
11	nc10	*ymfL (-9,16 Kcal/mol), ymfM (-7,14 Kcal/mol)*
12	SsrA	*smpB (-10,50 Kcal/mol)*

Tabla 5.2: La tabla muestra los resultados de predicción de los blancos de los sRNAs huérfanos, junto con la predicción de su energía de interacción sRNA-mRNA.

5.2.2. Identificación de las regiones de interacción sRNA-mRNA blanco.

El análisis anterior mostró que existían 298 interacciones en las cuales no se encontraba caracterizada la región de interacción sRNA-mRNA blanco. Para predecir estas regiones de interacción se utilizó el software IntaRNA, escogiendo la interacción de menor energía libre de Gibbs. Los resultados mostraron que las energías de interacciones predichas estuvieron en un rango de entre -24,65 kcal/mol y 13,66 kcal/mol con una media de -9,84 kcal/mol. Estos resultados son comparables a los observados en las energías de las interacciones validadas empíricamente las cuales presentan un rango que va desde -72,62 kcal/ml a 76,70 kcal/mol, con una media de 9,82 kcal/mol.

De esta manera, nuestra base de datos quedó conformada por 435 interacciones, caracterizadas en su región de unión sRNA-mRNA blanco por métodos empíricos (118; 27,1 %) y bioinformáticos (317; 72,8 %) (Figura 5.3).

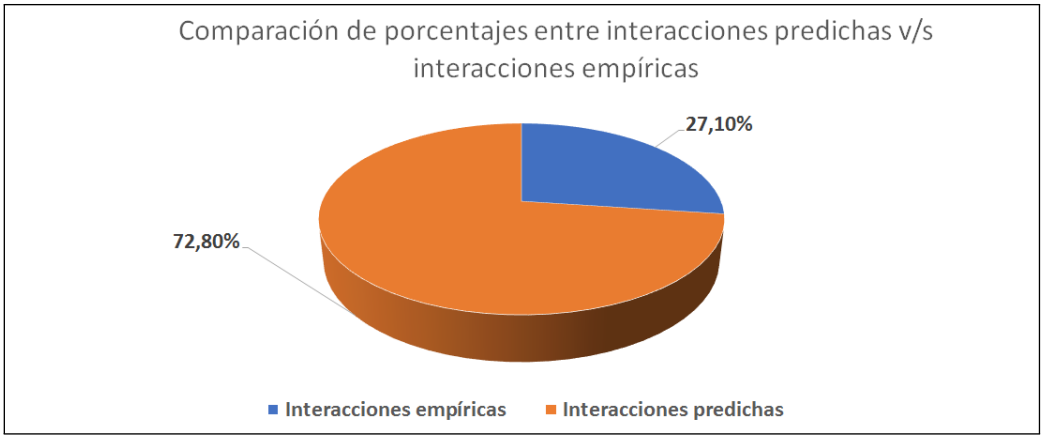


Figura 5.3: Porcentaje de interacciones predichas y empíricas en *E. coli.*

Las interacciones sRNA-mRNA blanco pueden ocurrir en distintas regiones del mensajero, 5'-UTR, región codificante, y la zona 3'-UTR. Por el mecanismo de regulación de la expresión génica de los sRNA sobre los mensajeros, la mayor parte de las interacciones empíricamente validadas ocurre en la región 5'-UTR o en el inicio de la traducción. Para averiguar si esto también ocurre en las regiones de interacción predichas, se procedió a clasificar las interacciones de acuerdo a su zona de interacción en el mRNA blanco. Los resultados muestran que las regiones predichas se localizan mayormente en el área codificante, luego vienen las interacciones en la región 5'-UTR. Es interesante destacar que el 30,1 % de las interacciones en la zona codificante se localiza en los primeros 150 nucleótido (5'-CDS) (ver tabla 3). Esta región ha sido definida como una rampla traduccional donde existe un incremento progresivo en la adaptación al uso codogénico, permitiendo un incremento gradual de la velocidad de avance del ribosoma[34].

Zona	**Interacciones empíricas**	**Interacciones predichas**	**Total**
5' UTR	31	38	69
5'UTR - CDS	34	33	67
5'- CDS	17	72	89
5'-CDS - CDS-M	1	8	9
CDS-M	25	157	182
CDS-3'	1	2	3
3'UTR	4	12	16
Total	113	322	435

Tabla 5.3: Distribución de las interacciones por zonas del mensajero

La presencia de un mayor número de interacciones en la zona codificante del mRNA blanco, podría ser resultado del mayor tamaño de esta zona respecto a las otras regiones. Para evaluar esta posibilidad se calculó la densidad de interacciones por región del mensajero, dividiendo el número de interacciones por la cantidad de interacciones totales y por el tamaño de la zona donde interactúa el sRNA (Sección 4.3.2.3 de Materiales y Métodos). Los resultados muestran que en las interacciones determinadas mediante métodos empíricos la mayor densidad se encuentra en la zona 5'UTR-CDS con un valor de 1,9 interacciones/base (Tabla 5.4). En las interacciones predichas la zona 5' UTR-CDS mostró un valor de densidad similar con 1,49 interacciones/base, sin embargo, en las interacciones localizadas en la región codificante las interacciones predichas mostraron una probabilidad 57 veces mayor. Lo que se ve reflejado en que el 48,8 % de las interacciones predicha se localizó en esta región (Tabla 5.4).

Zona	Densidad de Interacciones empíricas (Razón/Tamaño zona*1000)	Densidad de Interacciones Predichas (Razón/Tamaño zona*1000)	Razón (DIP/DIE)
5' UTR	5,42	2,33	0,43
5' UTR - CDS	13,55	5,20	0,38
5'- CDS	1,92	1,47	0,76
5'-CDS - CDS-M	8,05	1,26	0,15
CDS-M	0,009	0,56	57,69
CDS-3'	7,62	0,21	0,02
3'UTR	0,16	0,75	4,46

Tabla 5.4: Densidad de interacción en las diferentes regiones del mRNA.

Como se puede apreciar en la Tabla 5.4 para el caso de las interacciones empíricas e interacciones predichas de las zonas 5'-CDS, presentan una probabilidad de interacción similar pues las densidades de interacción son muy similares. Este resultado es distinto al observado en la zona CDS-M donde se obtuvo valores de densidad más altos para las interacciones predichas, esto a pesar de ser normalizado por el tamaño de la región de interacción y por la cantidad de interacciones predichas.

5.2.3. 5.2.3. Funciones biológicas reguladas por sRNA.

Para identificar las funciones biológicas de los mRNA blancos regulados por sRNA se utilizó la herramienta BLASTKOALA[12]. BLASTKOALA es una herramienta para identificar funciones biológicas a partir de un set secuencias aminoácidicas, el cual puede clasificar en 18 tipos funciones biológicas (como por ejemplo: Familia de enzimas, procesamiento de la información genética, procesos celulares, etc). Primero se analizaron los mRNA regulados por sRNA cuyas interacciones no han sido predichas por medio de RNAPredator (Un total de 303 mensajeros). De estos, BLASTKOALA permitió asignar una función biológica a 284 mRNAs. La mayor cantidad de mensajeros regulados por sRNA están relacionados con el procesamiento de la información ambiental (108), y en segundo lugar se encuentran los mensajeros relacionados a procesos celulares (48) (Figura 5.4)

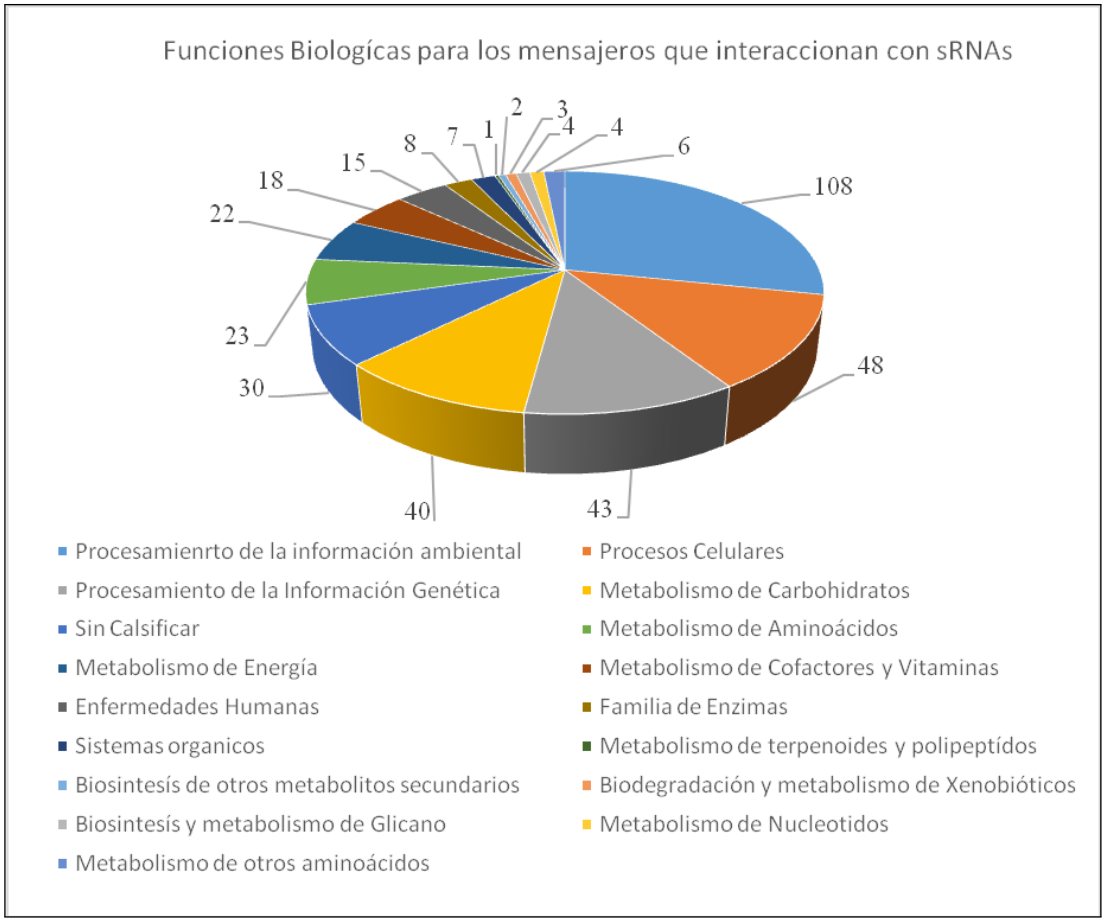


Figura 5.4: Funciones biológicas de los mRNAs regulados por sRNA. La figura muestra los procesos biológicos de E. coli regulados por sRNA, así como la cantidad de mensajeros regulados en cada uno de estos procesos.

En el caso de los mensajeros obtenidos por medio de predicciones bioinformáticas (24 mensajeros), estos participan en 9 proceso biológicos. La función biológica encargada del procesamiento de la información ambiental también presentó la mayor cantidad de mensajeros, con un total de 8 (equivalente al 36,36 % de los nuevos blancos identificados), en segundo lugar, con un total de 4, encontramos los mensajeros que no clasifican en alguna función biológica (equivalente a un 18,18 % de los nuevos blancos identificados), en tercer lugar se encuentran el metabolismo de aminoácidos, el metabolismo de energía y los procesos celulares con un total de 2 mensajeros por cada una de estas funciones (equivalentes a un 9,09 % para cada una de estas funciones biológicas).

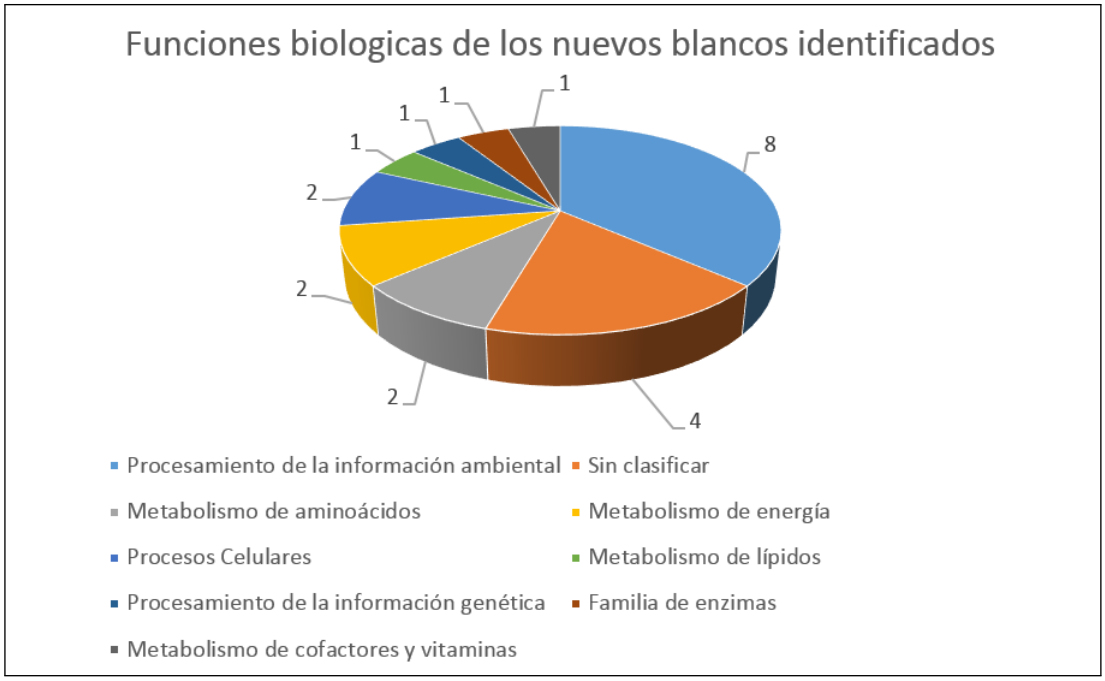


Figura 5.5: **Funciones biológicas de los mRNAs regulados por sRNA e identificados por métodos bioinformáticos** La figura muestra los posibles procesos biológicos de *E. coli* regulados por sRNA sin un blanco identificados por algún método empírico, y el número de mensajeros regulados en cada uno de estos procesos.

5.2.4. 5.2.4. Localización de sitios de interacción sRNA-mRNA en los sRNAs

En *E.coli* los sRNAs pueden interactuar con uno o varios mRNA blanco. Para determinar si en los sRNA que interactúan con múltiples blancos existe un patrón común de interacción se utilizó un análisis matricial en el cual se asignó un valor de 1 a la posición del sRNA equivalente al nucleótido que interactúa con el mRNA, y un valor de 0 en el caso contrario. Este análisis se aplicó a los sRNAs que interactúan con al menos 4 mRNAs diferentes y permitió obtener un mapa de frecuencia de interacción sRNA-mRNA a lo largo del sRNA. Se logró identificar un total 16 sRNA con al menos 4 interacciones, siendo OmrA el sRNA con mayor número de blancos. Al graficar el número de interacciones por cada nucleótido del sRNA (perfil de interacción) se observó que las interacciones no se distribuían homogéneamente a lo largo del sRNA, sino más bien existían zonas con mayor abundancia, y regiones donde no ocurre interacción alguna. Esta distribución sirvió para dividir el sRNA en zonas identificadas con las letras A, B, C, D y F, y en subzonas identificadas con números (Por Ejemplo A1, A2, etc) (Tabla 5.5). Estas zonas dentro del sRNA pueden interactuar con uno o varios mensajeros a la vez.

N°	sRNA	N° de mRNAs que Interactúan	Zona (N° de Interacciones)
1	OmrA	50	A1(39), A2(21), B (0)
2	RyhB	43	A1(16), A2(33), A3(8)
3	OmrB	26	A1(19), A2(18), B (0)
4	GcvB	26	A (19), B (0), C (7), D (0)
5	Spot42	23	A1(6), A2(8), A3(14), B (0), C (1)
6	OxyS	18	A (8), B (0), C1(3), C2(9), D (0)
7	MicA	14	A (14), B (0)
8	RybB	11	A (11), B (0)
9	FnrS	10	A1(2), A2(8), A3(3), B (0)
10	DsrA	10	A (0), B1(6), B2(8), B3(1), C (0)
11	SgrS	8	A (0), B (8), C (0)
12	MicC	8	A (2), B (0), C (6), D (0)
13	MicF	7	A1(5), A2(2), C (0)
14	RprA	6	A (0), B (6), C (0)
15	CyaR	4	A1(2), A2(3), B (0)
16	RygD	4	A(0), B(1), C(0), D(1), E(0), F(0), G(0), H(1), I(0)

Tabla 5.5: Organización de las interacciones en los sRNAs con más de 4 mRNA blancos.

5.2.4.1 Análisis de OmrA.

Para el caso de OmrA que interacciona con 50 mensajeros su estructura se dividió en 2 zonas principales (A y B) y 2 zonas secundarias (A1, A2) (Figura 5.6A). Solo en la zona A (Zona OmrA-A) ocurre la interacción sRNA-mRNA. Esta zona puede a su vez ser dividida en otras dos zonas secundarias OmrA-A1 y OmrA-A2. En OmrA-A1 ocurren un total de 35 interacciones mientras que en OmrA-A2 ocurren 15. Para determinar si este perfil de interacción está relacionado con la estructura secundaria del sRNA, se mapearon las diferentes zonas identificadas en un modelo de la estructura de OmrA construido por medio del software RNAfold WebServer (Figura 5.6B)

Como se puede observar la zona OmrA-B (que no presenta interacciones con mRNAs) tiene una estructura de horquilla rica en GC, con una alta probabilidad de estar apareado por lo tanto representaría una estructura más estable. En el caso de la zona OmrA, esta región presenta un menor contenido de GC y mayor probabilidad de formar regiones

desapareadas, las cuales podría servir parar formar las interacciones iniciales con los mRNA blanco. Es interesante que la sub-zona OmrA-A1 que presenta un mayor número de interacciones también es la sub-zona con menor contenido GC y menor probabilidad de formar estructura estable.

5.2.4.2 Análisis de RyhB.

Este sRNA interacciona con 43 mensajeros y el perfil de interacciones permitió dividir su estructura en 1 zona principal y 3 zonas secundarias RyhB-1A, RyhB-1B y RyhB-1C (figura 5.7A). A través de la zona 1A, RyhB interactúa con 16 mRNA s diferentes. Nueve de estos mRNA interactúan con RyhB exclusivamente a través de esta zona, mientras que los otros siete mRNAs interactúan con RyhB a través de las otras dos zonas (RyhB-1B y RyhB-1C). Para el caso de la zona RyhB-1B, esta región interactúa con 33 mRNAs, 20 de los cuales interactúan exclusivamente con esa zona mientras que los otros 13 presentan interacciones compartidas con zonas RyhB-1A y RyhB-1C. Para el caso de RyhB-1C, en esta zona ocurren un total de 8 interacciones, de las cuales solo 1 es exclusiva (Figura 5.7B).

Al mapear este perfil de interacción en la estructura secundaria de RyhB, se puede observar que las zonas RyhB-1A y RyhB-1B, presentan un menor contenido de CG y una mayor probabilidad de formar regiones desapareadas que interactúen con los mRNA blanco, consistente con esto la sub-zona RyhB-1B que presenta un menor contenido de GC también es la sub-zona que presenta el mayor número de interacciones. Por otro lado, la zona RyhB-1C presenta una estructura de horquilla rica en CG, con una alta probabilidad de estar apareado es también la zona con menor número de interacciones.

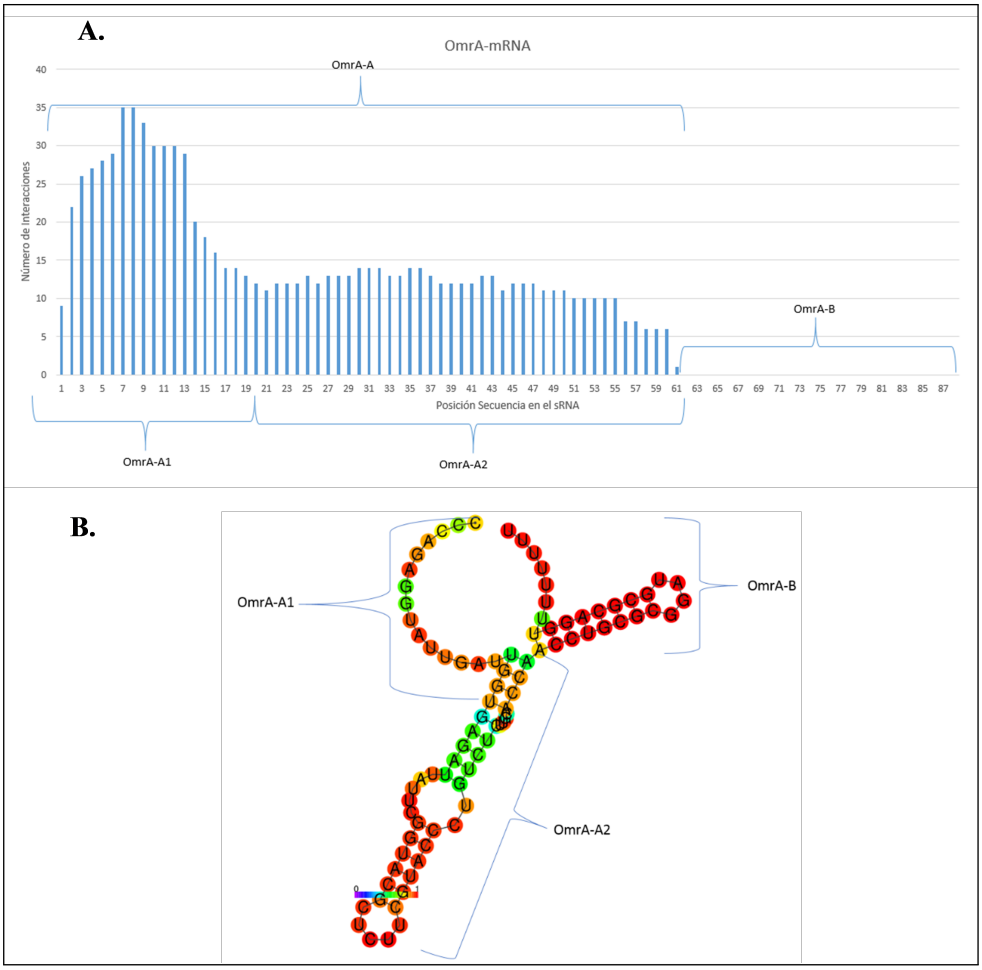


Figura 5.6: **Análisis matricial de OmrA.** En A se observa el gráfico correspondiente al perfil de interacción sRNA-mRNA de OmrA. En B se observa la estructura secundaria predicha de OmrA con el Software RNAfold WebServer, se muestra la localización de las regiones identificadas en A. La escala de colores representa la probabilidad de formar una estructura secundaria estable con ella misma

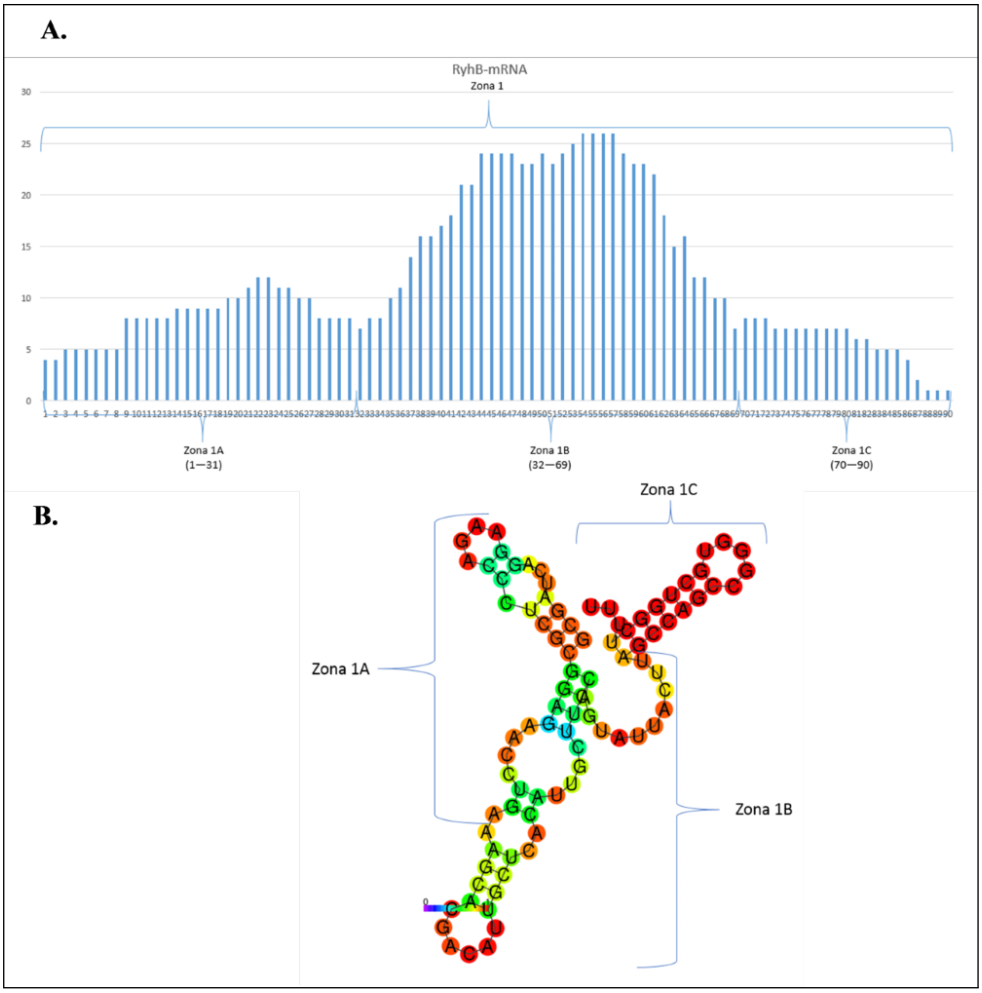


Figura 5.7: **Análisis matricial de RyhB.** En A se observa el gráfico correspondiente al perfil de interacción sRNA-mRNA de RyhB. En B se observa la estructura secundaria predicha de OmrA con el Software RNAfold WebServer, se muestra la localización de las regiones identificadas en A. La escala de colores representa la probabilidad de formar una estructura secundaria estable con ella misma

5.2.4.3 Análisis de OmrB.

OmrB, que interactúa con 26 mensajeros, puede ser dividido en 2 zonas principales OmrB-A y OmrB-B (Figura 5.8A). En OmrB-A se localizan todas las interacciones con los mRNA-blanco. La zona OmrB-A puede a su vez puede ser dividida en otras 2 zonas secundarias OmrB-A1 y OmrB-A2. En OmrB-A1 ocurren un total de 20 interacciones, de las cuales 7 corresponden a mRNA que interaccionan con OmrB exclusivamente a través de esta zona, mientras que los otros 13 corresponden a interacciones con mRNA que también interactúan con la otra zona secundaria (OmrB-A2). Por otra parte, en OmrB-

A2 es el blanco de 19 mRNAs, 6 de los cuales interaccionan con OmrB exclusivamente a través de esta sub-zona. Al mapear el número de interacciones en la estructura secundaria del sRNA, nuevamente las regiones con menor contenido GC y mayor probabilidad de estar desapareadas son las que presentan una mayor cantidad de interacciones(OmrB-A1). Mientras que las regiones con alta probabilidad de formar una horquilla no presentan interacciones con mRNAs (Figura 5.8B).

5.2.4.4 Análisis de GcvB.

Este sRNA Interactúa con 26 mensajeros. Su estructura pudo ser dividida en 4 zonas principales GcvB-A, GcvB-B, GcvB-C, GcvB-D (Figura 5.9A). En las zonas GcvB-A y GcvB-C, se localizan todas las interacciones sRNA-mRNA. En GcvB-A y GcvB-C se pudieron identificar un total de 19 y 7 interacciones, respectivamente. Como se puede observar al mapear el perfil dentro de la estructura secundaria de GcvB, las zonas GcvB-B y GcvB-D tienen una estructura de horquilla rica en GC, con una alta probabilidad de estar apareado por lo tanto representa una estructura más estable. En el caso de las zonas GcvB-A y GcvB-C, estas regiones presentan un menor contenido de GC y una mayor probabilidad de formar regiones desapareadas que interactúen con los mRNA blanco. La zona GcvB-A que presenta un mayor número de interacciones también es la zona con mayor contenido CG de las 2 zonas en donde ocurren interacciones (Fhigura 5.9B).

5.2.4.5 Análisis de Spot42.

El perfil de interacción de Spot42 con sus mRNA blanco permite dividir a este sRNA en 3 zonas principales (Spot42-A, Spot42-B, y Spot42-C) y 3 zonas secundarias (Figura 5.10A). En las zonas Spot42-A y Spot42-C se localizan las interacciones sRNA-mRNA, mientras que Spot42-B no presentan ninguna interacción. La zona Spot42-A puede ser dividida en otras 3 zonas secundarias Spot42-A1, Spot42-A2 y Spot42-A3. En Spot42-A1 se pueden identificar siete interacciones, de las cuales 3 corresponden a mRNA que interactúan exclusivamente con esta zona en esta zona y 4 presentan interacciones compartidas con otras zonas (Spot42-A2 y Spot42A3). En la zona Spot42-A2 ocurren un total de 8 interacciones, de las cuales 3 son exclusivas y las otras 5 son interacciones compartidas con las zonas Spot42-A2 o Spot42-A3. La zona Spot42-B tiene una estructura de horquilla rica en CG, con una alta probabilidad de estar apareado por lo tanto representa una estructura más estable. En el caso de la zona Spot42-A, esta región presenta un menor contenido de GC y mayor probabilidad de formar regiones desapareadas que interactúen con los mRNA blanco. Es interesante que la sub-zona Spot42-A3 que presenta un mayor

número de interacciones, también es la sub-zona con menor contenido GC y mayor probabilidad de formar una estructura estable (Figura 5.10B).

Los resultados obtenidos muestran que en las regiones con alto contenido en GC cumplirían roles de carácter estructural, pues no formarían interacciones con mRNAs. Para averiguar si esto se cumple en todos los casos de sRNA que interactúan con múltiples mRNAs, se cotejó el contenido GC vs el número de interacciones en las respectivas sub-zonas. Los resultados muestran que con excepción de RyhB, en todos los sRNA con múltiples interacciones, existe una región del sRNA con alto contenido en GC que no interactúa con ningún mensajero (Tabla 5.6, Anexo C)

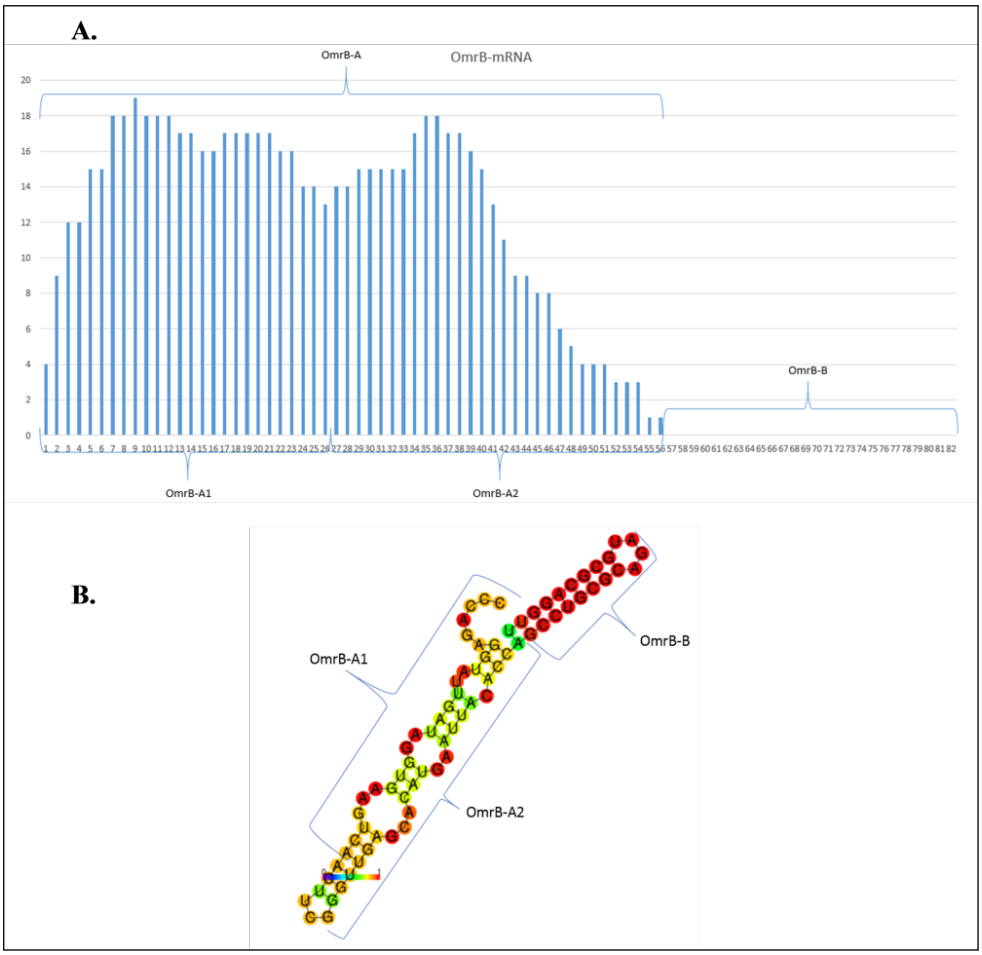


Figura 5.8: **Análisis matricial de OmrB.** En A se observa el gráfico correspondiente al perfil de interacción sRNA-mRNA de OmrB. En B se observa la estructura secundaria predicha de OmrB con el Software RNAfold WebServer, se muestra la localización de las regiones identificadas en A. La escala de colores representa la probabilidad de formar una estructura secundaria estable con ella misma

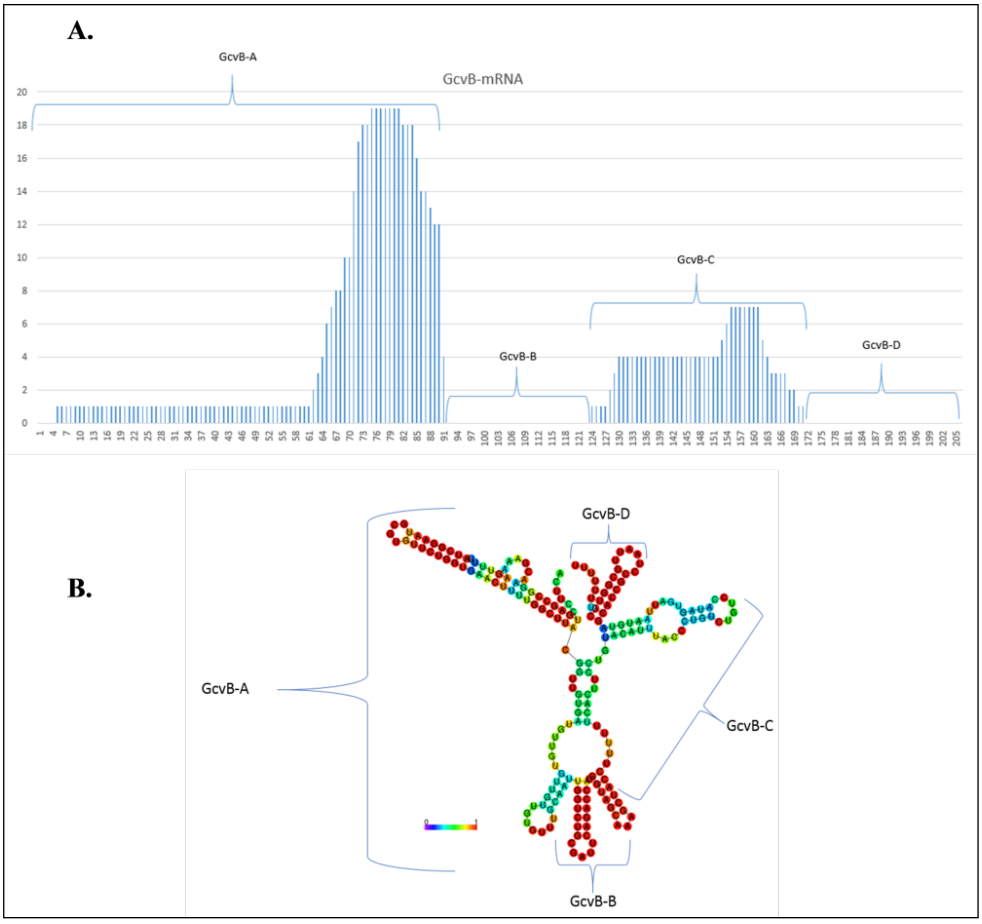


Figura 5.9: **Análisis matricial de GcvB.** En A se observa el gráfico correspondiente al perfil de interacción sRNA-mRNA de GcvB. En B se observa la estructura secundaria predicha de GcvB con el Software RNAfold WebServer, se muestra la localización de las regiones identificadas en A. La escala de colores representa la probabilidad de formar una estructura secundaria estable con ella misma

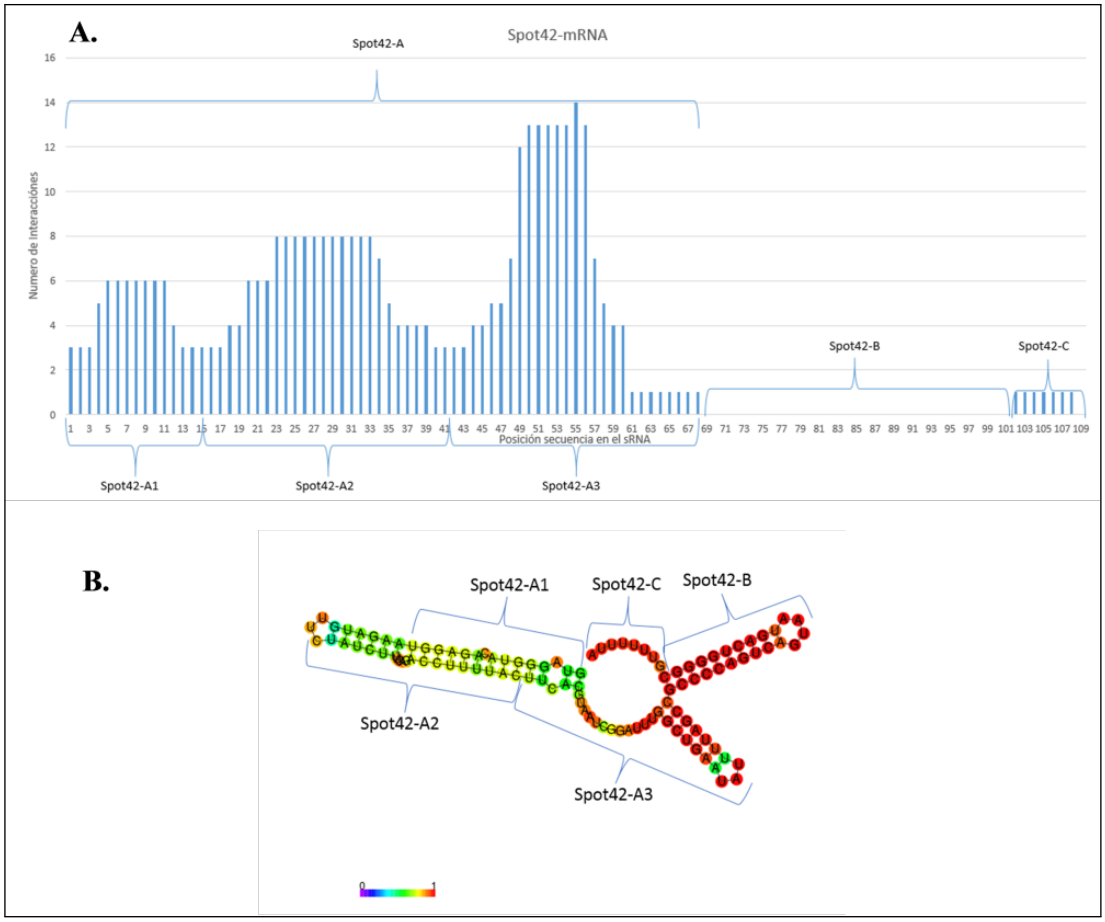


Figura 5.10: **Análisis matricial de Spot42.** En A se observa el gráfico correspondiente al perfil de interacción sRNA-mRNA de Spot42. En B se observa la estructura secundaria predicha de Spot42 con el Software RNAfold WebServer, se muestra la localización de las regiones identificadas en A. La escala de colores representa la probabilidad de formar una estructura secundaria estable con ella misma

sRNAs	**Presenta horquilla CG estable**
OmrA, OmrB, GcvB, Spot42, OxyS, MicA, RybB, FnrS, DsrA, SgrS, MicC, MicF, RprA, CyaR, RygD	Si
RyhB	No

Tabla 5.6: Resumen de la presencia de horquilla GC estables donde no ocurre interaccion con mensajeros en las estructuras secundarias de los sRNAs que interaccionan con 4 o más mensajeros, con la finalidad de identificar si en la estructura secundarias de los sRNAs se forma una zona en la cual no ocurre interacción sRNA-mRNA

5.3. OE3: Caracterización de la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción.

5.3.1. Análisis estadístico de los CAI y tAI de los mensajeros regulados por sRNAs.

Para determinar si existe una relación entre el uso de codones de mRNAs regulados por sRNA y aquellos que no, se caracterizó primero los valores de CAI de los genes a lo largo de la región codificante y de las regiones 5'-CDS, CDS-M, CDS-3', de ambos grupos. Al comparar los genes regulados por sRNA vs los no regulados por sRNAs solo se encontró diferencias significativas entre los valores de CAI o tAI de la región 3' CDS ($p < 0{,}05$) siendo los valores de CAI de los genes regulados por sRNA levemente mayores a los no regulados. En el caso de los valores de tAI, en esta región también se observa la misma tendencia, pero no alcanza a lograr una significancia estadística (Tabla 5.7). Cuando se analizó si a lo largo de los genes regulados existían diferencia entre sus regiones, se observó que la región 5'CDS presentaba diferencias significativas con las tres regiones, esto también se repitió en los genes no regulados (datos no mostrado) sugiriendo que en los genes regulados por sRNA no existe una composición codogénica diferente a la presente en los genes no regulados. Un resultado interesante apareció cuando se comparó el valor de CAI o tAI de la zona de interacción con cada una de las regiones de CDS. La región de interacción mostró valores de CAI mayores a la región 5'-CDS, y CDS-3', esto implica que en estas regiones se utilizan codones que se encuentran con mayor frecuencia a lo largo del genoma, por otra parte, al realizar el mismo análisis usando los valores de tAI, se observó que la región de interacción presenta menores valores de tAI que cada una de las regiones analizadas (Tabla 5.8). En su conjunto estos resultados sugieren que en las regiones de interacción se utilizan preferencialmente codones de alto uso, pero que tienen pocos tRNA que los decodifican. Esto implica que en las regiones donde ocurre la interacción sRNA-mRNA los ribosomas tendrían una tasa de avance más lento en relación con el resto de la región codificante. Con la finalidad de corroborar estos resultados se realizaron los mismos test, pero pareados para ver si las diferencias significativas son por casos individuales (Tabla 5.9), en donde se observó que la tendencia mostrada en la Tabla 5.8 se repite en la Tabla 5.9

Comparación genes regulados vs no regulados en sus valores de CAI y tAI				
Índice	**Región**	**Media Regulados por sRNA**	**Media No regulados por sRNA**	**Valor P**
Codon Adaptation Index (CAI)	5'CDS	0.6885 ± 0.003794	0.6870 ± 0.001086	0.7153
	CDS-M	0.7159 ± 0.004670	0.7144 ± 0.001315	0.7560
	CDS-3'	0.6994 ± 0.007894	0.6806 ± 0.002111	0.0217*
tRNA Adaptation Index (tAI)	5'CDS	0.2460 ± 0.001705	0.2438 ± 0.0004673	0.2197
	CDS-M	0.2564 ± 0.001528	0.2553 ± 0.0007802	0.5183
	CDS-3'	0.2636 ± 0.003571	0.2569 ± 0.0009786	0.0709

Tabla 5.7: Tabla comparativa CAI y tAI de los mensajeros que interaccionan con sRNA v/s los mensajeros que no interaccionan con sRNA. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney no pareado. El asterisco sugiere significancia estadística

A partir de los datos obtenidos en la Tabla 5.7 la hipótesis planteada para esta memoria es rechazada por lo que con la finalidad de encontrar un factor que indique como se forma la interacción sRNA-mRNA se realizaron los siguientes estudios:

Comparación de regiones en genes regulados por sRNAs				
Índice	**Región**	**Media Región Interacción**	**Media de Región CDS**	**Valor P**
Codon Adaptation Index (CAI)	5'-CDS	0,728 ± 0,159	0,688 ± 0,068	<0,001*
	CDS-M	0,728 ± 0,159	0,715 ± 0,060	0,2161
	CDS-3'	0,728 ± 0,159	0,699 ± 0,134	0,0076*
	CDS	0,728 ± 0.159	0,713 ± 0,061	0,1159
tRNA Adaptation Index (tAI)	5'-CDS	0,216 ± 0,006	0,246 ± 0,002	<0,0001*
	CDS-M	0,216 ± 0,006	0,256 ± 0,002	<0,0001*
	CDS-3'	0,216 ± 0,006	0,266 ± 0,007	<0,0001*
	CDS	0,216 ± 0,006	0,262 ± 0,001	<0,0001*

Tabla 5.8: Comparación de los valores de CAI y tAI entre la región de interacción y los valores presentes en las regiones codificantes. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney no pareado. El asterisco sugiere significancia estadística.

Comparación de regiones en genes regulados por sRNAs (Análisis Pareado)			
Índice	**Región**	**Media de las Diferencias**	**Valor-P**
Codon Adaptation Index (CAI)	5'-CDS	0.03084 ± 0.1501	0.0002*
	CDS-M	-0.01108 ± 0.1498	0.1792
	CDS-3'	0.02990 ± 0.1820	0.0030*
	CDS	-0.000096 ± 0.1477	0.9905
tRNA Adaptation Index (tAI)	5'-CDS	0.4733 ± 0.1530	<0.0001*
	CDS-M	0.4625 ± 0.1541	<0.0001*
	CDS-3'	0.4609 ± 0.1623	<0.0001*
	CDS	0.4631 ± 0.1540	<0.0001*

Tabla 5.9: Comparación de los valores de CAI y tAI entre la región de interacción y los valores presentes en las regiones codificantes. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney pareado. El asterisco sugiere significancia estadística.

5.3.2. Regiones conservadas en mRNAs reconocidos por el mismo sRNA.

El resultado del análisis de densidad de interacciones sRNA-mRNA indicó que existían regiones dentro del sRNA que reconocen específicamente un conjunto de mRNA. Para determinar si en estos mRNA existe una región conservada o motivo que reconozca a esta región del sRNA se decidió realizar un análisis de búsqueda de motivos conservados por cada uno de los mensajeros que interaccionan con sRNAs. Por medio del uso de MEMEsuit (http://meme-suite.org) se ingresaron la secuencia de las regiones de interacciones, y el software por medio de uso de cadenas de Markov localizó una posible secuencia consenso. A continuación, se muestran los resultados de este análisis.

5.3.2.1. OmrA

En la Tabla 5.10, se presenta los motivos identificados en todos los mRNAs que interactúan con OmrA o con sus zonas (OmrA-A1 y OmrA-A2):

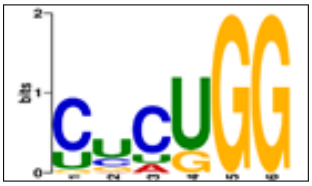
sRNA	**Valor-P**	**Tamaño**	**Mensajeros en donde está el Motivo**	**Motivo**	**N° de Secuencias**
OmrA	1,7x10^{-1}	6	*14 (sufD, cheZ, folP, csgB, yadL, yfB, csiE, clpB, ygjN, glcD, yzgL, yrfC, hisM, cydD, xylH, osmB, rumA, liT)*		14/50
OmrA-A1	9,6x10^{-10}	6	21 (gntP, ygjN, hokB,ssuC, csgD, ybeT, mipA, clpB, yzgL,ompT, narH, fecA, xylH, fepA, uup, csgB, csiE, folP, sufD, glmM)		21/39
OmrA-A2	7,9x10^{-3}	6	20 (malK, gmhB, xylH, yrfC, deoR, btuB, yccS, cydD, glcD,hokD, yhbE, yaeP, cheZ, yeaZ, yadD, hisM, ydhT, ydbC, fdoI, fimF)		20/20

Tabla 5.10: Secuencias consenso para OmrA, junto con las zonas que fueron localizadas en éste.

Para el caso de la región consenso de OmrA de todos los mensajeros que interaccionan con este se encontró una secuencia consenso que presenta un Valor-p de 1,7x10^{-1} con un tamaño de 6 y está localizado en 14 de los 50 casos utilizados para buscar la secuencia consenso de los mensajeros que interaccionan con OmrA, lo que representa a un 36.73 % del total de los casos utilizados, estadísticamente la secuencia consenso no es válida. Para la región OmrA-A1 la secuencia consenso identificada presenta un Valor-P de 9,6x10^{-10} con un tamaño de 6 y está presente en 21 de los 35 casos utilizados para buscar la secuencia consenso de los mensajeros que interaccionan con OmrA, lo que representa a un 53,8 % del total de los casos utilizados, estadísticamente la secuencia consenso es válida y como está presente en más de un 50 % de los casos utilizados se puede considerar una secuencia consenso. Para el caso de OmrA-A2 la secuencia consenso identificada presenta un Valor-P

de 7,9x10^{-3} con un tamaño de 6 y está presente en 20 de los 20 casos utilizados para buscar la secuencia consenso de los mensajeros que interaccionan con OmrA, lo que representa a un 100 % del total de los casos utilizados, estadísticamente la secuencia consenso no es válida y aunque esté presente en más de 50 % de los casos, como estadísticamente no es significativo esta secuencia no se considera una secuencia consenso para este caso.

5.3.2.2. RyhB

Para el caso de la región consenso de RyhB de todos los mensajeros que interaccionan con este se encontró una secuencia consenso que presenta un Valor-P de 1x10^{-8} con un tamaño de 8 pb y está localizado en 20 de los 43 mensajeros que interactúan con RyhB, lo que porcentualmente representa un 46,51 % del total de los casos utilizados. Para la región RyhB-A1 la secuencia consenso identificada presenta un Valor-P de 0,26 con un tamaño de 7 y está localizado en el 87,5 % de las secuencias analizadas. Para el caso de la región RyhB-A2 la secuencia consenso identificado tiene un Valor-P de 8,9x10^{-13} con un tamaño de 8 y está presente 19 de 33 mensajeros que interaccionan con esta región, lo que representa a un 57,57 % del total de mensajeros utilizados. Para el caso de la región RyhB-A3 la secuencia consenso identificado presenta un Valor-P de 1,4x10^{-3} con un tamaño de 8 e identificada en 5 de los 8 mensajeros que interaccionan en esta región, lo que representa a un 55,5 % del total de los mensajeros. A continuación, la Tabla 5.11 resume los resultados para cada una de las regiones presentes en RyhB:

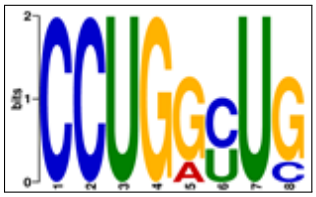
sRNA	Valor-P	Tamaño	Mensajeros en donde está el Motivo	Motivo	N° de Secuencias
RyhB	$1x10^{-8}$	8	*20(sucD, proA, ygiQ, yagJ, yegK, sucB, yadS, sugE, bfr, ydaN, ompC, metI, yiaM, perM, yheF, galK, motA, icd, mdh, sodB)*		20/43
RyhB-A1	$2,6x10^{-1}$	7	14 (fur, ompF, ftn, sucC, ompA, ptsG, ygeZ, hns, ynfF, kdpA, perM, metI, ybjG, sdhD)		14/16
RyhB-A2	$8,9x10^{-13}$	8	19 (sucD, proA, ygiQ, yagJ, yegK, yiaM, bfr, ydaN, sucB, sugE, yadS, metH, ygiT, mdh, metI, citG, yagT, fumA, shiA)		19/33
RyhB-A3	$1,4x10^{-3}$	8	5 (ompC, galK, metI, sodB, iscS)		5/8

Tabla 5.11: Secuencias consenso para RyhB, junto con las zonas que fueron identificadas para este sRNA.

5.3.2.3. OmrB

Para el caso de OmrB, no se identificaron motivos estadísticamente significativos. Sin embargo, se identificó un motivo común a todos los mRNA que interactúan con OmrB. Este motivo tiene un tamaño de 6 nucleótidos, pero un valor-P de $2x10^{2}$ (Tabla 5.12).

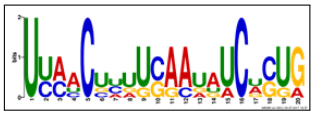
sRNA	Valor-P	Tamaño	Mensajeros en donde está el Motivo	Motivo	N° de Secuencias
OmrB	2,0x10^{2}	6	*25 (yeeE, ypdD, osmB, csgD, ompA, fecA, csgB, folP, yhdN, mutM, ykfI, ybhT, fimF, yaiY, fldA, ybeT, yphD, trxC, gntP, ybjE, fepA, srlB,yjhI, glmM, yaeH)*		25/26
OmrB-A1	6,2x10^{-1}	20	5 (fimF, ypdD, yaiY, fldA, ybjE)		5/19
OmrB-A2	1,5x10^{-1}	8	9 (yphD, ypdD, ybeT, yaeH, ykfI, folP, trxC, ybjE, yhdN)		9/18

Tabla 5.12: Secuencias consenso para OmrB, junto con las zonas que fueron identificadas para este sRNA.

5.3.2.4. GcvB

En relación a los mRNA que interactúan con GcvB, se identificó un motivo con un Valor-P de 2,9 x 10^{-3} en 14 de los 26 mensajeros que interaccionan con GcvB, lo que significa que está presente a un 53,8 % de los mensajeros. Para el caso de la región GcvB-A se encontró una secuencia consenso que presenta un Valor-P de 7,3x10^{-6} y está presente en un 89,4 % (17/19) de los mensajeros que interaccionan con esta región. La secuencia consenso identificada para GcvB fue la misma que se identificó en el análisis de GcvB-A, probablemente debido a que el mayor número de interacciones que establece GcvB lo hace a través de la región GcvB-A. Para el caso de la región GcvB-C se identificó una secuencia consenso que presenta un Valor-P de 1,1x10^{-1} con un tamaño de 6 y está localizado en el 100 % de los mensajeros utilizados en este caso (Tabla 5.13).

sRNA	Valor-P	Tamaño	Mensajeros en donde está el Motivo	Motivo	N° de Secuencias
GcvB	2,9x10^{-3}	14	*14 (oppA, ydbH, ilvC, livK, ompA, thrL, dppA, iciA, ompF, ompC, sodB, livJ, ndK, ptsG)*		14/26
GcvB-A	7,3x10^{-6}	17	17 (oppA, ybdH, ilvC, livK, thrL, ompA, iciA, ompF, dppA, ompC, sodB, ndK,,livJ, ptsG, yaeC, gltI, sstT)		17/19
GcvB-C	1,1x10^{-1}	7	7 (brnQ, hns, gdhA, lrp, cycA, dksA, serA)		7/7

Tabla 5.13: Secuencias consenso para GcvB, junto con las zonas que fueron identificadas para este sRNA

5.3.2.4. Spot42

En el caso de mensajeros que interactúan con Spot42 se identificó un motivo cuyo Valor-P es de 5,4x10^{-10} y está localizado en el 91,3 % de los mensajeros que interaccionan con este sRNA. Para el caso de la región S42-A1 se encontró una secuencia consenso de 8 nucleótidos con un Valor-P de 5,1x10^{-5} presente en el 100 % de los mensajeros que interaccionan con esta región. Para el caso de la región S42-A2 se identificó un motivo de 6 nucleótidos (Valor-P = 1,6x10^{-1}) localizado también en el 100 % de los mensajeros que interaccionan con esta región. A pesar de esto, debido a su valor P no se puede asignar a este motivo significancia estadística. En la región S42-A3 también se identificó un motivo de 8 nucleótidos (Valor-P = 1,7x10^{-11}) localizado en 10 de los 14 mensajeros que interaccionan en esta región (76,9 %) (Tabla 5.14)

sRNA	Valor-P	Tamaño	Mensajeros en donde está el Motivo	Motivo	N° de Secuencias
Spot42	5,4x10^{-10}	8	*21 (fucP, ascF, ompC, ompA, galK, ompF, puuE, nanC, dppA, glpF, fucI, gltA, atoD, xylF, sucC, srlA, ybbW , paaK, caiA, nanT, sthA)*		21/23
S42-A1	5,1x10^{-5}	8	6 (gltA, atoD, nanC, paaK, galK, xylF)		6/6
S42-A2	1,6x10^{-1}	6	8 (puuE, sucC, caiA, galK, srlA, paaK, ybbW, xylF)		8/8
S42-A3	1,7x10^{-11}	8	10 (fucP, ompC, ompA, glpF, fucI, ompF, puuE, ascF, dppA, galK)		10/14

Tabla 5.14: Secuencias consenso para Spot42, junto con las zonas que fueron identificadas para este sRNA

5.4. OE4: Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo y de la región de interacción.

Los sRNAs participan formado una red de interacción que regula la expresión génica en E. coli. Para tener una adecuada comprensión de esta red, es útil usar aproximaciones

que permitan su visualización incorporando diversos tipos de información. En este caso, el tipo de regulación sobre la expresión génica, la adaptación al uso de codones, adaptación al pool de tRNA, regiones presentes dentro de cada sRNA, procesos biológicos involucrados y motivos presentes en sus mRNA blanco. Para la generación de la red de interacción se utilizó al software Cytoscape junto con el software Microsoft Excel para crear el script que creará la red de interacción sRNA-mRNA en E. coli. Primero se determinó una red donde se incorporó el tipo de regulación, sRNAs, mRNA blanco, y tipo de interacción (experimental o predicha). Los resultados de esta red se muestran en la Figura 5.11, en esta red se incluyeron las 435 interacciones que fueron establecidas en esta tesis.

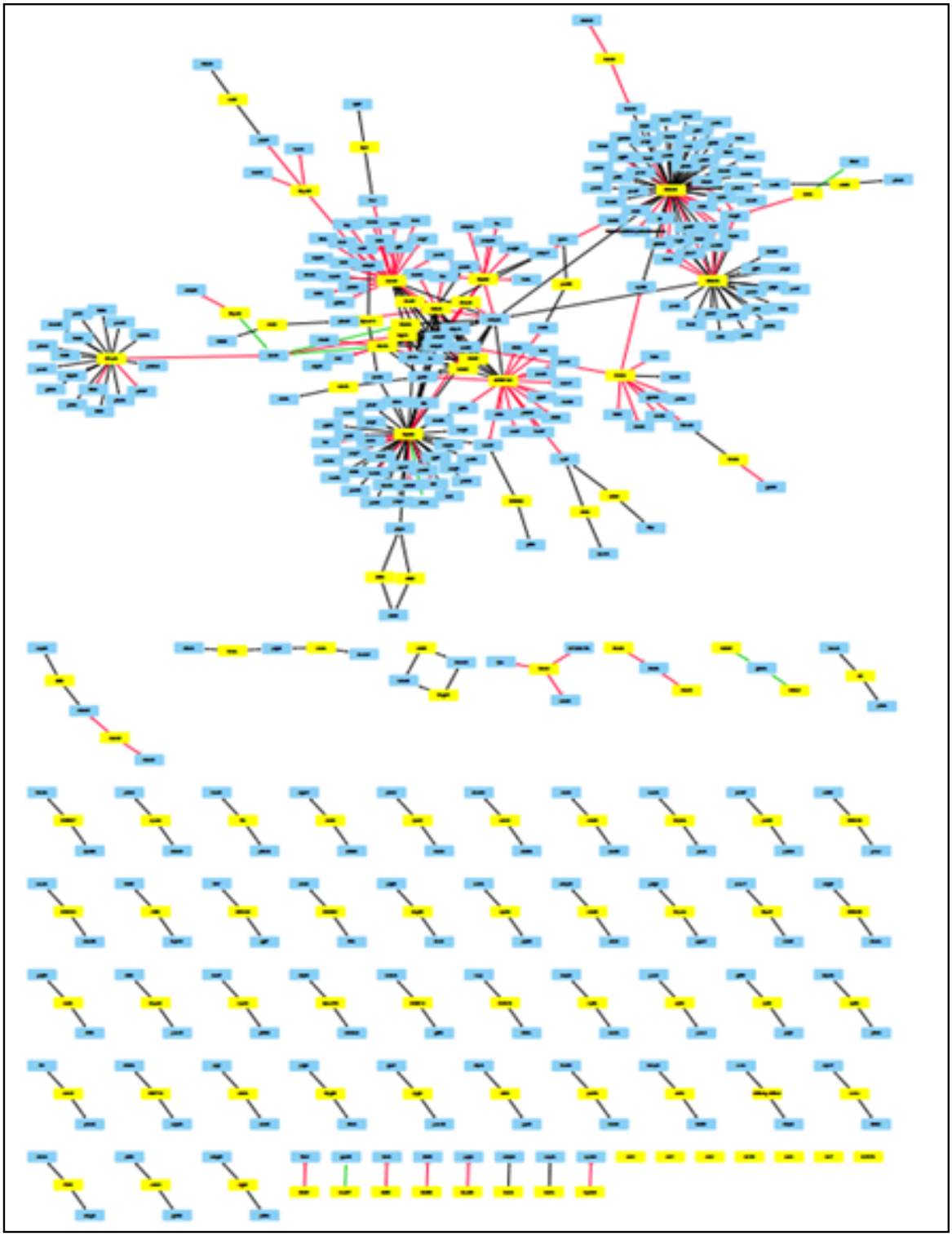


Figura 5.11: **Modelo de red de interacción sRNA-mRNA en *E. coli*** En donde se puede visualizar un área de mayor densidad en donde se ubican la mayor cantidad de interacciones comunes y otra zona en donde se encuentran los sRNAs que interaccionan con 2 mensajeros y que no comparten interacciones entre sRNAs. Como se puede apreciar los rectángulos amarillos representan los sRNAs, mientras que los rectángulos azules representan a los mensajeros que forman interacción sRNA-mRNA. Para el caso de las líneas las líneas rojas representan represiones, las verdes inducciones y las negras representan interacciones no definidas.

El siguiente paso fue incorporar en la red, la información sobre las zonas identificadas en cada uno de los mRNA que establecen múltiples interacciones.

Para crear una visualización más completa de la red de interacción en *E. coli* se crearon modelos de los sRNAs que interaccionaban con sobre 4 mensajeros con la finalidad de crear una representación de los gráficos obtenidos en el OE2 los cuales se muestran a continuación:

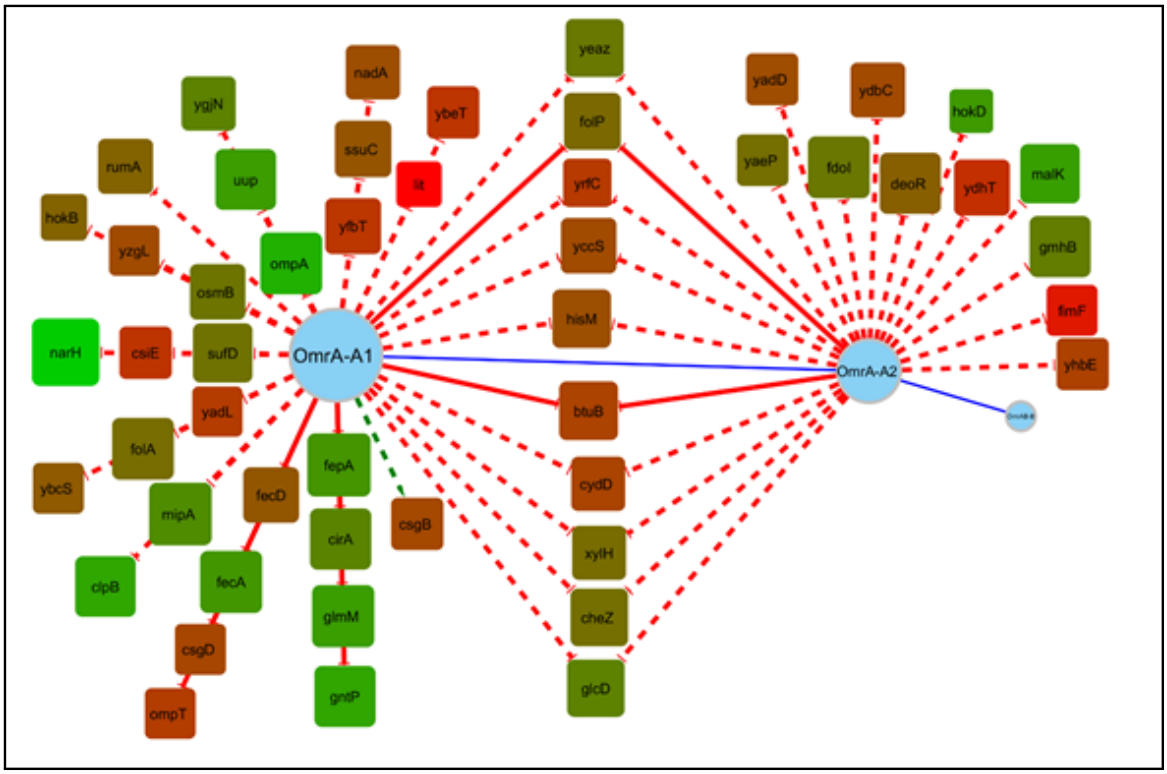


Figura 5.12: **Modelo de red de interacción de OmrA** Se puede apreciar la distribución de zonas de OmrA, junto con esto se pueden apreciar las 3 zonas de OmrA y el tamaño de estas se ve reflejado por las cantidades de interacciones sRNA-mRNA, las líneas rojas punteadas representan represiones predichas por métodos bioinformáticos, las líneas verdes punteadas representan inducciones predichas por métodos bioinformáticas y las rojas sólidas representan represiones identificadas por métodos empíricos. Los valores de CAI son indicados en una escala de colores en el nodo (color verde representa CAI bajos y el color rojo representa a CAI más altos). El tamaño del nodo representa el valor de tAI con respecto al máximo y mínimo identificados en *E. coli*.

Para el caso de RyhB que posee 43 interacciones se puede apreciar en la figura 5.13 como está organizada:

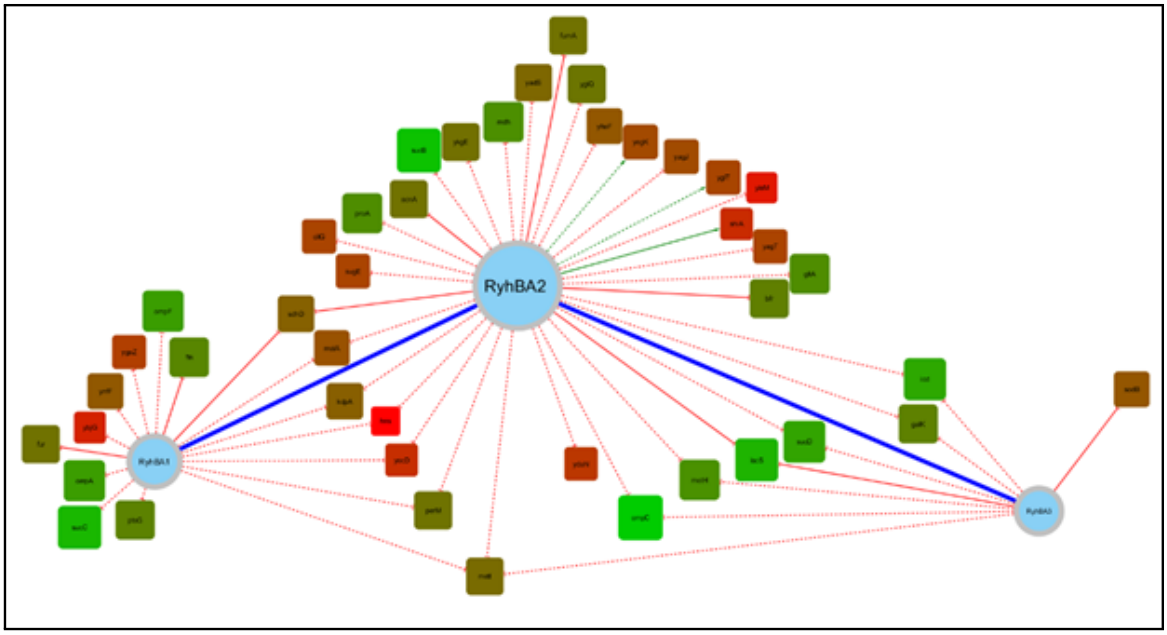


Figura 5.13: **Modelo red Interacción RyhB** Modelo red Interacción RyhB. Se puede apreciar las 3 zonas de RyhB, donde el tamaño de estas se ve reflejado por las cantidades de interacciones sRNA-mRNA que forman, las líneas rojas punteadas representan represiones predichas por métodos bioinformáticos, las líneas verdes punteadas representan inducciones predichas por métodos bioinformáticos, las rojas sólidas representan represiones identificadas por métodos empíricos y las líneas verdes sólidas representan inducciones identificadas por métodos empíricos. Los colores de los nodos representan los valores de CAI de los mensajeros con respecto al CAI máximo y al CAI mínimo identificados en el genoma de *E. coli* (0,924 valor máximo y 0,3613 valor mínimo). Para el caso de los tamaños de los mensajeros estos reflejan al tAI con respecto al máximo y mínimo identificados en *E. coli.* (0,3779 valor máximo y 0,1501 valor mínimo)

Como puede apreciarse en la Figura 5.13 se puede observar la distribución de las interacciones alrededor de las zonas del RyhB, como también la presencia de mensajeros que interaccionan en 2 zonas a la vez (13 interacciones que interaccionan entre 2 zonas), también se puede apreciar mensajeros que interaccionan con 3 zonas a la vez (1 interacción que interactúa con las 3 zonas a la vez) también se puede apreciar que en la zona Ryhb-A2 existen la mayor cantidad de interacciones que están registradas para RyhB (un total de 20) de los cuales están divididos en 18 represiones y 2 inducciones, en el caso de la zona RyhB-A1 se pueden encontrar un total de 9 interacciones las cuales las 9 son represiones y en el caso de la zona RyhB-A3 es la zona que presenta menos cantidad de interacciones, con un total de 1 interacción.

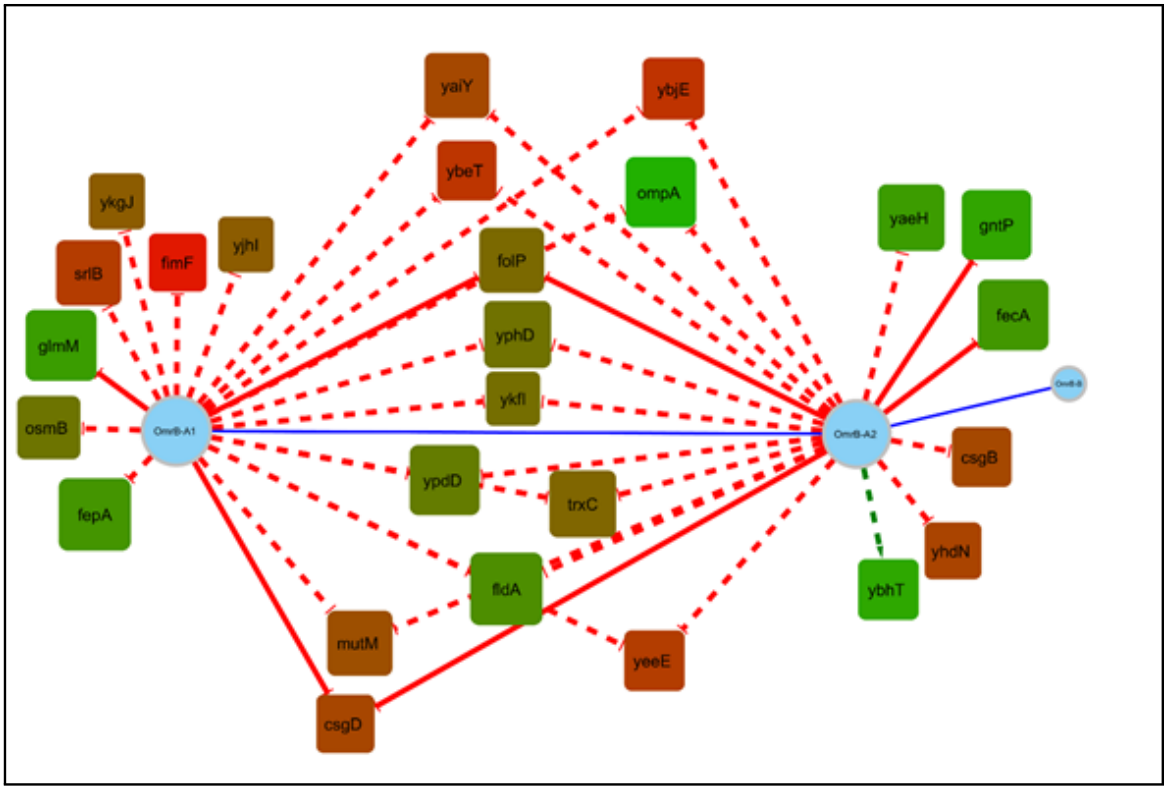


Figura 5.14: **Modelo de red de interacción OmrB** Se puede apreciar las 3 zonas de OmrB, donde el tamaño de estas se ve reflejado por las cantidades de interacciones sRNA-mRNA que forman, las líneas rojas punteadas representan represiones predichas por métodos bioinformáticos, las líneas verdes punteadas representan inducciones predichas por métodos bioinformáticos y las líneas rojas sólidas representan represiones identificadas por métodos empírico. Los colores de los nodos representan los valores de CAI de los mensajeros con respecto al CAI máximo y al CAI mínimo identificados en el genoma de *E. coli* (0,924 valor máximo y 0,3613 valor mínimo). Para el caso de los tamaños de los mensajeros estos reflejan al tAI con respecto al máximo y mínimo identificados en *E. coli* (0,3779 valor máximo y 0.1501 valor mínimo)

Como puede apreciarse en la **Figura 5.14** se puede observar la distribución de las interacciones alrededor de las zonas del OmrB. También se puede observar la presencia de mensajeros que interaccionan en 2 zonas a la vez (13 interacciones que interaccionan entre 2 zonas), también se puede apreciar que en la zona OmrB-A1 existen las mayores cantidades de interacciones que están registradas para OmrB (un total de 7) de los cuales todas son represiones, en el caso de la zona OmrB-A2 se pueden encontrar un total de 6 interacciones las cuales las 5 son represiones y solo 1 inducción. En el Caso de la zona OmrB-B no presenta ninguna interacción esto se puede deber a que es en esa zona donde la proteína HFQ interacciona con el sRNA.

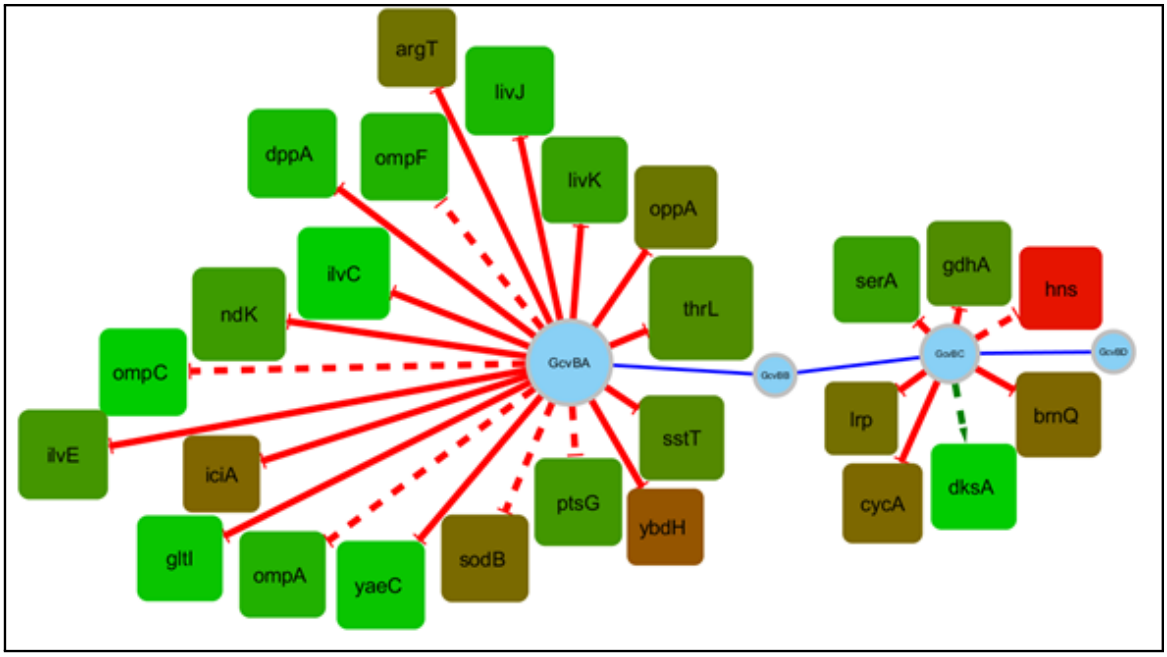


Figura 5.15: **Modelo de red de interacción GcvB** Se puede apreciar las 4 zonas de GcvB, donde el tamaño de estas se ve reflejado por las cantidades de interacciones sRNA-mRNA que forman, las líneas rojas punteadas representan represiones predichas por métodos bioinformáticos, las líneas verdes punteadas representan inducciones predichas por métodos bioinformáticos y las líneas rojas sólidas representan represiones identificadas por métodos empírico. Los colores de los nodos representan los valores de CAI de los mensajeros con respecto al CAI máximo y al CAI mínimo identificados en el genoma de *E. coli* (0,924 valor máximo y 0,3613 valor mínimo). Para el caso de los tamaños de los mensajeros estos reflejan al tAI con respecto al máximo y mínimo identificados en *E. coli* (0,3779 valor máximo y 0,1501 valor mínimo).

Para el caso de GcvB (Figura 5.15), se puede observar que este sRNA está dividido en 4 zonas entre las cuales 2 realizan interacciones (GcvB-A y GcvB-C) y 2 zonas en las que no existe ningún tipo de interacción sRNA-mRNA (GcvB-C y GcvB-D). En el caso de las zonas que existen interacciones sRNA y mRNA, en la zona GcvB-A existe la mayor cantidad de interacciones en esa zona registradas para GcvB (18 interacciones) de las cuales todas son represiones, en comparación en la zona GcvB-C existe la menor cantidad de interacciones registradas para GcvB (7 interacciones) las cuales están dividida en 6 represiones y 1 inducción.

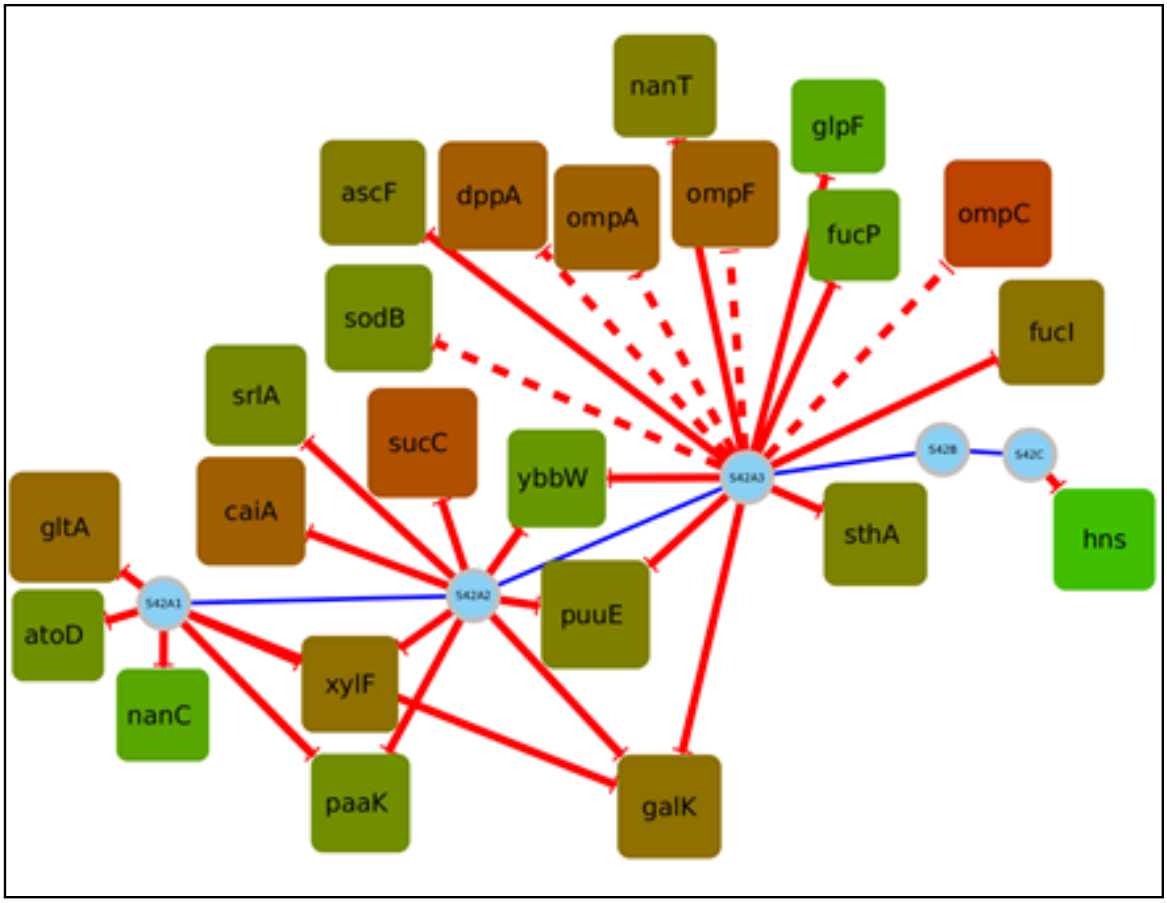


Figura 5.16: **Modelo de red de interacción Spot42** Se puede apreciar las 5 zonas de Spot42, donde el tamaño de estas se ve reflejado por las cantidades de interacciones sRNA-mRNA que forman, las líneas rojas punteadas representan represiones predichas por métodos bioinformáticos y las líneas rojas sólidas representan represiones identificadas por métodos empírico. Los colores de los nodos representan los valores de CAI de los mensajeros con respecto al CAI máximo y al CAI mínimo identificados en el genoma de *E. coli* (0.924 valor máximo y 0.3613 valor mínimo). Para el caso de los tamaños de los mensajeros estos reflejan al tAI con respecto al máximo y mínimo identificados en *E. coli* (0.3779 valor máximo y 0,1501 valor mínimo).

Para el caso de Spot42 (Figura 5.16), se puede observar que este sRNA está dividido en 5 zonas entre las cuales 4 realizan interacciones (S42A1, S42A2, S42A3, S42C) y 1 zona en las que no existe ningún tipo de interacción sRNA-mRNA (S42B). En el caso de las zonas que existen interacciones sRNA y mRNA, en la zona S42A3 existe la mayor cantidad de interacciones registradas para Spot42 (10 interacciones) de las cuales todas son represiones, en la zona S42A1 y S42A2 para cada uno de estos casos existen un total de 3 interacciones sRNA-mRNA para cada zona respectivamente, siendo todas estas de tipo represión y en la zona S42C solo se registra 1 interacción sRNA-mRNA que al igual que los

casos anteriormente mencionados también es una represión. Cabe mencionar que existe 4 mensajeros que interaccionan con 2 zonas a la vez (2 interacciones mRNA-S42A1; S42A2 y 2 interacciones mRNA-S42A2; S42A3) y solo 1 caso en el que el mensajero interacciona con 3 zonas a la vez (galK con S42A1, S42A2 y S42A3).

5.4.1. Funciones Biológicas de los genes que interactúan con los módulos de los sRNAs.

Para determinar si los genes que reconocen un determinado módulo al interactuar con un sRNA se encuentran relacionados funcionalmente se decidió comparar conjunto de procesos biológicos que regulan. Se aplicó la herramienta Blast Koala tanto para todos los genes que son regulados por un mismo sRNA como para los genes que interactúan con un mismo módulo. Los resultados del primer análisis muestran que principalmente los sRNA regulan genes relacionados al procesamiento de la información ambiental, seguido de procesos celulares y procesamiento de la información genética (Figura 5.17).

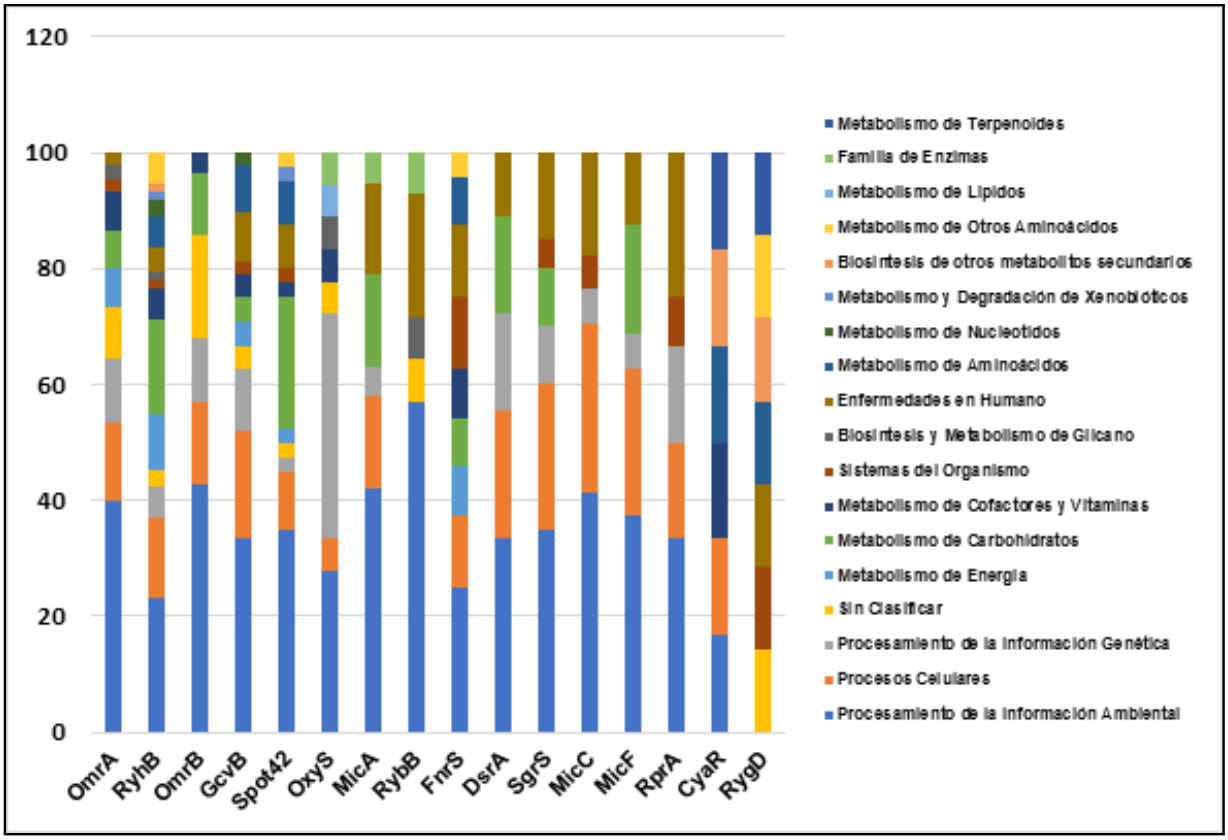


Figura 5.17: **Procesos biológicos asociados a los genes regulados por sRNA** La figura muestra la frecuencia de los diferentes procesos biológicos en los genes regulados por sRNA que reconocen múltiples blancos. Como cada sRNA reconoce un número diferente de genes, la frecuencia de ocurrencia del proceso biológicos fue normalizada por el número total de procesos identificados en los genes regulados por un mismo sRNA.

Este análisis se repitió para cada uno de los genes que son reconocidos por un mismo módulo. Desde el punto de vista funcional la mayor parte de módulos interactúa con genes que participan en diferentes procesos biológicos. Solo los procesos biológicos relacionados a Metabolismo de Lípidos, Biosíntesis de Metabolitos secundarios, y Metabolismo de terpenos y poliketidos interactúan solo con un módulo, (OxyS-C2, y RygD-D) en el sRNA blanco. En el caso de genes relacionados al Metabolismo y Biosíntesis de Glicanos, el 75 % de los módulos que regula esta función lo hace en forma exclusiva. Desde la perspectiva de los módulos, solo Spot42C, DsrA-B3 y RygD-F y H, interactúan exclusivamente con mRNAs que pertenecen a una sola función biológica. Nuestros resultados también muestran que en casi todos los sRNA existen módulos que interactúan específicamente con un conjunto de genes relacionados a un solo tipo de función biológica. En el caso de OmrA, el módulo A1 interactúa específicamente con 3 funciones biológicas (Familia

de enzimas, Sistemas del organismo, y Enfermedades humanas), mientras que el módulo A2 interactúa con genes relacionados al Metabolismo y síntesis de glicanos. Para RyhB, el módulo A1 interactúa exclusivamente con dos procesos (Metabolismos y síntesis de glicanos, y Metabolismo y degradación de xenobióticos), el módulo A2 con Biosíntesis de metabolitos secundarios, y el módulo A3 con sistemas del organismo. En el caso de OmrB, el Módulo A1 interactúa exclusivamente con mRNAs relacionados al Metabolismo de Carbohidratos (Figura 5.18).

Los resultados anteriores muestran que en los sRNA existen módulos que reconocen específicamente mRNA asociadas a determinadas funciones biológicas. Para determinar si estos módulos también estan asociados a genes que presentan un determinado sesgo codogénico, se analizó para cada sRNA si entre los mRNA que reconocen los diferentes módulos existían valores de adaptación al uso codogénico diferentes o si presentaban una adaptación diferente al pool de tRNA. El análisis mostró que solo para RyhB y GcvB se identificaron diferencias estadísticamente significativas en los valores de adaptación al uso de codones (CAI), mientras que en la adaptación a pool de tRNA solo se identificaron diferencias estadísticamente significativas para RyhB y OxyS (Figura 5.19).

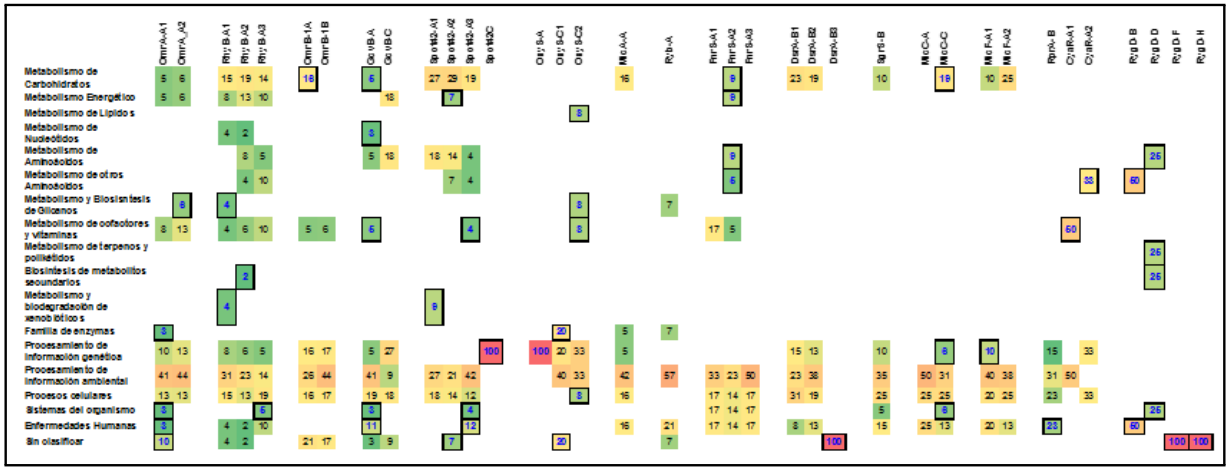


Figura 5.18: **Funciones biológicas asociadas a los módulos**La figura muestra el tipo de función biológica presente en los mRNA que interactúan con los diversos módulos identificados. Los valores numéricos representan el porcentaje de mRNA que interactúan con cada módulo y que codifican para proteínas asociadas a la función biológica. Los valores fueron normalizados respecto al número total de funciones biológicas asociadas a cada módulo del sRNA. Las casillas enmarcadas en línea negra y destacadas en letras de color azul representan los procesos biológicos presentes en un solo módulo de cada sRNA.

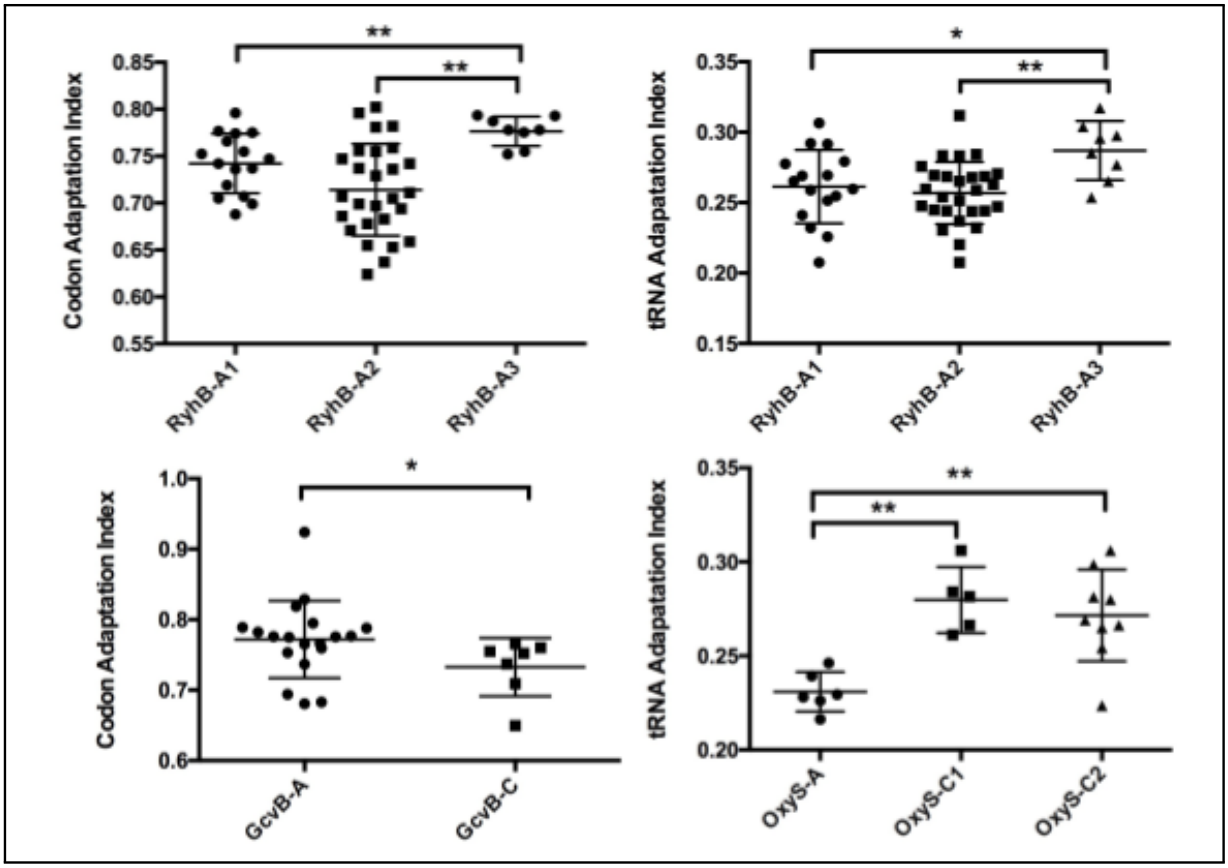


Figura 5.19: **Sesgo en el uso de codones en los módulos de los sRNA**La figura muestra los valores de CAI (Codon Adaptation Index) y tAI (tRNA Adaptation Index) de los mRNA que interactúan con los módulos de RyhB, OxyS y GcvB. Se compararon las diferencias estadísticas entre grupos usando el test Mann-Whitney no pareado.

6. Discusión

6.1. Discusión

La traducción de los mensajeros a proteínas es un complejo proceso biológico en el cual los sRNA actúan como reguladores afectando este proceso de manera positiva o negativa dependiendo el tipo de interacción que conforme. La traduccción tambien es afectada por la composisción codogénica de la región codificante. Genes que presentan codones que poseen una mayor disponibilidad de tRNA son traducidos mas eficientementes que aquellos compuestos genes que poseen codones de baja frecuencia o con baja disponibilidad de tRNAs que los decodifican. Para predecir la eficiencia de la traducción de un gen existen 2 índices el CAI y tAI, siendo el CAI el que evalua la adaptación codogénica y tAI evalua los tRNAs disponibles para cada codón. Con la finalidad de establecer una relación entre el tAI y CAI de un gen con su regulación por sRNA se estudiaron los CAI y tAI de los mensajeros que interaccionan con sRNA y se compararon con los valores de CAI y tAI de los mensajeros que no interactuaban con sRNAs con la finalidad de establecer si la interacción sRNA-mRNA esta directamente relacionada con los valores de estos ícomo ndices.

Para demostrar esto se planteó la siguiente hipótesis: En Escherichia coli K12 existe una red de regulación mediada por sRNA, en la cual los genes regulados por sRNA presentan un patrón de codones y adaptación al pool de tRNA diferente al presente en los genes no regulados por sRNAs. Con la finalidad de responder a esta hipótesis se plantearon los siguientes objetivos:

1. Identificación mediante búsqueda en las bases de datos las parejas de sRNAs y mRNA en *Escherichia coli* K-12 MG1655.

2. Identificación y predicción de las regiones de interacción entre mRNA y sRNA.

3. Caracterización de la adaptación al repertorio de tRNA de los genes regulados por sRNA y de sus regiones de interacción.

4. Generación de un mapa de interacciones sRNA-mRNA que incluya la adaptación del repertorio de tRNA y uso codogénico del gen completo y de la región de interacción.

Respecto al OE1 a partir de los datos obtenidos en las bases de datos se identificaron 108 sRNAs en *E. coli*, de los cuales 96 presentaron interacciones sRNA-mRNA con 304 mensajeros y 12 no registraron ningún tipo de interacción. Cabe destacar que en este punto se identificó una gran cantidad de interacciones sRNA-mRNA presentes bibliografía en las cuales se conoce el sRNA y su mRNA blanco, pero no la zona donde ocurre esta interacción. Para lograr esto se utilizaron 2 bases de datos sRNATarBase y BSRD. sRNATarBase es una base de datos que contiene la información de interacciones sRNA-mRNA de 17 organismos bacterianos [5], donde se encuentra la información de *E. coli* K-12 MG1655, de las cuales se encuentran 230 interacciones de las cuales algunas de estas se les conoce el sRNA y su blanco, pero no la zona en donde ocurre esta interacción. Por otra parte BSRD es una base de datos que contiene la información de 783 especies bacterianas que dan cuenta de 957 interacciones[15]. Para *E. coli* K-12 MG1655, en esta base de datos están contenidos todos los sRNAs identificados para este organismo, junto con esto también se encuentran las interacciones extraídas de sRNATarBase, e interacciones predichas con la identificación del software que se utilizó, como también se incluye la energía de la posible interacción. Por otra parte, la base de datos creada en este trabajo (DB_Ecoli_V3) es una base de datos que solo se encarga de almacenar la información de las interacciones sRNA-mRNA de *E. coli* K-12 MG1655, contiene información relevante como la energía de interacción, el método que se utilizó para describir la interacción, el efecto, entre otros. En comparación con sRNATarBase y BSRD, DB_Ecoli_V3 es más completa en relación a la información de las interacciones de *E. coli* K-12 MG1655 ya que esta base contiene todos los datos relevantes de los sRNAs (nombre sinónimo y división de zonas del sRNA), como también de los mensajeros (secuencia del mensajero con las zonas UTR, Valores de CAI y TAI).

En el caso del OE2 a partir de los resultados obtenidos anteriormente se procedió a caracterizar las interacciones que faltaban y a predecir las interacciones para los sRNAs que faltaban, esto llevó a poder completar de esta manera a todos los sRNAs que están presentes *E. coli*, por lo tanto, quedó un total de 435 interacciones con un total de 326 mensajeros, divididos en 322 interacciones predichas (equivalente a un 59,31 %) y a 113 interacciones empíricas (equivalentes a un 40,68 %). Cabe mencionar que la identificación de los mensajeros con sRNAs llevó al proceso de identificación de zona de interacción en los sRNAs, como lo estableció Taylor B. Updegrove y otros en su artículo [35], que a diferencia de su investigación se utilizaron todos los mensajeros que interaccionan con un sRNA en específico, dando de esta manera para el caso de Spot42 un resultado muy similar al que ellos obtuvieron por medio de su investigación.

La metodología creada por Taylor B. en su artículo[35], la cual utiliza la energía libre de Gibbs por nucleótido para obtener un patrón de interacción en el sRNA. Esta metodología se utilizó en investigaciones como la que estudia al sRNA GlmZ el cual es dependiente de Hfq y modula la traducción de glmS [38] y en otro estudio para comprobar la dependencia de la proteína ProQ para el sRNA RaiZ en Salmonella Typhimurium[29]. En comparación la metodología desarrolla por Ávalos F, Tello M y otros, utiliza un script el cual comprende el largo del sRNA y marca con 1 y con 0 en donde hay interacción sRNA-mRNA, la ventaja de este método es que utiliza todos los mensajeros que interactúan con el sRNA, siendo el único parámetro para crear el patrón de interacción la posición de la interacción en el sRNA, esto después se gráfica y se procede a identificar las zonas de interacción en el sRNA. La ventaja de esta metodología es la creación de la red de interacción a partir del gráfico en la cual se puede observar como el mensajero interactúa con el sRNA. La desventaja de esta metodología en comparación con la desarrollada por Updegrove es la cantidad de pasos que fueron necesarios para llevar a un resultado similar.

Para el caso del OE3 se realizó análisis estadísticos con la finalidad de encontrar diferencias entre los valores de CAI y tAI para los genes regulados y no regulados por sRNAs, lo cual nos mostró para el caso de este tipo de comparación ninguna diferencia significativa, sin embargo, al comprar las sitios de interacción que son regulados por sRNAs se logran encontrar diferencias en esto sobre todo en las regiones 5'CDS y 3'CDS para el caso del valor de CAI, es decir que estas regiones tienen la tendencia de utilizar codones que se encuentran con mayor frecuencia a lo largo del genoma, para el caso del valor de tAI se encontró que la región de interacción presenta valores de este índice menores, lo que sugiere que los codones utilizados para formar interacción son de un alto uso pero que presenta un poca cantidad de tRNA para realizar la traducción.

En el proceso de traducción los ribosomas son los encargados de trasformar los codones de la secuencia codificante en aminoácidos. Durante este proceso la velocidad de traducción es variable dependiendo del codón que esté presente en la secuencia codificante, es decir, si el codón que se estÃ¡ traduciendo tiene un alto uso y a la vez una alta disponibilidad de tRNA la velocidad de traducción es alta, en el caso de que el codón sea de alto uso y a la vez tenga una baja disponibilidad de tRNA la velocidad de traducción será más lenta. La forma de evaluar esto es por medio de los índices de CAI y tAI y de esta manera se podría identificar las posibles zonas del área codificante en donde la traducción avanza más lento. La identificación de zonas de baja velocidad de traducción podría ayudar a encontrar nuevos sRNAs que interacciones con esta, ya que los sRNAs modulan la traducción en zonas del área codificante que presente codones de alto uso y baja disponibilidad de tRNAs ya que al presentar el ribosoma una menor velocidad de traducción le entrega una mayor

cantidad de tiempo al sRNA para formar la interacción sRNA-mRNA y de esta manera lograr reprimir el proceso de traducción de este. Este principio puede ser utilizado para identificar nuevas interacciones sRNA-mRNA en otros organismos bacterianos como por ejemplo bacterias que afectan a especies de pescados (como, por ejemplo: *Salmo Salar, Oncorhynchuis mykiss*, entre otras especies acuáticas), como también en otras cepas de *E. coli*, como también en distintas cepas de *Salmonella* o cualquier organismo bacterianas que presenten en su genoma RNA no codificante (específicamente sRNAs).

En el caso del OE4 para expresar los resultados obtenidos se creó una red de interacción en donde se pudieron identificar todas las interacciones registradas para *E. coli*, junto con esto se crearon redes de interacción para cada sRNA de manera individual, lo que ayudó a identificar la organización de las interacciones por medio del CAI y tAI que presentan los mensajeros que interaccionan con estos sRNAs en particular y como se distribuye el CAI y el tAI de los mensajeros a lo largo del sRNA y como existen zonas los sRNAs que tienen preferencias específicas de interacción con mensajeros que presentan ciertos valores de tAI y CAI. Este proceso se podría expandir hacia otros organismos bacterianos con la finalidad de comprender como se comportan los valores de CAI y tAI de los mensajeros que interaccionan con sRNAs de estos organismos en específico.

Para el estudio de interacciones sRNA-mRNA la creación de redes de interacción es importante, ya que por este método de representación gráfica se puede tener una visión general la forma en que los sRNAs interactúan con mensajeros. Varios autores han logrado representar las interacciones sRNA-mRNA de manera exitosa por medio del uso de redes de interacción sRNA-mRNA, tomemos por ejemplo el caso de los autores Stortz et. al [31] en el cual representaron las interacciones de 20 sRNAs (lo cual equivale a 18.5 % de los sRNAs identificados) con 99 mensajeros, en el cual solo muestra la sRNA con su blanco, no muestran información como el efecto de dicha interacción, si la interacción es empírica o predicha, etc. En el caso de la red de interacción elaborada por Ávalos, F., Tello, M, la red de interacción elaborada por ellos muestra a los 108 sRNAs con sus respectivos blancos, el efecto de la interacción, si la interacción que se describe es empírica o predicha, junto con esto también incluye los valores de CAI y tAI representados por color y tamaño del nodo del mensajero respectivamente. En este aspecto la red de interacción descrita por los dos autores anteriores muestra mucha más información que la primera red que describía la interacción sRNA-mRNA en *E. coli*, abriendo una nueva ventana hacia la comprensión de las interacciones sRNA-mRNA.

7. Conclusión

7.1. Conclusión

A partir de la información recopilada en la Tabla 5.7, sugiere el rechazo de la hipótesis planteada, sin embargo, los nuevos resultados obtenidos en esta memoria de título permiten sugerir que en *E. coli*, el mecanismo de regulación por RNA pequeño, ocurre en zonas del área codificante que contienen codones de alto uso y baja disponibilidad de tRNA, lo que implica que la regulación ocurre en regiones en donde la velocidad de avance del ribosoma es menor.

Bibliografía

[1] Shoshy Altuvia, Dalit Weinstein-Fischer, Aixia Zhang, Lisa Postow, and Gisela Storz. A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. *Cell*, 90(1):43–53, 1997.

[2] Gary D. Bader, Doron Betel, and Christopher W.V. Hogue. BIND: The Biomolecular Interaction Network Database. *Nucleic Acids Research*, 31(1):248–250, 2003.

[3] Marie Bouvier, Cynthia M. Sharma, Franziska Mika, Knud H. Nierhaus, and Jörg Vogel. Small RNA Binding to 5 mRNA Coding Region Inhibits Translational Initiation. *Molecular Cell*, 32(6):827–837, 2008.

[4] Jader M. Caldonazzo Garbelini, André Y. Kashiwabara, and Danilo S. Sanches. Sequence motif finder using memetic algorithm. *BMC Bioinformatics*, 19(1):4, 2018.

[5] Yuan Cao, Jiayao Wu, Qian Liu, Yalin Zhao, Xiaomin Ying, Lei Cha, Ligui Wang, and Wuju Li. sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments. *RNA (New York, N.Y.)*, 16(11):2051–7, 2010.

[6] Ronan K Carroll, Andy Weiss, William H Broach, Richard E Wiemels, Austin B Mogen, Kelly C Rice, and N Shaw. Analysis of Regulatory or Small RNA Gene Expression in Staphylococcus aureus. 7(1):1–13, 2016.

[7] Fabienne F V Chevance, Soazig Le Guyon, and Kelly T. Hughes. The Effects of Codon Context on In Vivo Translation Speed. *PLoS Genetics*, 10(6), 2014.

[8] Mario dos Reis, Renos Savva, and Lorenz Wernisch. Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Research*, 32(17):5036–5044, 2004.

[9] Thomas A Geissmann and Danièle Touati. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *The EMBO journal*, 23(2):396–405, 2004.

[10] Ruth Hershberg, Shoshy Altuvia, and Hanah Margalit. A survey of small RNA-encoding genes in Escherichia coli. *Nucleic Acids Research*, 31(7):1813–1820, 2003.

[11] Intisar Husain, Aziz Sancar, and Chapel Hill. The codon adaptation index - a measure of directional synonymous codon usage bias and its potential application. 15(3):1281–1295, 1987.

[12] Minoru Kanehisa, Yoko Sato, and Kanae Morishima. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4):726–731, 2016.

[13] Daniel Lai and Irmtraud M Meyer. A comprehensive comparison of general RNA-RNA interaction prediction methods. *Nucleic acids research*, page 1477, 2015.

[14] R a Lease and M Belfort. A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures. *Proceedings of the National Academy of Sciences of the United States of America*, 97(18):9919–9924, 2000.

[15] Leo Lei Li. Bacterial Small Regulatory RNA Database.

[16] Jane M. Liu and Andrew Camilli. A broadening world of bacterial small RNAs. *Current Opinion in Microbiology*, 13(1):18–23, 2010.

[17] Jonathan Livny, Anja Brencic, Stephen Lory, and Matthew K. Waldor. Identification of 17 Pseudomonas aeruginosa sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Research*, 34(12):3484–3493, 2006.

[18] Eric Massé and Susan Gottesman. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7):4620–4625, 2002.

[19] Thorleif Moller, Thomas Franch, Peter Hojrup, Douglas R. Keene, Hans Peter Bchinger, Richard G. Brennan, and Poul Valentin-Hansen. Hfq: A Bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular Cell*, 9(1):23–30, 2002.

[20] Mor Nitzan, Rotem Rehani, and Hanah Margalit. Integration of Bacterial Small RNAs in Regulatory Networks. *Annual Review of Biophysics*, 46(1):131–148, 2017.

[21] Eva Maria Novoa and Llus Ribas de Pouplana. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends in Genetics*, 28(11):574–581, 2012.

[22] Keiichiro Ono, Barry Demchak, and Trey Ideker. Cytoscape tools for the web age: D3.js and Cytoscape.js exporters. *F1000Research*, 3:143, 2014.

[23] Kai Papenfort, Verena Pfeiffer, Sacha Lucchini, Avinash Sonawane, J. C D Hinton, and Jörg Vogel. Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Molecular Microbiology*, 68(4):890–906, 2008.

[24] S. Peri. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Research*, 32(90001):497D–501, 2004.

[25] Monica Riley, Takashi Abe, Martha B. Arnaud, Mary K B Berlyn, Frederick R. Blattner, Roy R. Chaudhuri, Jeremy D. Glasner, Takashi Horiuchi, Ingrid M. Keseler, Takehide Kosuge, Hirotada Mori, Nicole T. Perna, Guy Plunkett, Kenneth E. Rudd, Margrethe H. Serres, Gavin H. Thomas, Nicholas R. Thomson, David Wishart, and Barry L. Wanner. Escherichia coli K-12: A cooperatively developed annotation snapshot - 2005. *Nucleic Acids Research*, 34(1):1–9, 2006.

[26] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. (Karp 2001):2498–2504, 2003.

[27] Cynthia M. Sharma, Fabien Darfeuille, Titia H. Plantinga, and Jörg Vogel. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes and Development*, 21(21):2804–2817, 2007.

[28] D D Sledjeski, a Gupta, and S Gottesman. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *The EMBO journal*, 15(15):3993–4000, 1996.

[29] Alexandre Smirnov, Chuan Wang, Lisa L Drewry, and Jörg Vogel. Molecular mechanism of mRNA repression in trans by a ProQ dependent small RNA. *The EMBO Journal*, 36(8):1029–1045, 2017.

[30] Chunxu Song, Menno van der Voort, Judith van de Mortel, Karl A. Hassan, Liam D H Elbourne, Ian T. Paulsen, Joyce E. Loper, and Jos M. Raaijmakers. The Rsm regulon of

plant growth-promoting Pseudomonas fluorescens SS101: Role of small RNAs in regulation of lipopeptide biosynthesis. *Microbial Biotechnology*, 8(2):296–310, 2015.

[31] Gisela Storz, Jörg Vogel, and Karen M. Wassarman. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, 43(6):880–891, 2011.

[32] Gang Su, John H. Morris, Barry Demchak, and Gary D. Bader. *Biological Network Exploration with Cytoscape 3*, volume 72. 2011.

[33] Mario Tello, Francisco Vergara, and Eugenio Spencer. Genomic adaptation of the ISA virus to Salmo salar codon usage. *Virology journal*, 10(1):223, 2013.

[34] Tamir Tuller, Asaf Carmi, Kalin Vestsigian, Sivan Navon, Yuval Dorfan, John Zaborske, Tao Pan, Orna Dahan, Itay Furman, and Yitzhak Pilpel. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2):344–354, 2010.

[35] Taylor B Updegrove, Svetlana A Shabalina, and Gisela Storz. How do base-pairing small RNAs evolve ? (March):379–391, 2015.

[36] Sandra C. Viegas, Verena Pfeiffer, Alexandra Sittka, Ines J. Silva, Jörg Vogel, and Cecília M. Arraiano. Characterization of the role of ribonucleases in Salmonella small RNA decay. *Nucleic Acids Research*, 35(22):7651–7664, 2007.

[37] Sandra Cristina Viegas and Cecília Maria Arraiano. Regulating the regulators: How ribonucleases dictate the rules in the control of small non-coding RNAs. *RNA biology*, 5(4):230–243, 2008.

[38] James D. Watson, Tania A. Baker, Stephen P. Bell, Alexander Gann, Michael Levine, and Richard Losik. *Molecular biology of the gene*. 2013.

[39] Nikolai Windbichler, Frederike von Pelchrzim, Oliver Mayer, Edina Csaszar, and Renée Schroeder. Isolation of small RNA-binding proteins from E. coli: evidence for frequent interaction of RNAs with RNA polymerase. *RNA biology*, 5(1):30–40, 2008.

ANEXOS

A. Resumen Congreso Somich Noviembre 2016

Caracterización de red de interacciones sRNA-mRNA en Escherichia coli K12 MG1655

Autores: Felipe Ávalos Díaz[1], Juan Pablo Maldonado[2], Mario Tello Reyes[1], Omar Orellana[2]

1: Laboratorio de Metagenómica Bacteriana, Centro de Biotecnología Acuícola, Facultad de Química y Biología, Universidad de Santiago de Chile.

2: Laboratorio de Biología Molecular Bacteriana, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile.

Introducción: Los sRNA son un tipo de ncRNA (RNA no codificante) en donde el tamaño característico va desde los 50 a 500 nucleótidos de longitud y por lo general son dependientes de la proteína Hfq. El efecto de la unión sRNA-mRNA puede ser de inducción y represión dependiendo de la zona en donde el sRNA está interactuando con el mRNA. Un mismo sRNA pude ser inductor de la traducción de un mRNA específico y la represión de la traducción de otro mRNA de forma simultánea. En el caso de los mRNA están divididos en zonas las cuales son 5'UTR, Zona Codificante y 3'UTR, pero en el caso de los sRNA no se ha determinado una forma de dividir su estructura en zonas de interés, lo cual nos lleva a preguntarnos si existe algún modo de identificar zonas en el sRNA.

Hipótesis: Los sRNAs están divididos en zonas que interactúan específicamente con un conjunto de mRNAs con la finalidad de regular la traducción de este. Metodología: Se recopilo la información obtenidas en las bases de datos sRNATarBAse y la base de datos BSRD, la cual posteriormente se complementó con las predicciones de IntaRNA para las

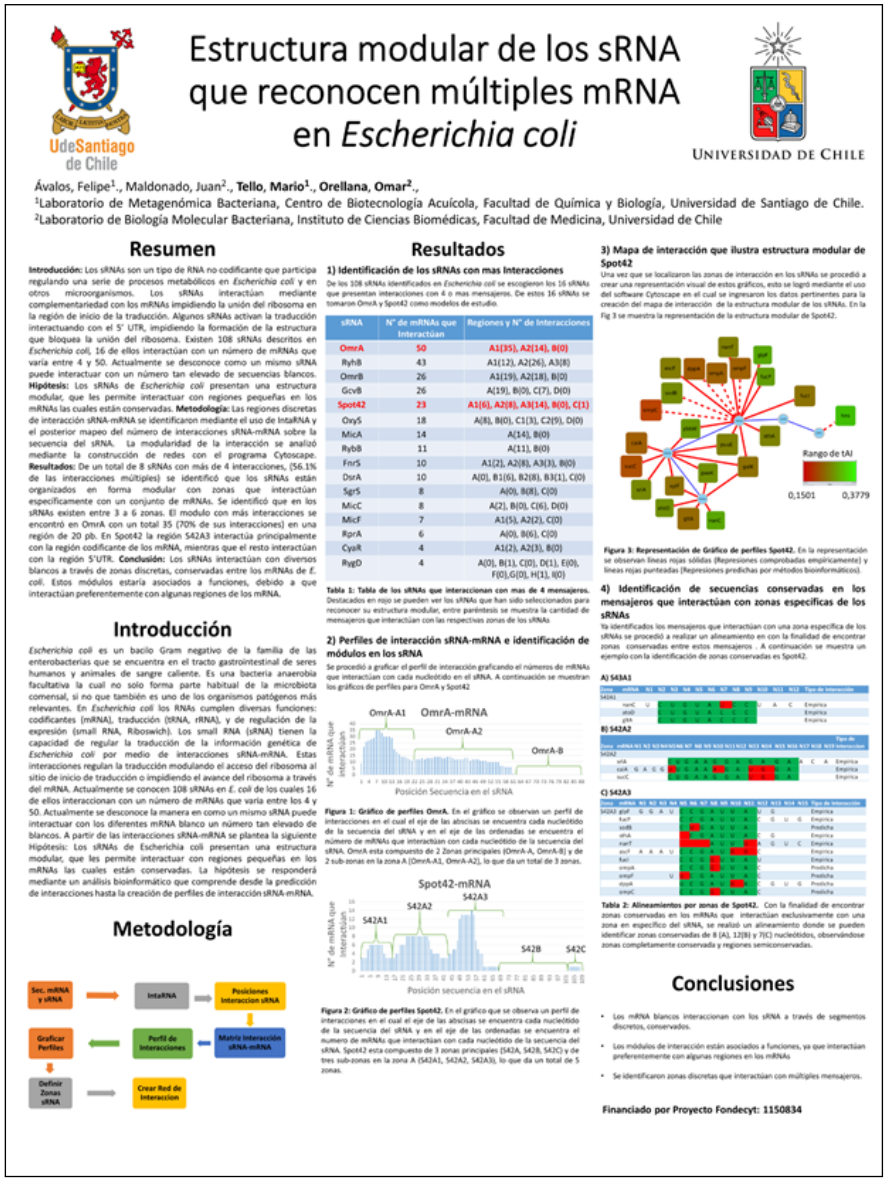
interacciones que no poseían lugar de interaccion en el sRNA. Posteriormente se tomaron los datos de las posiciones de interaccion con el sRNA para ejecutar un script que crea una matriz con las posiciones de interaccion, posteriormente la matriz se graficó por medio de Microsoft Excel para posteriormente proceder a identificar las zonas del sRNA.

Resultados: Se identificó que los sRNAs están organizados en forma modular con zonas que interactúan específicamente con un conjunto de mRNAs. Se identificó que en los sRNAs existen entre 2 a 6 zonas. De un total de 18 sRNAs analizados que dan cuenta de 167 interacciones equivalentes a un 38,30 % del total de interacciones identificadas, se identificó que los sRNAs con tres zonas son los más comunes de las 18 interacciones estudiadas con un total de 8 interacciones que equivale a un 44.4 % del total. En el caso de SPOT42 se identificaron 5 zonas de la cual la zona S42A3 interactúa principalmente con la región codificante de los mensajeros, mientras que en el resto de las zonas interactúan principalmente con la región 5'UTR.

Conclusión: Por medio del estudio de los resultados se lograron identificar zonas en los sRNA que son altamente utilizadas para las interacciones con mRNA, como también se han identificado zonas en las cuales no hay existencia de interacciones con mRNA y que posiblemente tengan un rol estructural para la determinación de la estructura del sRNA.

Financiamiento: Proyecto Fondecyt 1150834

B. Poster Presentado Somich 2016



UdeSantiago de Chile

Estructura modular de los sRNA que reconocen múltiples mRNA en *Escherichia coli*

UNIVERSIDAD DE CHILE

Ávalos, Felipe[1]., Maldonado, Juan[2]., **Tello, Mario[1]., Orellana, Omar[2].**,
[1]Laboratorio de Metagenómica Bacteriana, Centro de Biotecnología Acuícola, Facultad de Química y Biología, Universidad de Santiago de Chile.
[2]Laboratorio de Biología Molecular Bacteriana, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile

Resumen

Introducción: Los sRNAs son un tipo de RNA no codificante que participa regulando una serie de procesos metabólicos en *Escherichia coli* y en otros microorganismos. Los sRNAs interactúan mediante complementariedad con los mRNAs impidiendo la unión del ribosoma en la región de inicio de la traducción. Algunos sRNAs activan la traducción interactuando con el 5' UTR, impidiendo la formación de la estructura que bloquea la unión del ribosoma. Existen 108 sRNAs descritos en *Escherichia coli*, 16 de ellos interactúan con un número de mRNAs que varía entre 4 y 50. Actualmente se desconoce como un mismo sRNA puede interactuar con un número tan elevado de secuencias blancos. **Hipótesis:** Los sRNAs de *Escherichia coli* presentan una estructura modular, que les permite interactuar con regiones pequeñas en los mRNAs las cuales están conservadas. **Metodología:** Las regiones discretas de interacción sRNA-mRNA se identificaron mediante el uso de IntaRNA y el posterior mapeo del número de interacciones sRNA-mRNA sobre la secuencia del sRNA. La modularidad de la interacción se analizó mediante la construcción de redes con el programa Cytoscape. **Resultados:** De un total de 8 sRNAs con más de 4 interacciones, (56.1% de las interacciones múltiples) se identificó que los sRNAs están organizados en forma modular con zonas que interactúan específicamente con un conjunto de mRNAs. Se identificó que en los sRNAs existen entre 3 a 6 zonas. El modulo con más interacciones se encontró en OmrA con un total 35 (70% de sus interacciones) en una región de 20 pb. En Spot42 la región S42A3 interactúa principalmente con la región codificante de los mRNA, mientras que el resto interactúan con la región 5'UTR. **Conclusión:** Los sRNAs interactúan con diversos blancos a través de zonas discretas, conservadas entre los mRNAs de *E. coli*. Estos módulos estaría asociados a funciones, debido a que interactúan preferentemente con algunas regiones de los mRNA.

Introducción

Escherichia coli es un bacilo Gram negativo de la familia de las enterobacterias que se encuentra en el tracto gastrointestinal de seres humanos y animales de sangre caliente. Es una bacteria anaerobia facultativa la cual no solo forma parte habitual de la microbiota comensal, si no que también es uno de los organismos patógenos más relevantes. En *Escherichia coli* los RNAs cumplen diversas funciones: codificantes (mRNA), traducción (tRNA, rRNA), y de regulación de la expresión (small RNA, Riboswich). Los small RNA (sRNA) tienen la capacidad de regular la traducción de la información genética de *Escherichia coli* por medio de interacciones sRNA-mRNA. Estas interacciones regulan la traducción modulando el acceso del ribosoma al sitio de inicio de traducción o impidiendo el avance del ribosoma a través del mRNA. Actualmente se conocen 108 sRNAs en *E. coli* de los cuales 16 de ellos interaccionan con un número de mRNAs que varía entre los 4 y 50. Actualmente se desconoce la manera en como un mismo sRNA puede interactuar con los diferentes mRNA blanco un número tan elevado de blancos. A partir de las interacciones sRNA-mRNA se plantea la siguiente Hipótesis: Los sRNAs de Escherichia coli presentan una estructura modular, que les permite interactuar con regiones pequeñas en los mRNAs las cuales están conservadas. La hipótesis se responderá mediante un análisis bioinformático que comprende desde la predicción de interacciones hasta la creación de perfiles de interacción sRNA-mRNA.

Metodología

Resultados

1) Identificación de los sRNAs con mas Interacciones

De los 108 sRNAs identificados en *Escherichia coli* se escogieron los 16 sRNAs que presentan interacciones con 4 o mas mensajeros. De estos 16 sRNAs se tomaron OmrA y Spot42 como modelos de estudio.

sRNA	N° de mRNAs que Interactúan	Regiones y N° de Interacciones
OmrA	50	A1(35), A2(14), B(0)
RyhB	43	A1(12), A2(26), A3(8)
OmrB	26	A3(19), A2(18), B(0)
GcvB	26	A(19), B(0), C(7), D(0)
Spot42	23	A1(6), A2(8), A3(14), B(0), C(1)
OxyS	18	A(8), B(0), C1(3), C2(9), D(0)
MicA	14	A(14), B(0)
RybB	11	A(11), B(0)
FnrS	10	A1(2), A2(8), A3(3), B(0)
DsrA	10	A(0), B1(6), B2(8), B3(1), C(0)
SgrS	8	A(0), B(8), C(0)
MicC	8	A(2), B(0), C(6), D(0)
MicF	7	A1(5), A2(2), C(0)
RprA	6	A(0), B(6), C(0)
CyaR	4	A1(2), A2(3), B(0)
RyeD	4	A(0), B(1), C(0), D(1), E(0), F(0),G(0), H(1), I(0)

Tabla 1: Tabla de los sRNAs que interaccionan con mas de 4 mensajeros. Destacados en rojo se pueden ver los sRNAs que han sido seleccionados para reconocer su estructura modular, entre paréntesis se muestra la cantidad de mensajeros que interactúan con las respectivas zonas de los sRNAs

2) Perfiles de interacción sRNA-mRNA e identificación de módulos en los sRNA

Se procedió a graficar el perfil de interacción graficando el números de mRNAs que interactúan con cada nucleótido en el sRNA. A continuación se muestran los gráficos de perfiles para OmrA y Spot42

Figura 1: Gráfico de perfiles OmrA. En el gráfico se observan un perfil de interacciones en el cual el eje de las abscisas se encuentra cada nucleótido de la secuencia del sRNA y en el eje de las ordenadas se encuentra el número de mRNAs que interactúan con cada nucleótido de la secuencia del sRNA. OmrA esta compuesto de 2 Zonas principales (OmrA-A, OmrA-B) y de 2 sub-zonas en la zona A (OmrA-A1, OmrA-A2), lo que da un total de 3 zonas.

Figura 2: Gráfico de perfiles Spot42. En el gráfico que se observa un perfil de interacciones en el cual el eje de las abscisas se encuentra cada nucleótido de la secuencia del sRNA y en el eje de las ordenadas se encuentra el numero de mRNAs que interactúan con cada nucleótido de la secuencia del sRNA. Spot42 esta compuesto de 3 zonas principales (S42A, S42B, S42C) y de tres sub-zonas en la zona A (S42A1, S42A2, S42A3), lo que da un total de 5 zonas.

3) Mapa de interacción que ilustra estructura modular de Spot42

Una vez que se localizaron las zonas de interacción en los sRNAs se procedió a crear una representación visual de estos gráficos, esto se logró mediante el uso del software Cytoscape en el cual se ingresaron los datos pertinentes para la creación del mapa de interacción de la estructura modular de los sRNAs. En la Fig 3 se muestra la representación de la estructura modular de Spot42.

Figura 3: Representación de Gráfico de perfiles Spot42. En la representación se observan líneas rojas sólidas (Represiones comprobadas empíricamente) y líneas rojas punteadas (Represiones predichas por métodos bioinformáticos).

4) Identificación de secuencias conservadas en los mensajeros que interactúan con zonas especificas de los sRNAs

Ya identificados los mensajeros que interactúan con una zona especifica de los sRNAs se procedió a realizar un alineamiento en con la finalidad de encontrar zonas conservadas entre estos mensajeros . A continuación se muestra un ejemplo con la identificación de zonas conservadas es Spot42.

A) S43A1

Zona	mRNA	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	Tipo de Interacción
S42A1														
narC	U	C	U	G	U	A		C		G	U	A	C	Empirica
atoD		C	U	G	U	A		C		G				Empirica
gltA		C	U	G	U	A		C		G				Empirica

B) S42A2

Zona	mRNA	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	Tipo de Interacción
S42A2																					
srlA				U	G	A	A				A	U	A	C	A		A		C	A	Empirica
cusA	G	A	G	G	C	G	A	A		C		A		A		C		C		A	Empirica
ssuC					G	A	A	C		G		A				C		C		A	Empirica

C) S42A3

Zona	mRNA	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	Tipo de Interacción
S42A3	gyaF	G	G	A	U	C	G	A	U	A	U	A	U	G			Empirica
kaiA						C	G	A	U	U	A	C		G	U	G	Empirica
sodB						C	G	A	U	U	A	C					Predicha
ulaA						C	G	A	U	U	A	C	G				Empirica
narX						C	G	A	U	U	A	C	G	U	C		Empirica
ascF	A	A	A	U		C		A	U		A	C					Empirica
faeJ						C	C	G	U		A	C	U				Empirica
ompA						C	C	G	U		A	C					Predicha
ompF		U		C		C	G	A	U		A	C					Predicha
dppA				C		C	G	A	U		C	C	G	U	G		Predicha
ompC				C		C	G	A	U		A	C					Predicha

Tabla 2: Alineamientos por zonas de Spot42. Con la finalidad de encontrar zonas conservadas en los mRNAs que interactúan exclusivamente con una zona en especifico del sRNA, se realizó un alineamiento donde se pueden identificar zonas conservadas de 8 (A), 12(B) y 7(C) nucleótidos, observándose zonas completamente conservada y regiones semiconservadas.

Conclusiones

- Los mRNA blancos interaccionan con los sRNA a través de segmentos discretos, conservados.
- Los módulos de interacción están asociados a funciones, ya que interactúan preferentemente con algunas regiones en los mRNAs
- Se identificaron zonas discretas que interactúan con múltiples mensajeros.

Financiado por Proyecto Fondecyt: 1150834

Figura B.1: Poster Presentado en la Somich, Valdivia 2016

C. Gráficos, Estructuras y Representación sRNAs

C.0.1. OxyS

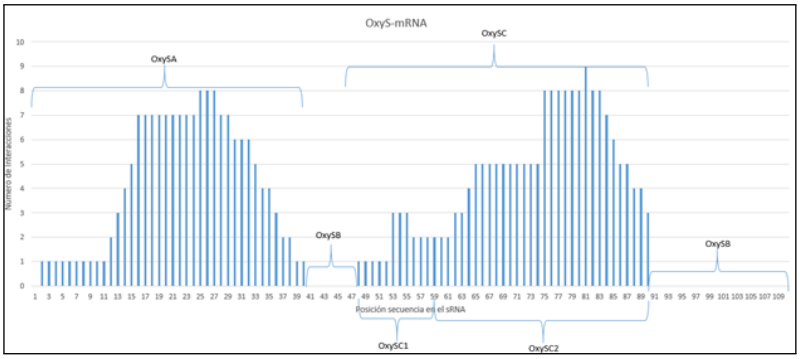


Figura C.1: Gráfico de la Estructura OxyS

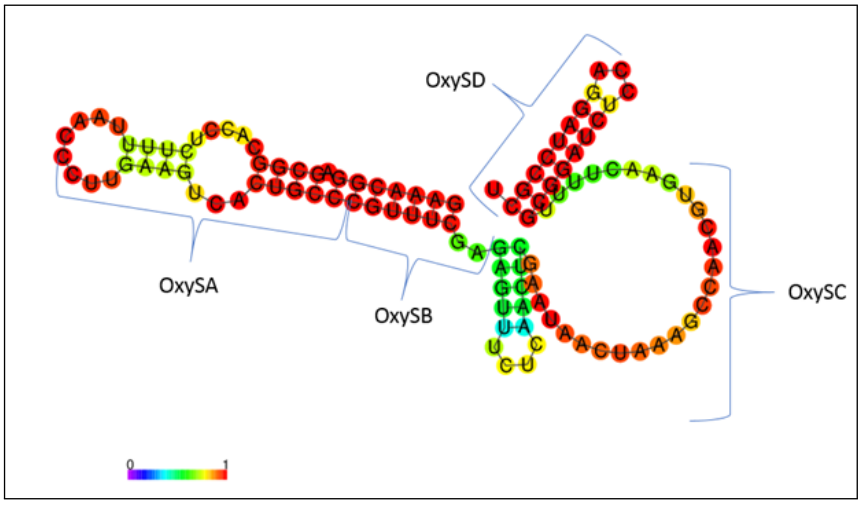


Figura C.2: Estructura Secundaria se OxyS, con sus zonas de interacción definidas

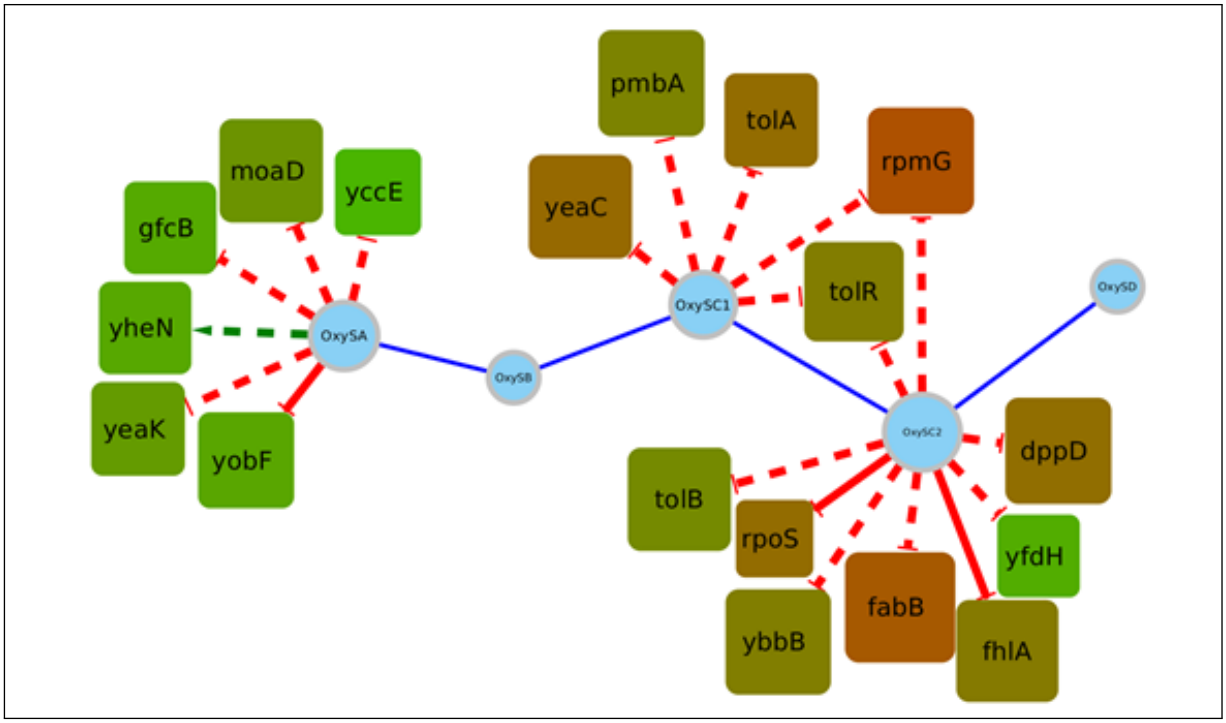


Figura C.3: Modelo de red de interacción OxyS

C.0.2. MicA

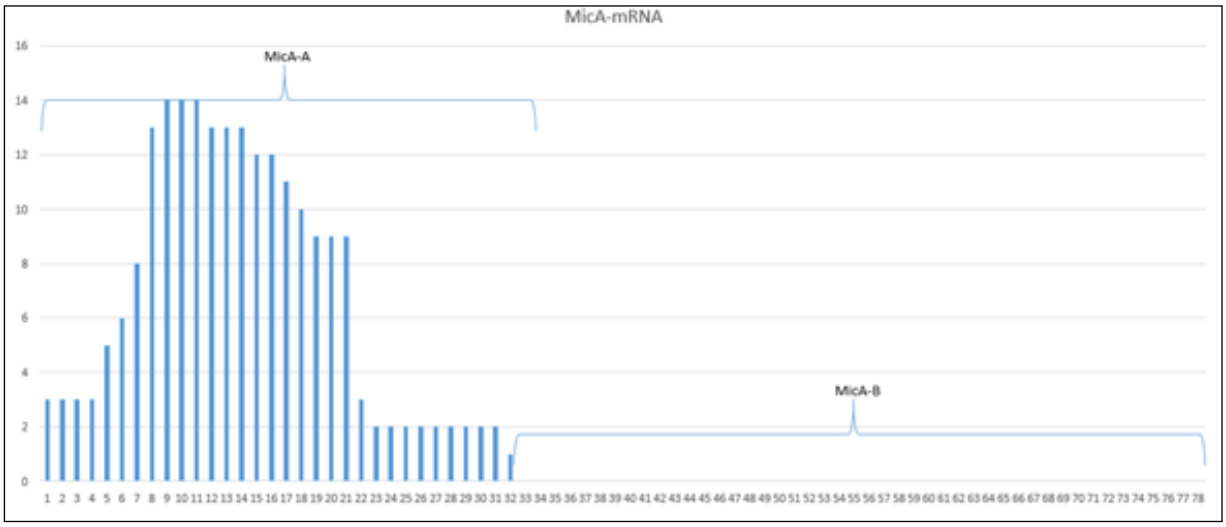


Figura C.4: Gráfico de la Estructura MicA.

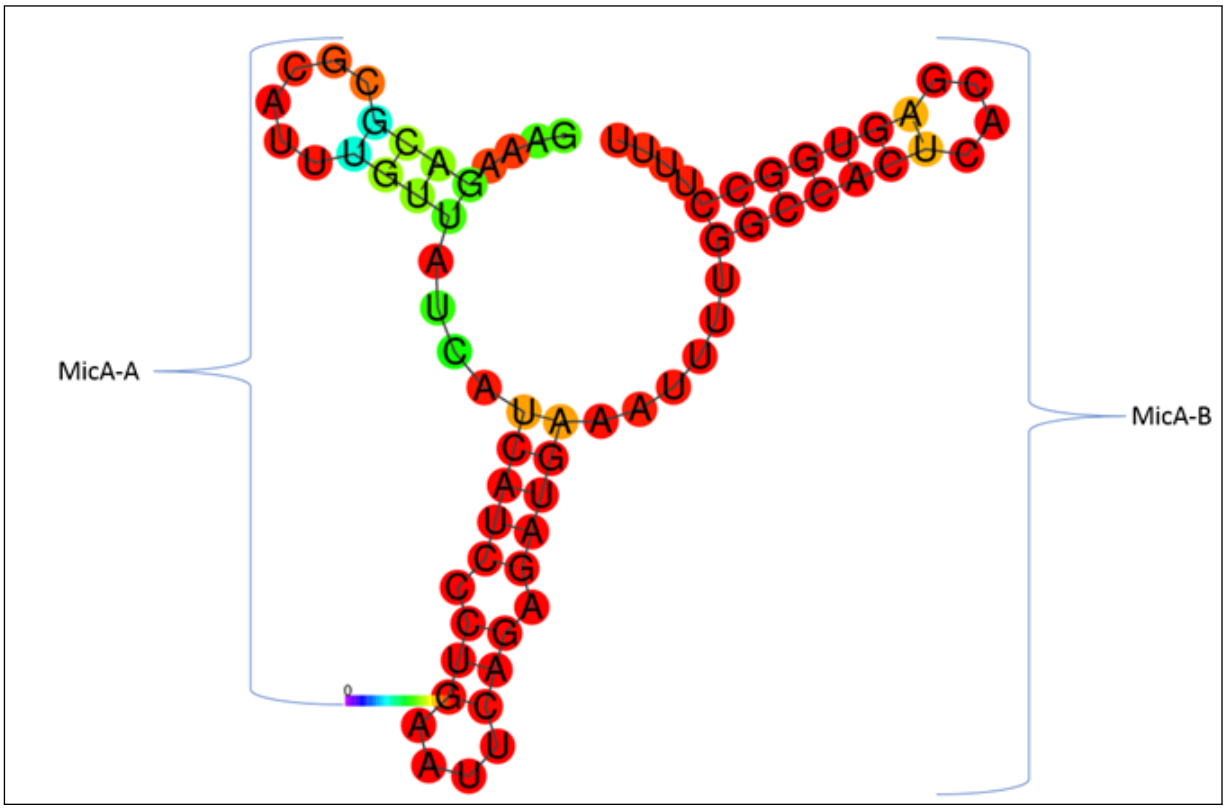


Figura C.5: Estructura Secundaria MicA, con sus zonas de interacción definidas.

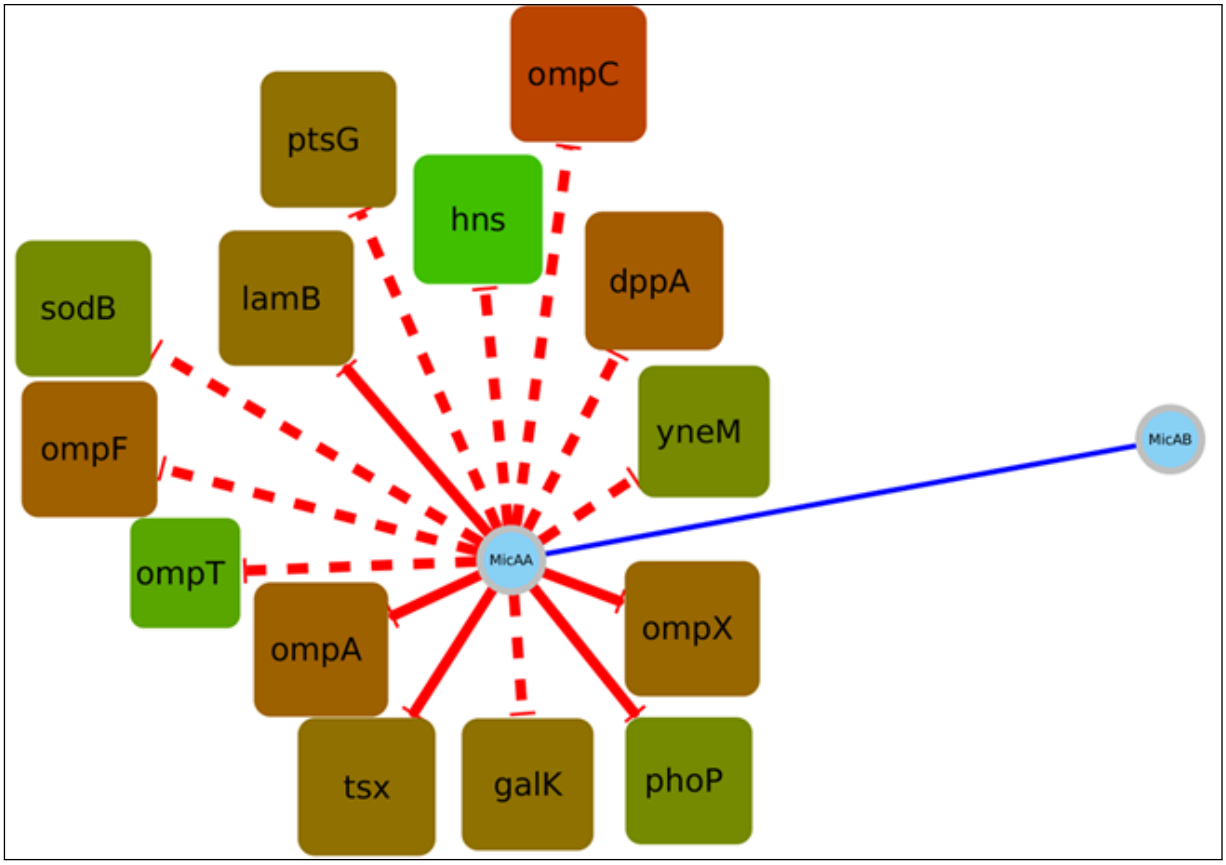


Figura C.6: Modelo de red de Interacción MicA.

C.0.3. RybB

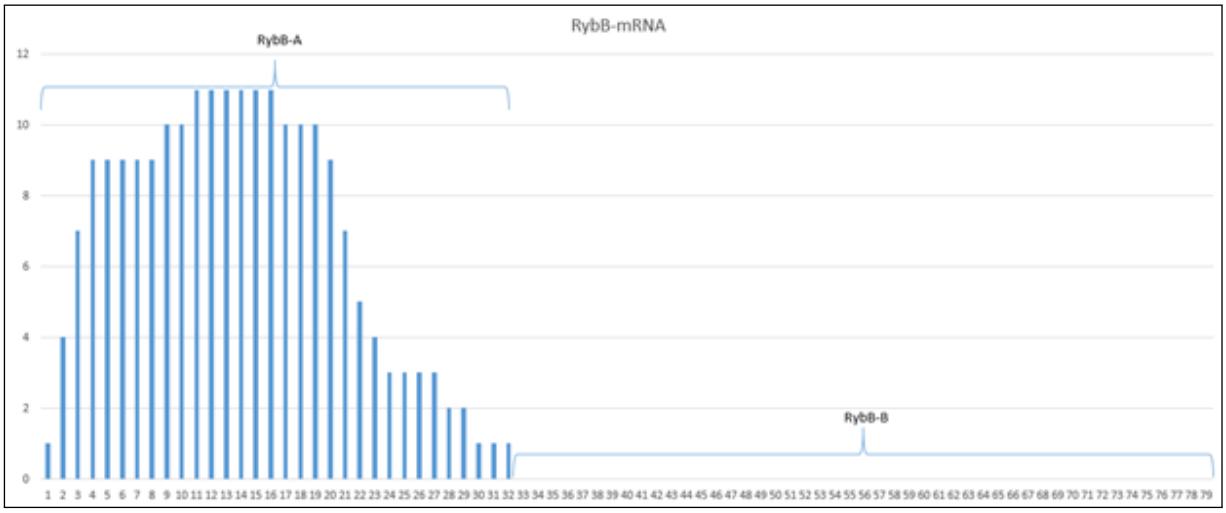


Figura C.7: Gráfico de la Estructura RybB.

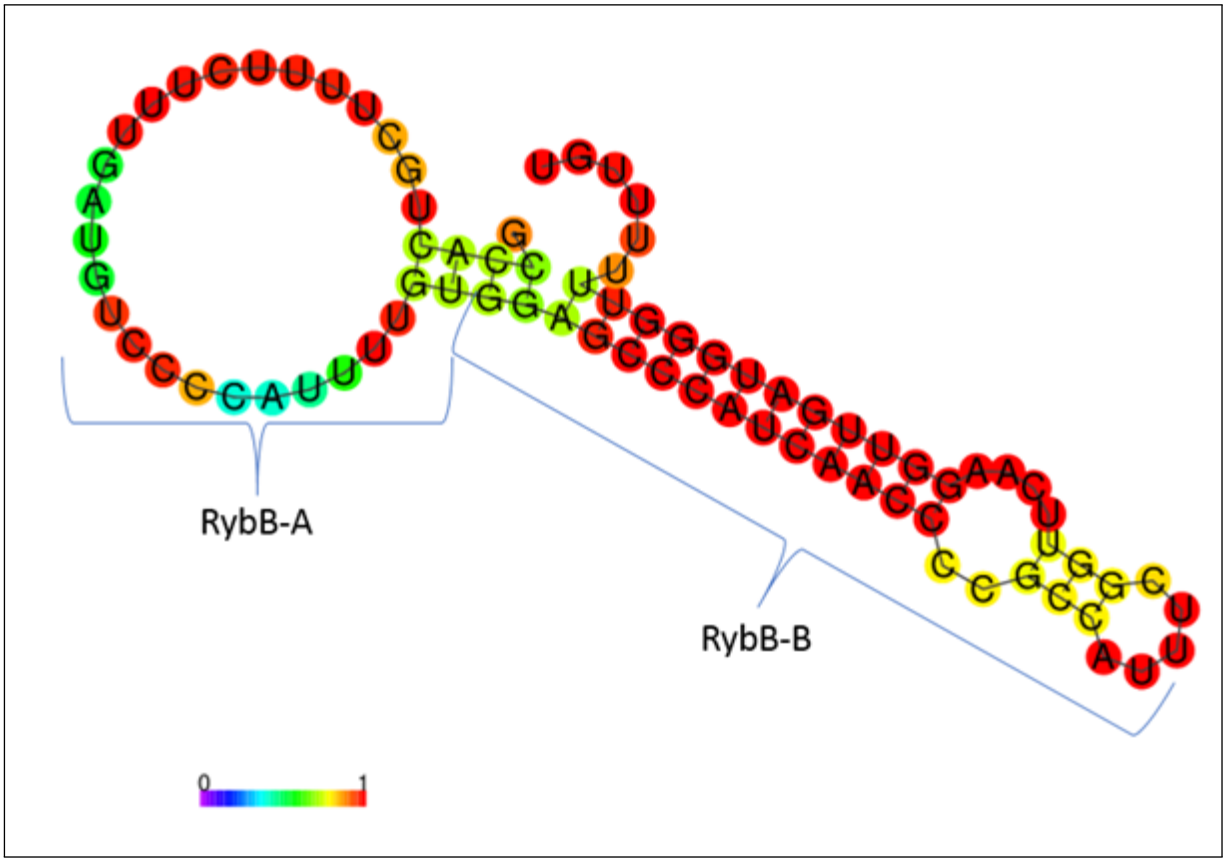


Figura C.8: Estructura Secundaria RybB, con sus zonas de interacción definidas.

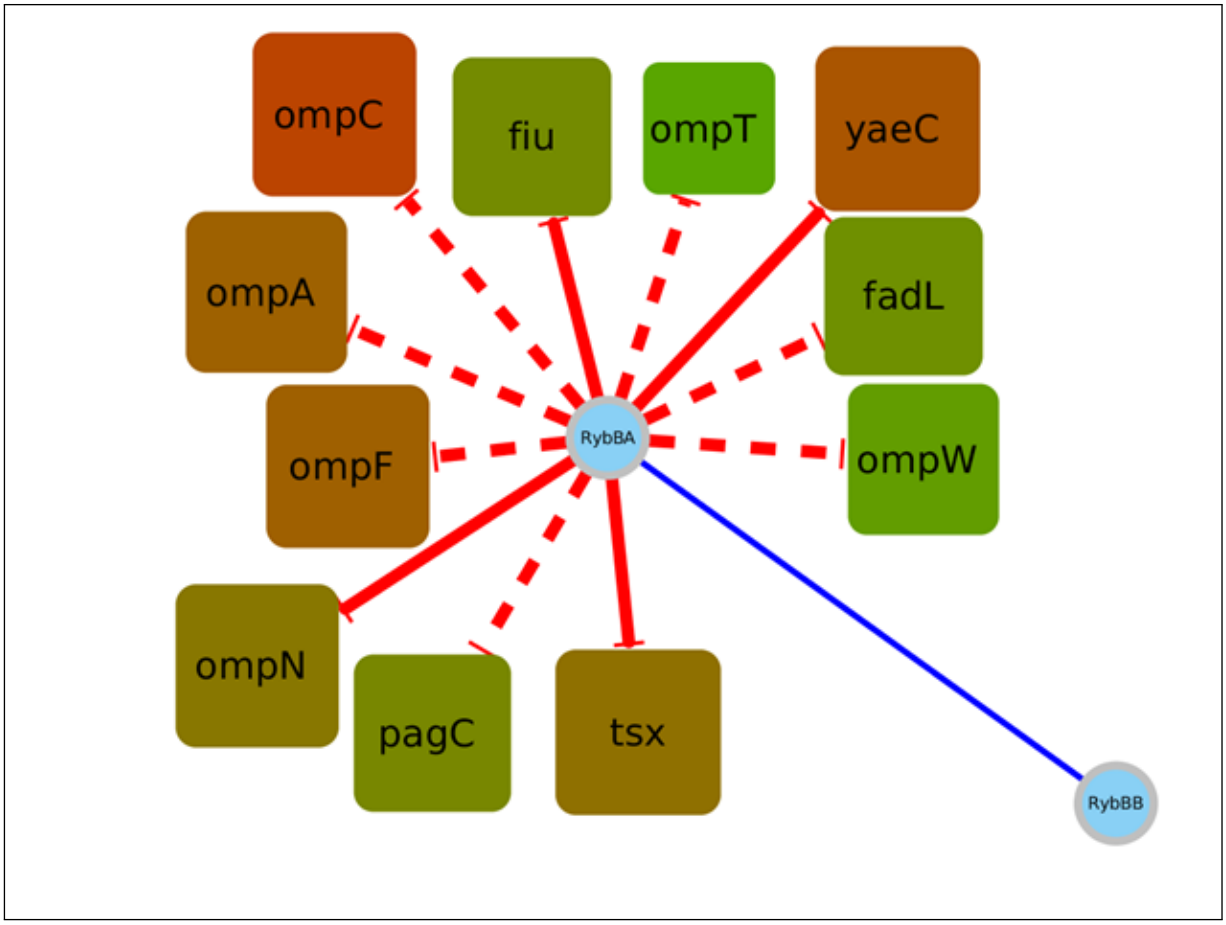


Figura C.9: Modelo de red de Interacción RybB.

C.0.4. FnrS

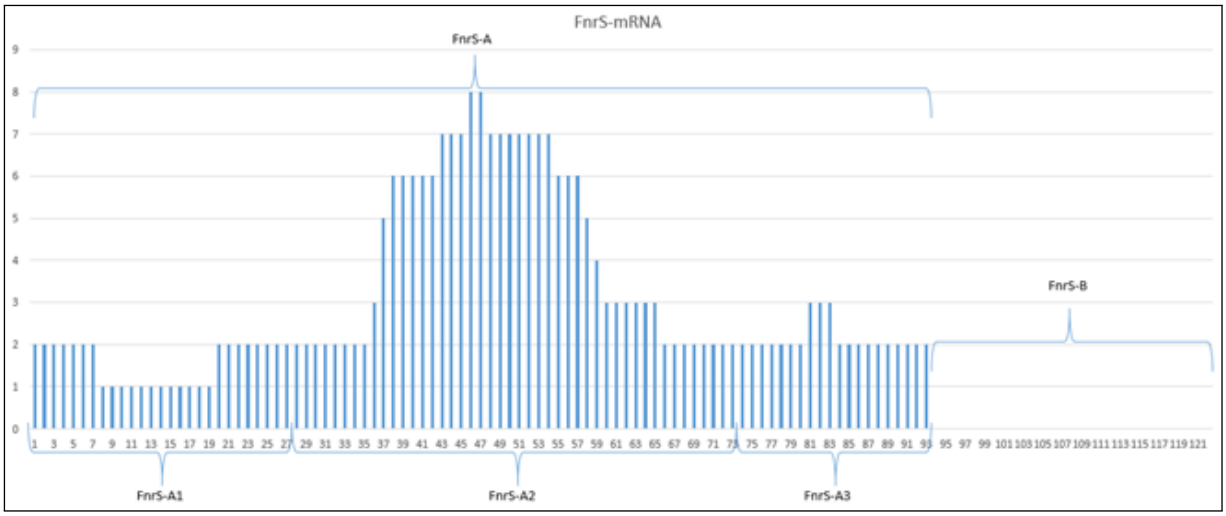


Figura C.10: Gráfico de la Estructura FnrS.

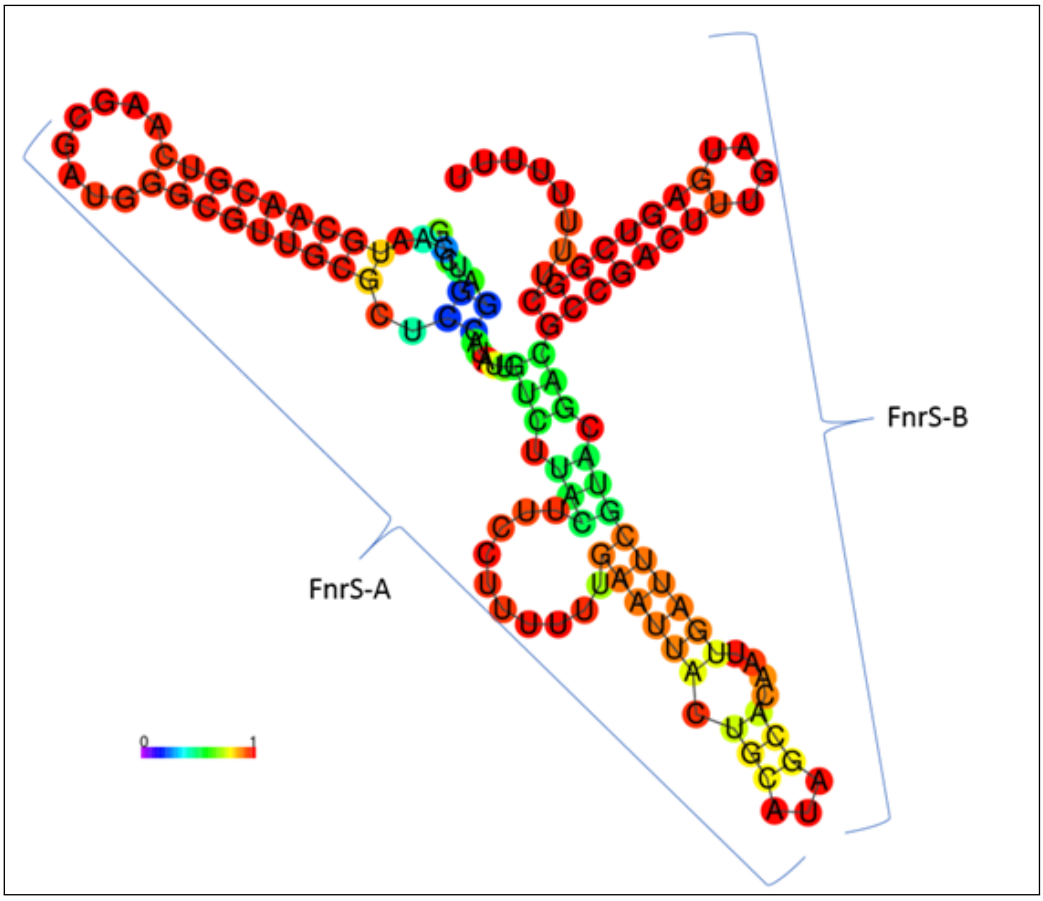


Figura C.11: Estructura Secundaria FnrS, con sus zonas de interacción definidas.

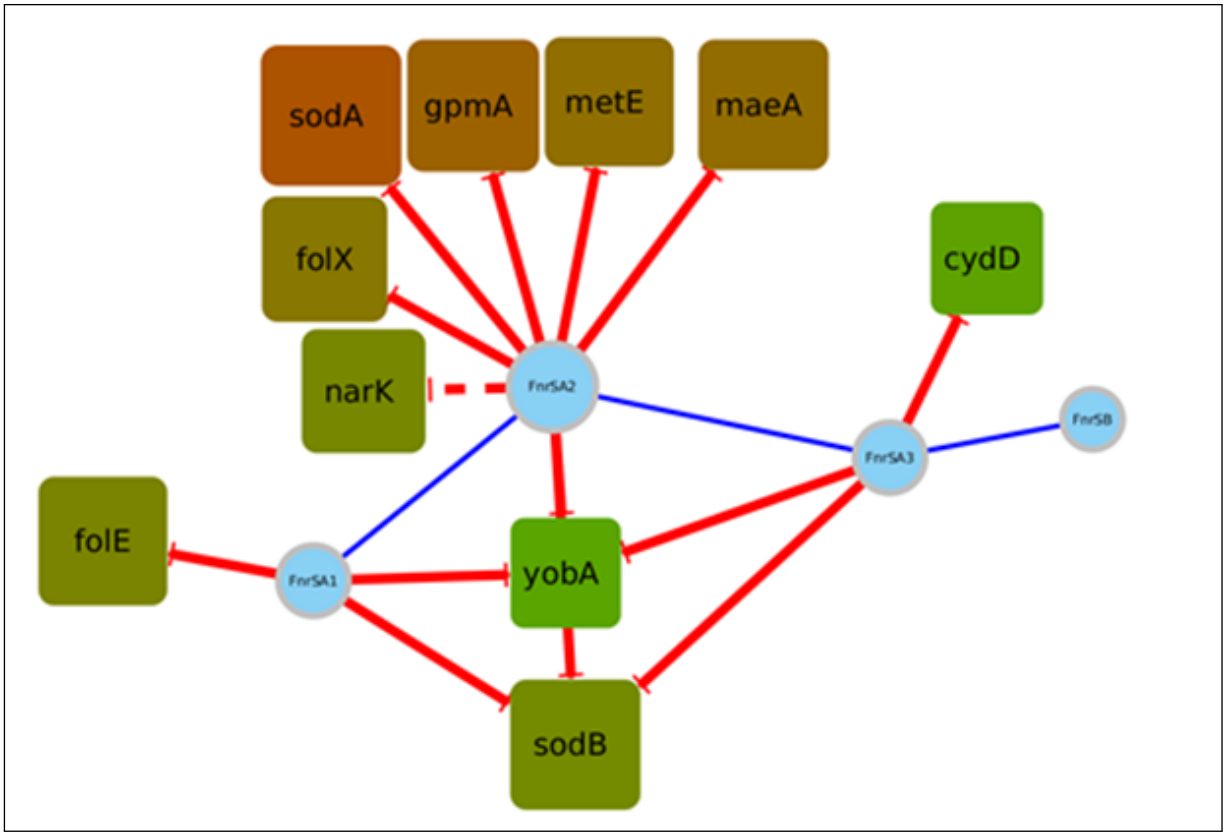


Figura C.12: Modelo de red de Interacción FnrS.

C.0.5. DsrA

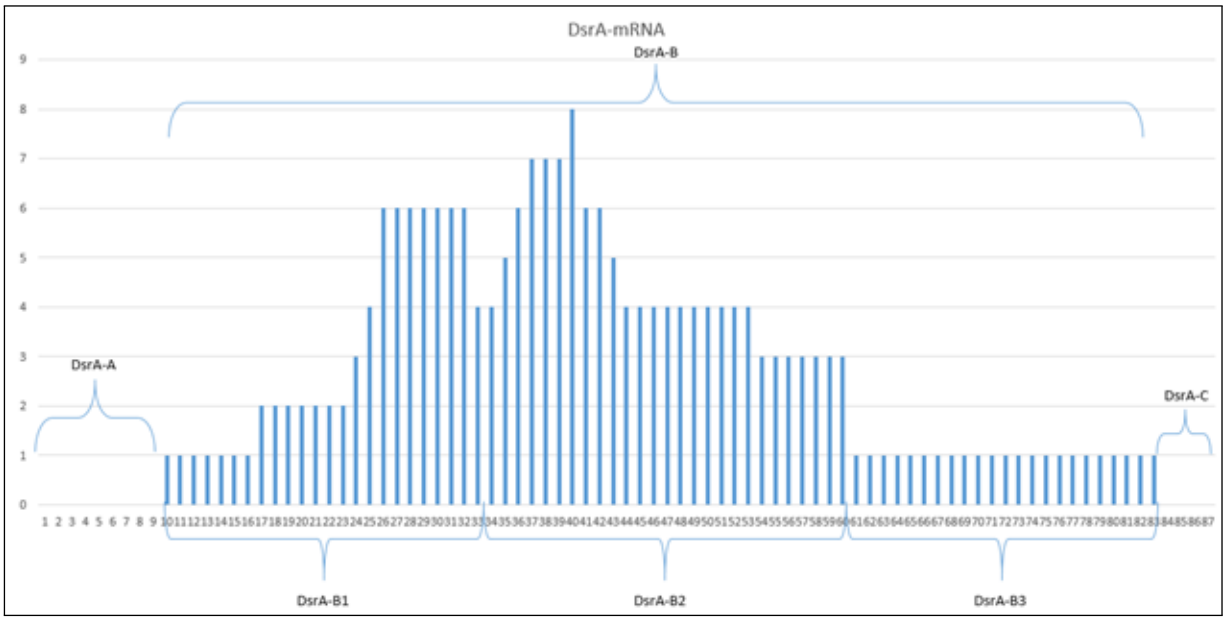


Figura C.13: Gráfico de la Estructura DsrA.

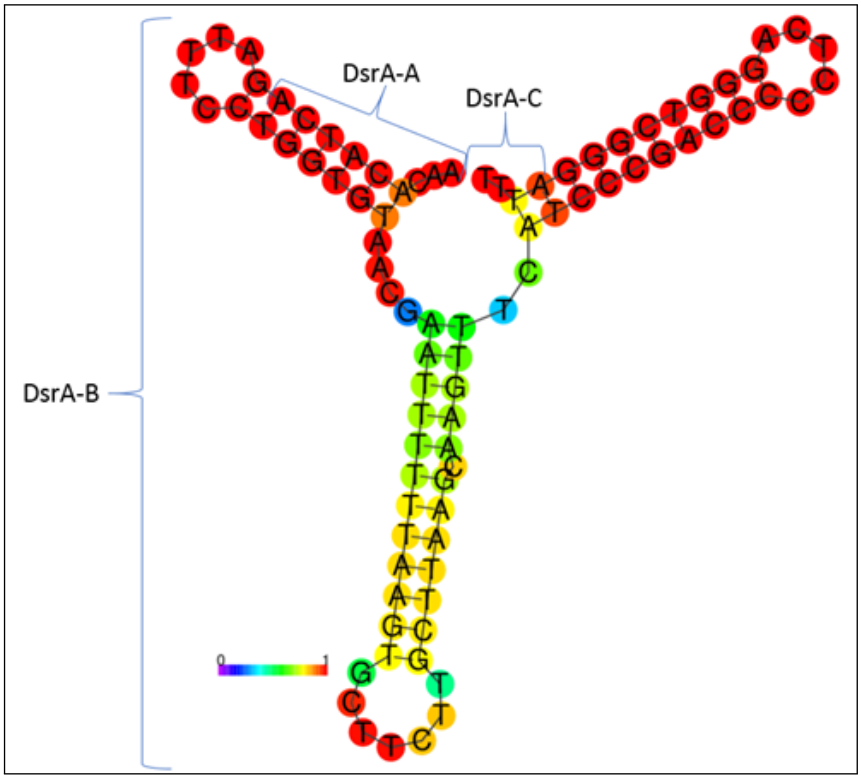


Figura C.14: Estructura Secundaria DsrA, con sus zonas de interacción definidas.

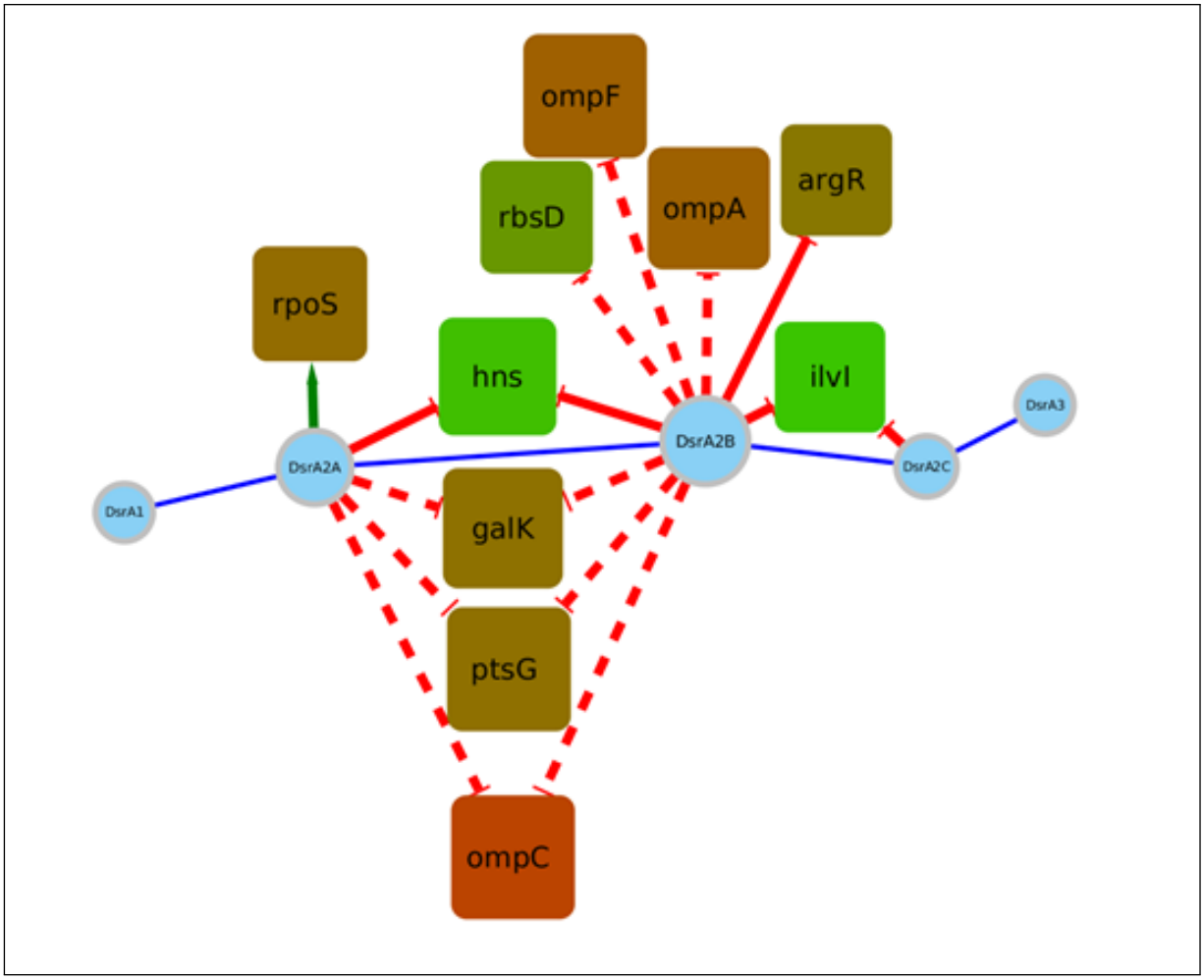


Figura C.15: Modelo de red de Interacción DsrA.

C.0.6. SgrS

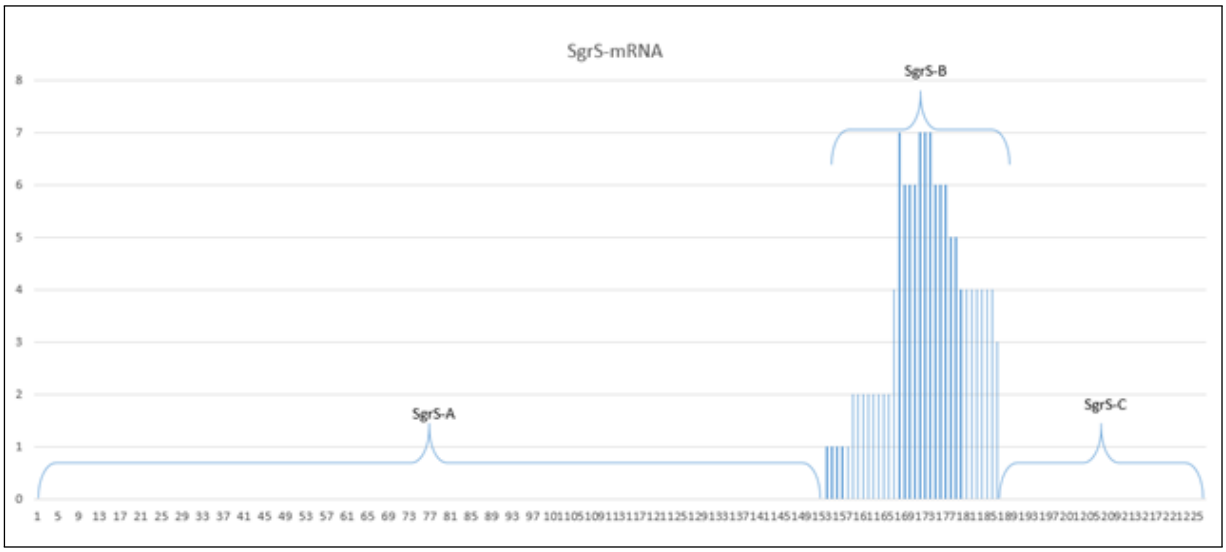


Figura C.16: Gráfico de la Estructura SgrS.

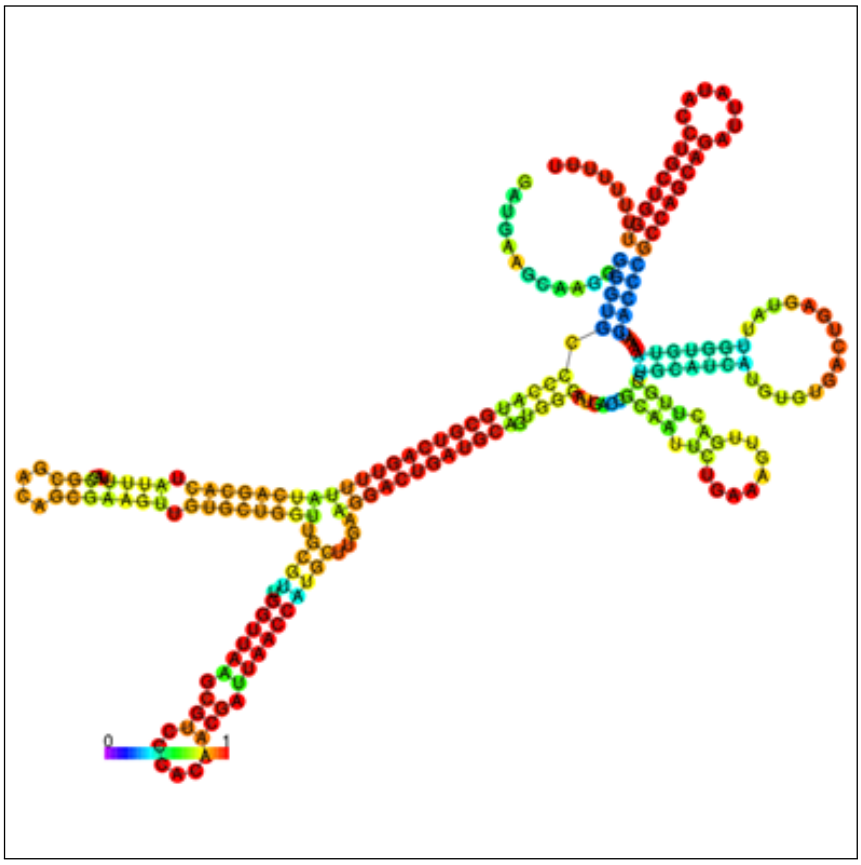


Figura C.17: Estructura Secundaria SgrS, con sus zonas de interacción definidas.

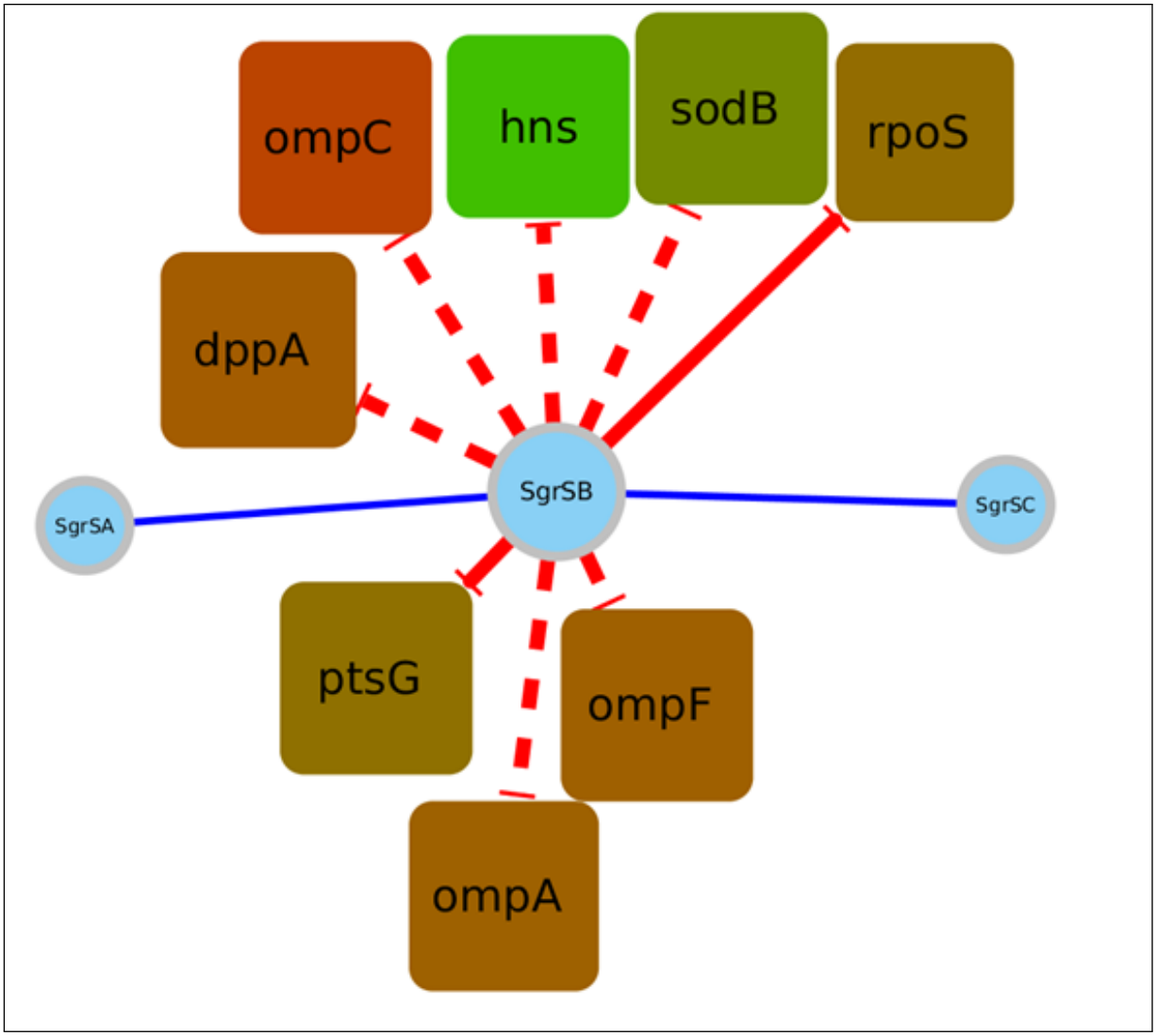


Figura C.18: Modelo de red de Interacción SgrS.

C.0.7. MicC

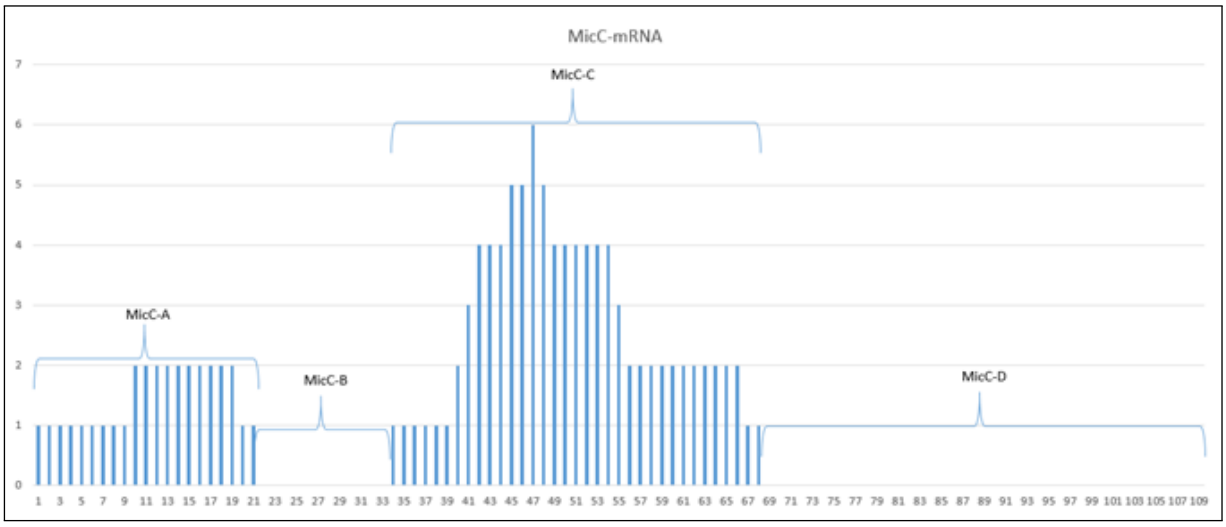


Figura C.19: Gráfico de la Estructura MicC.

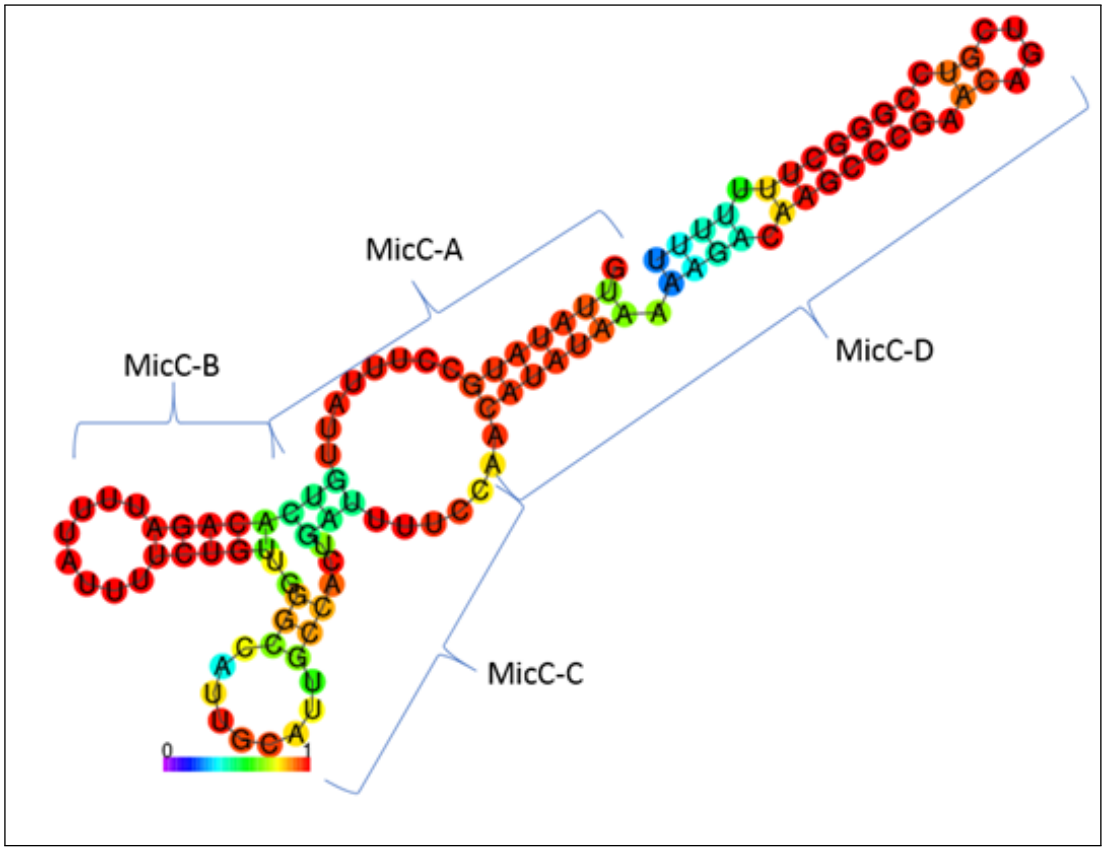


Figura C.20: Estructura Secundaria MicC, con sus zonas de interacción definidas.

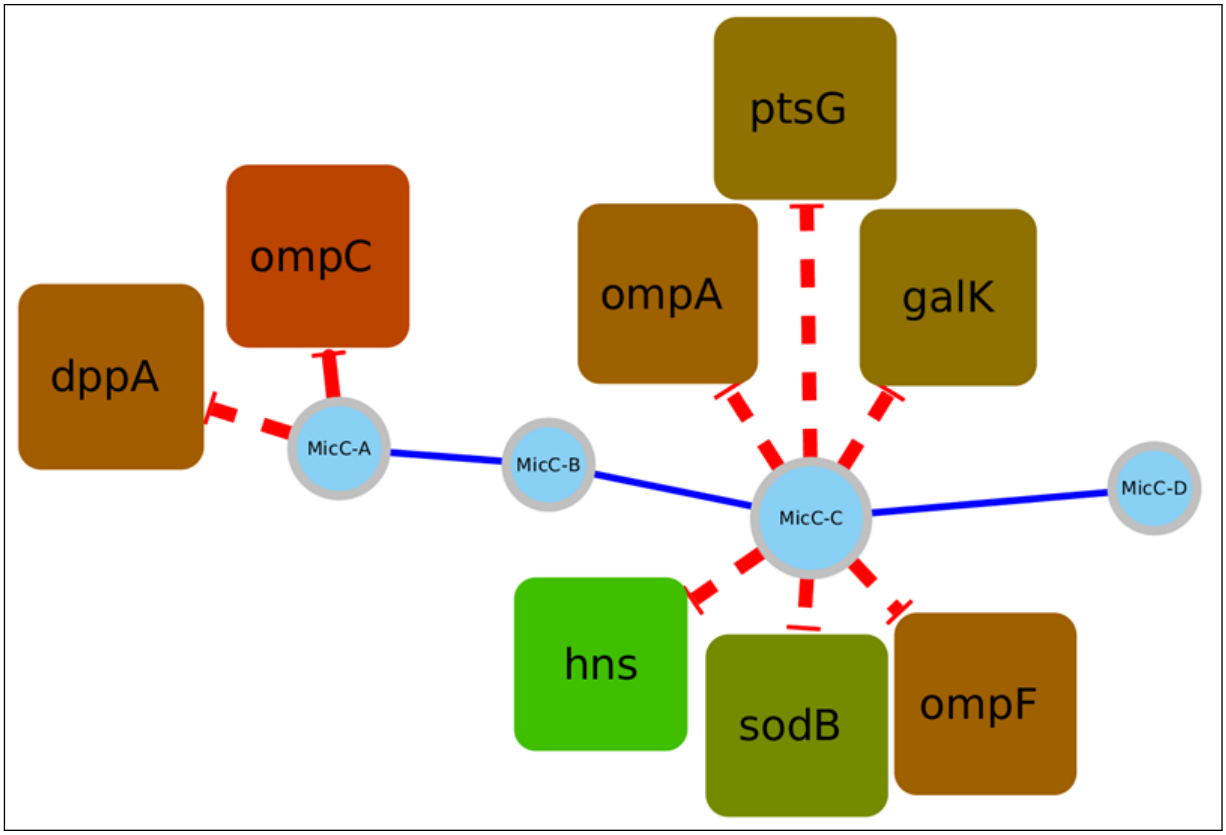


Figura C.21: Modelo de red de Interacción MicC.

C.0.8. MicF

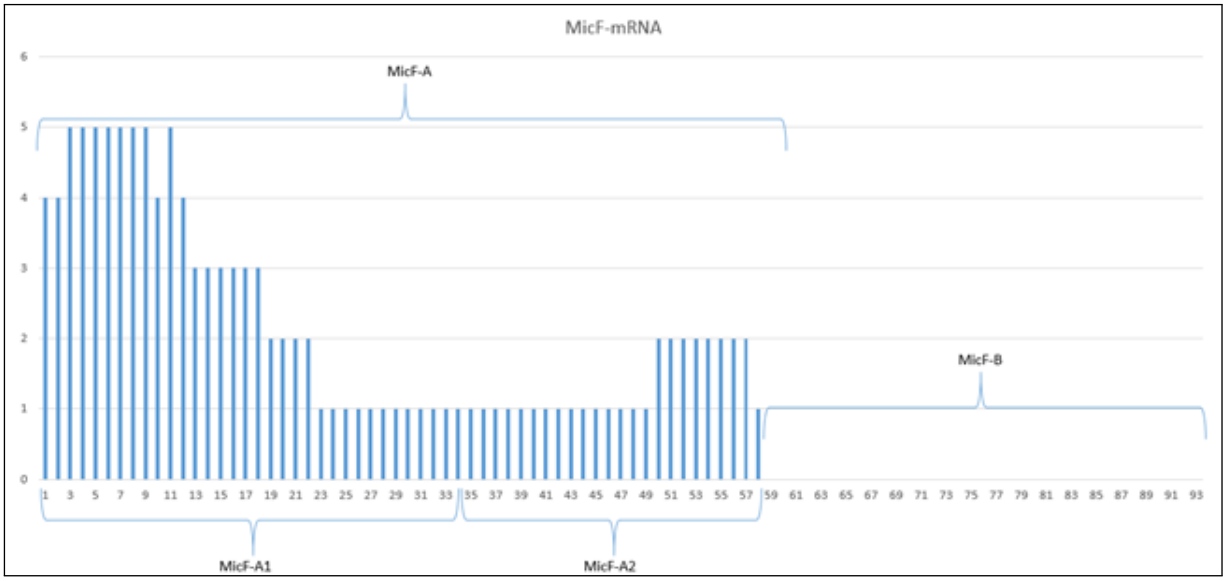


Figura C.22: Gráfico de la Estructura MicF.

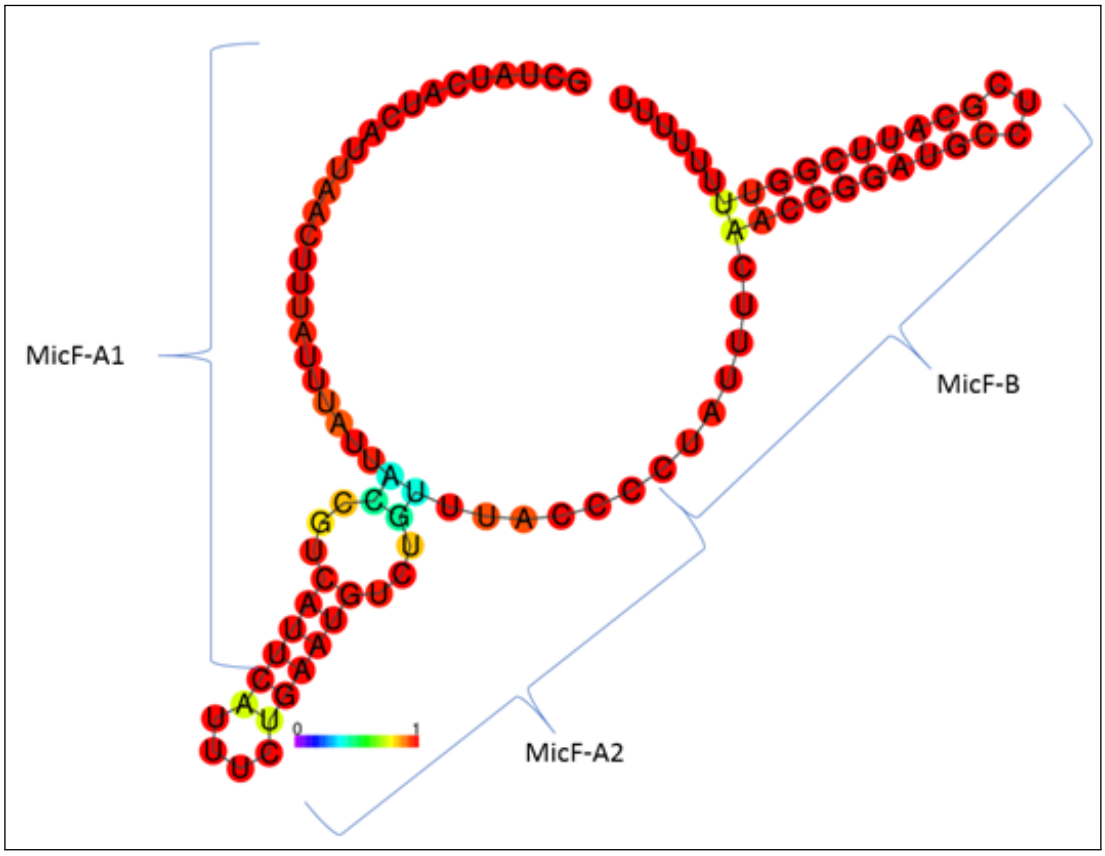


Figura C.23: Estructura Secundaria MicF, con sus zonas de interacción definidas.

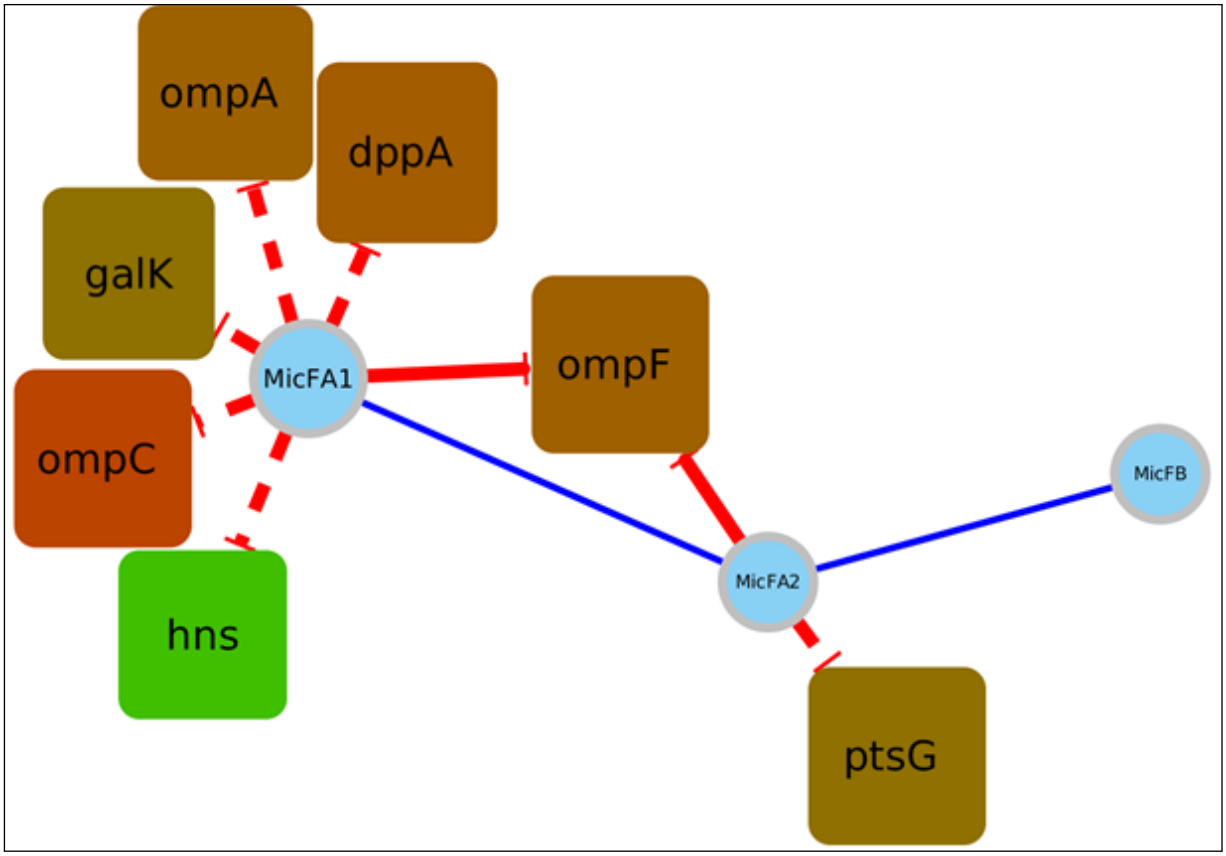


Figura C.24: Modelo de red de Interacción MicF.

C.0.9. RprA

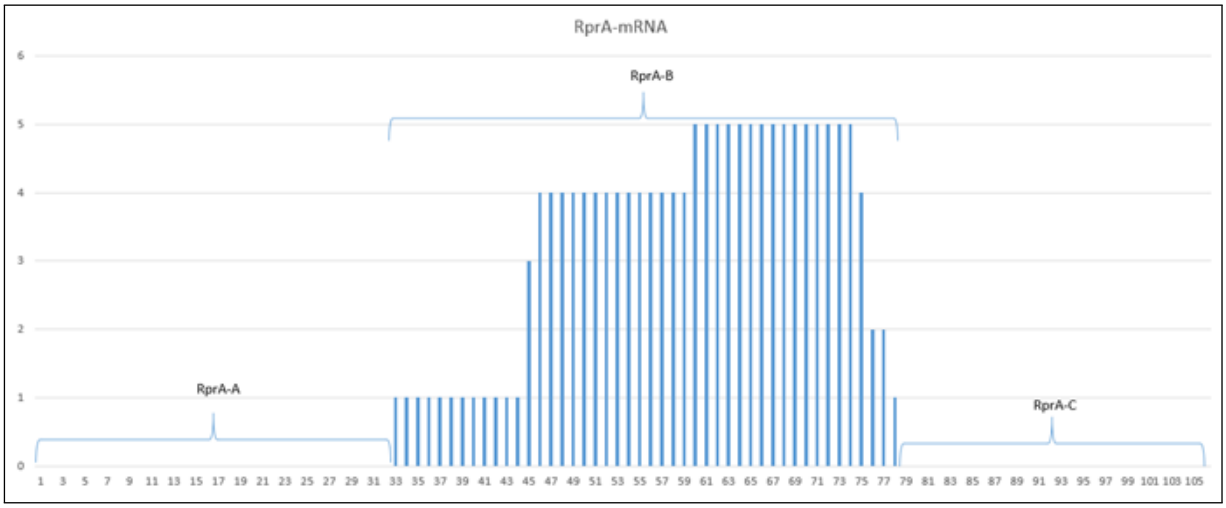


Figura C.25: Gráfico de la Estructura RprA.

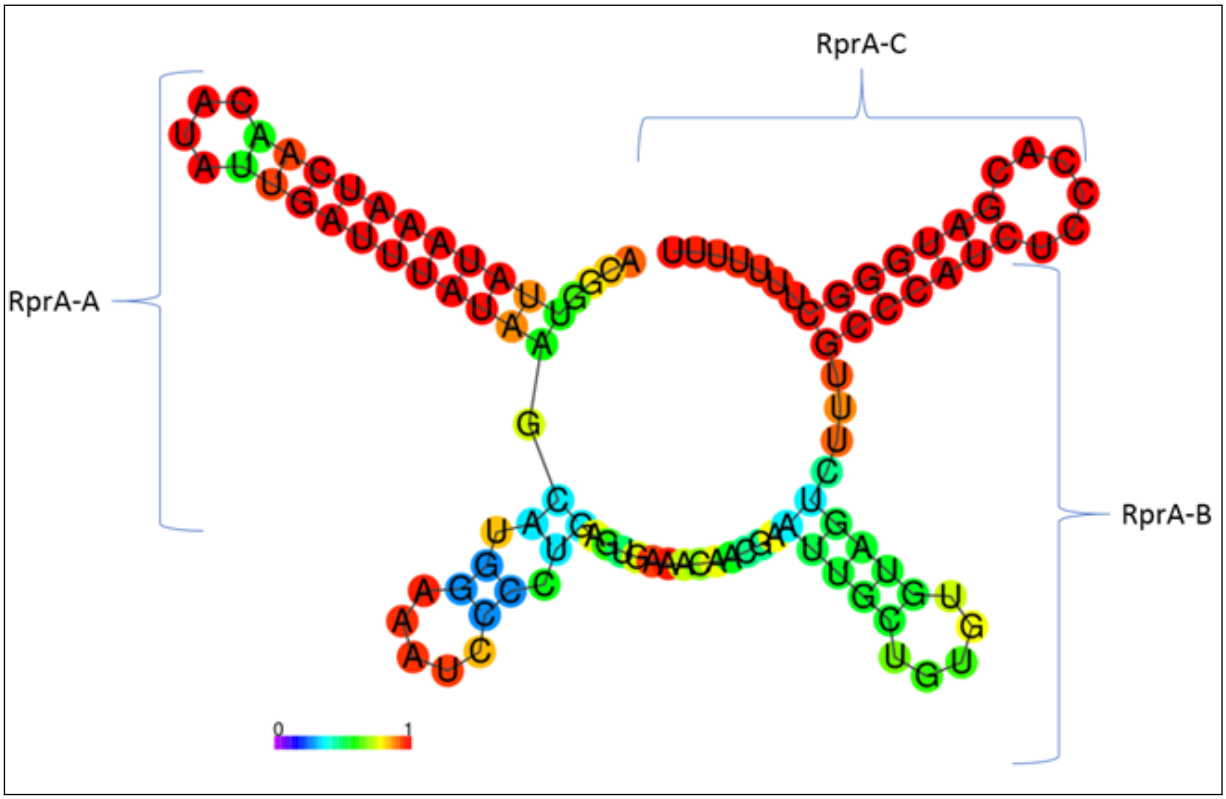


Figura C.26: Estructura Secundaria RprA, con sus zonas de interacción definidas.

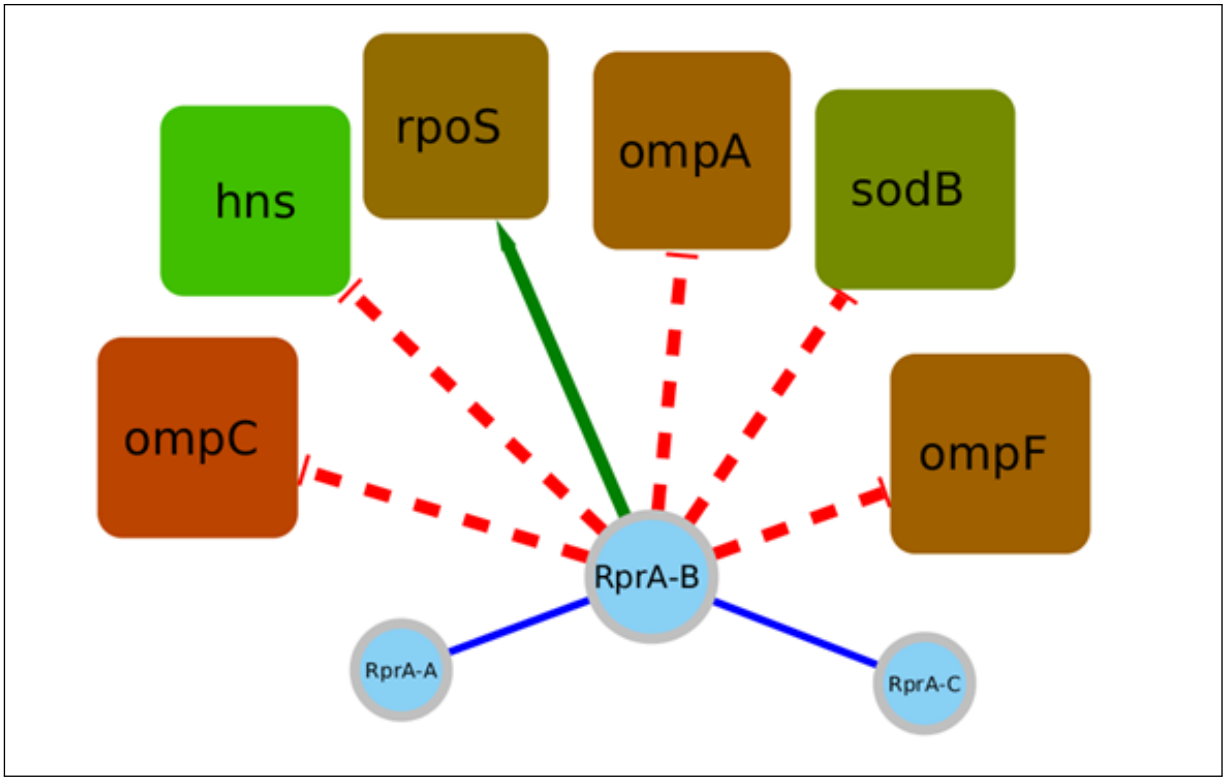


Figura C.27: Modelo de red de Interacción RprA.

C.0.10. CyaR

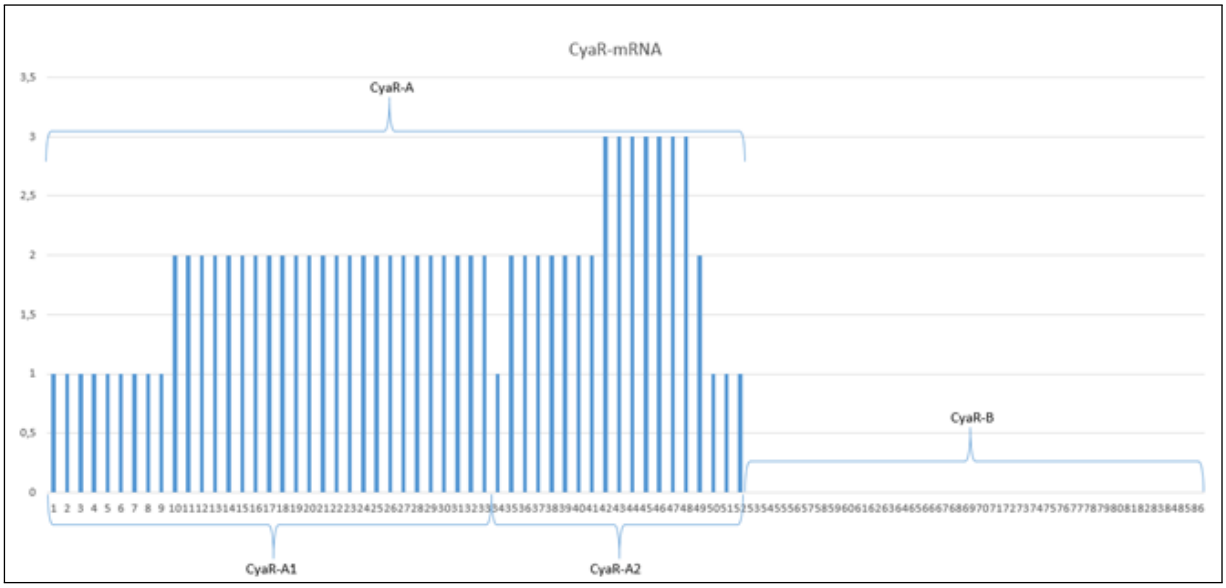


Figura C.28: Gráfico de la Estructura CyaR.

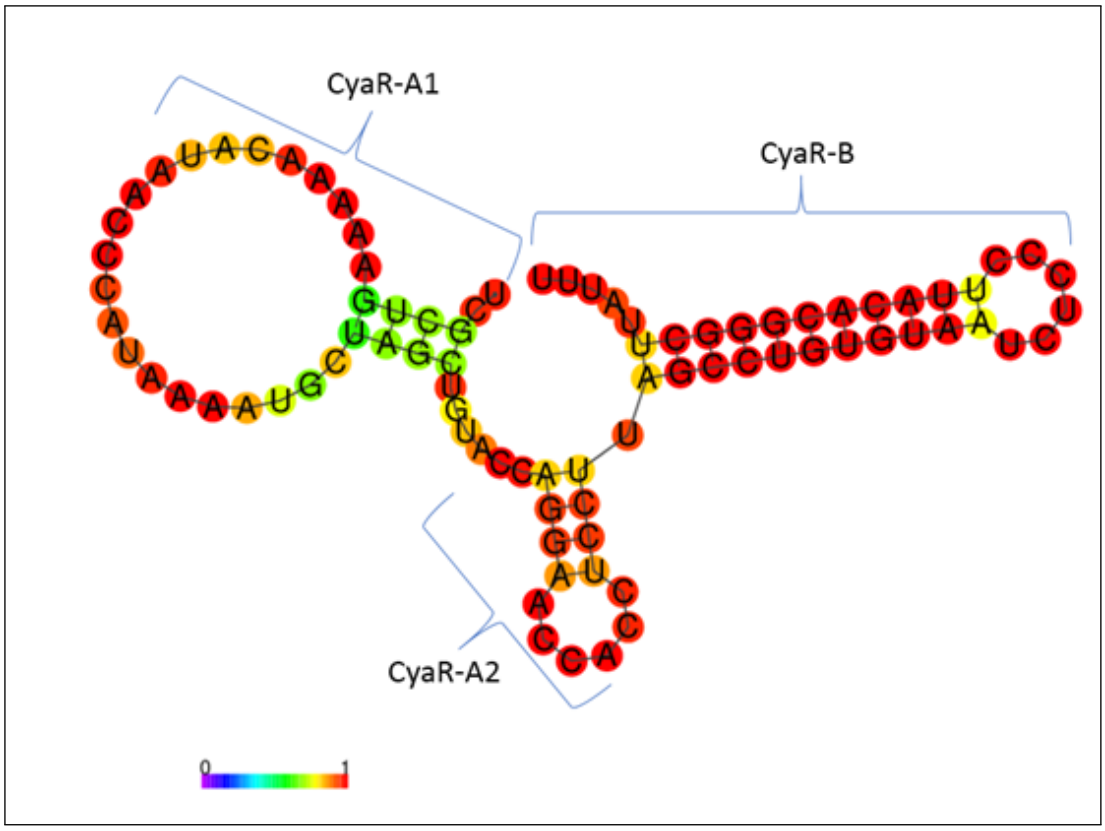


Figura C.29: Estructura Secundaria CyaR, con sus zonas de interacción definidas.

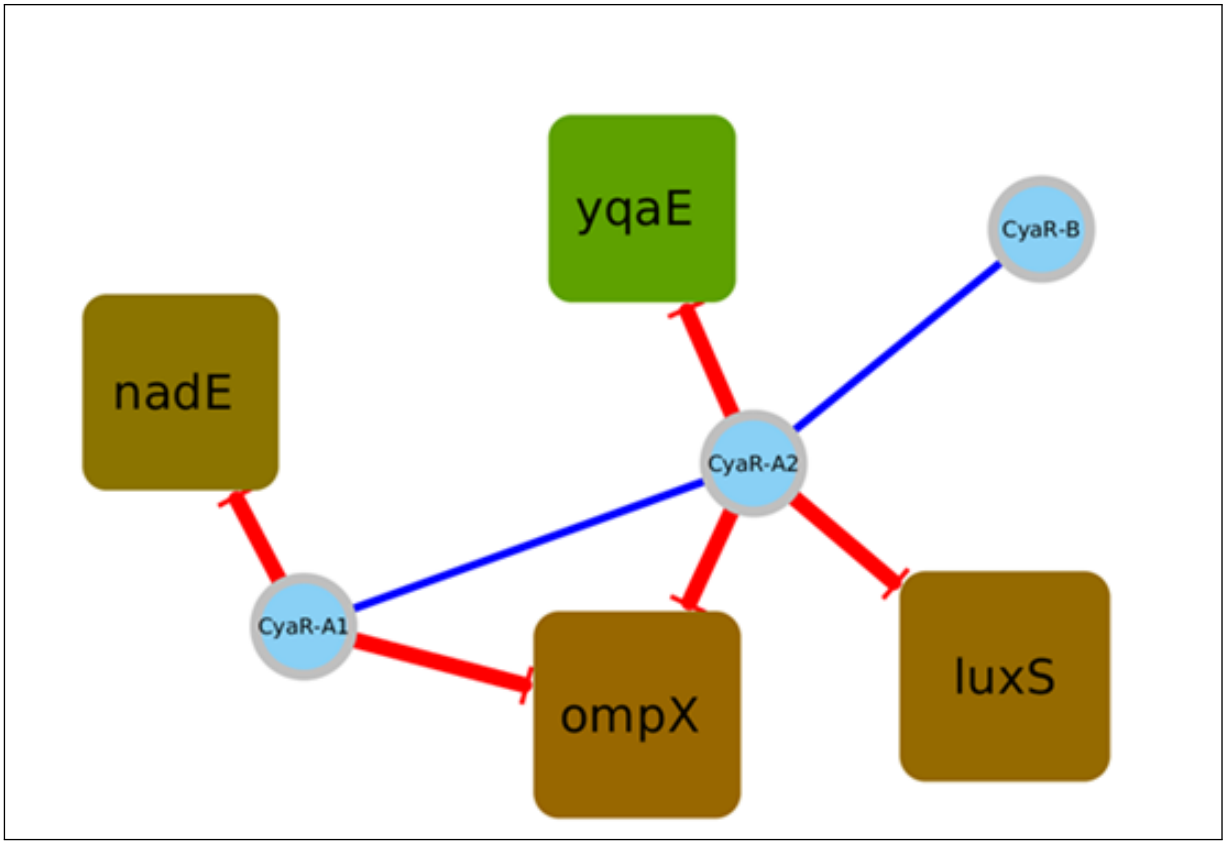


Figura C.30: Modelo de red de Interacción CyaR.

C.0.11. RygD

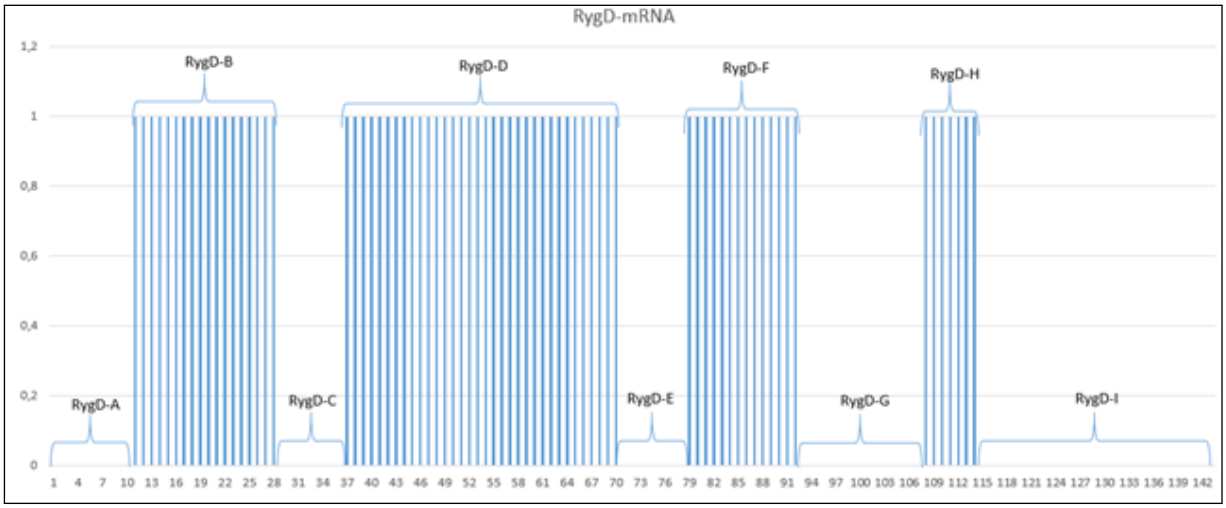


Figura C.31: Gráfico de la Estructura RygD.

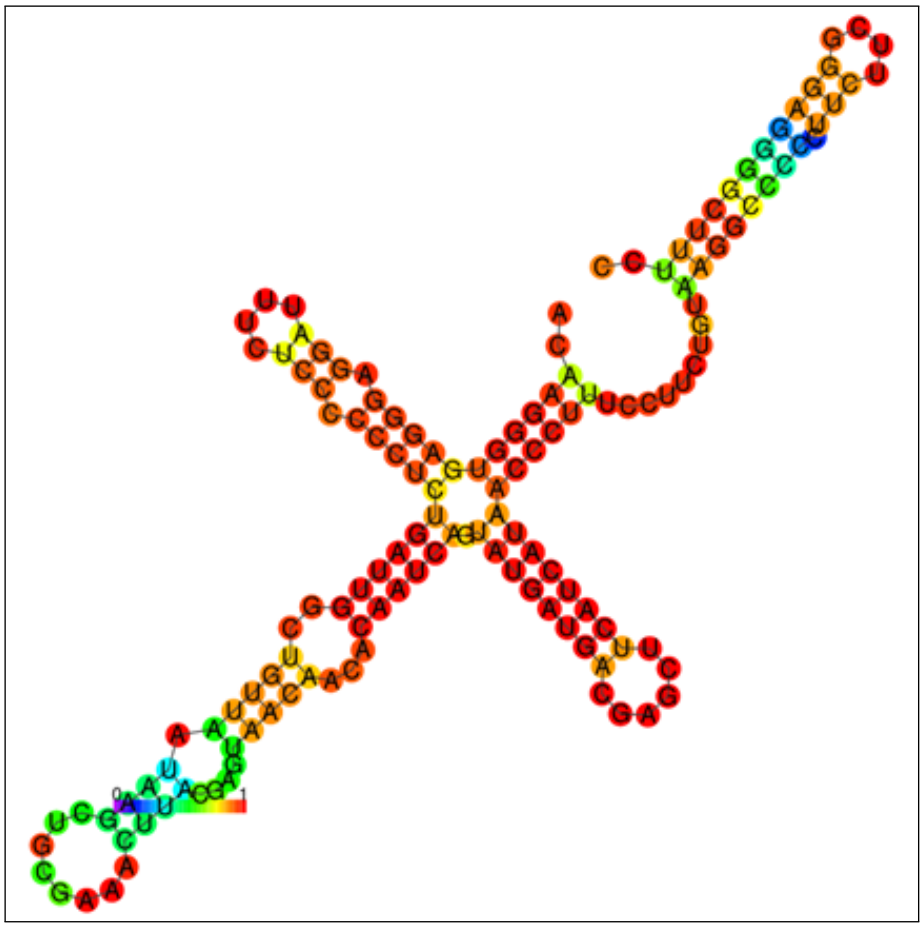


Figura C.32: Estructura Secundaria RygD, con sus zonas de interacción definidas.

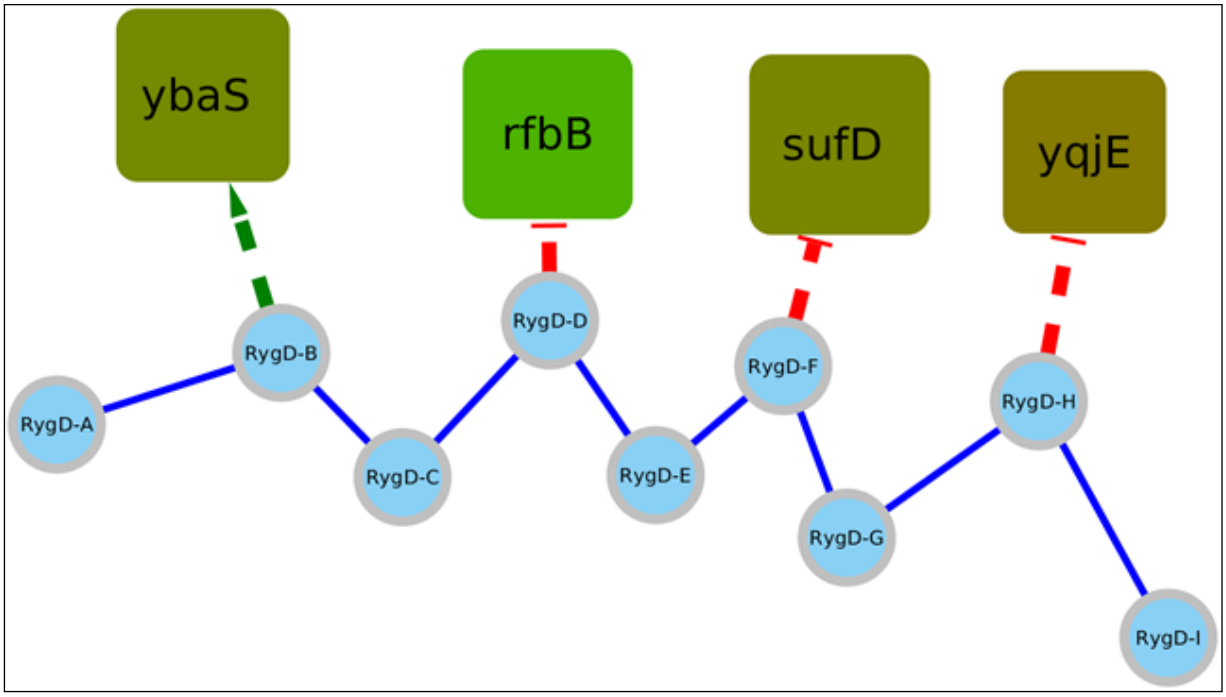


Figura C.33: Modelo de red de Interacción RygD.

D. Modelo Base de Datos

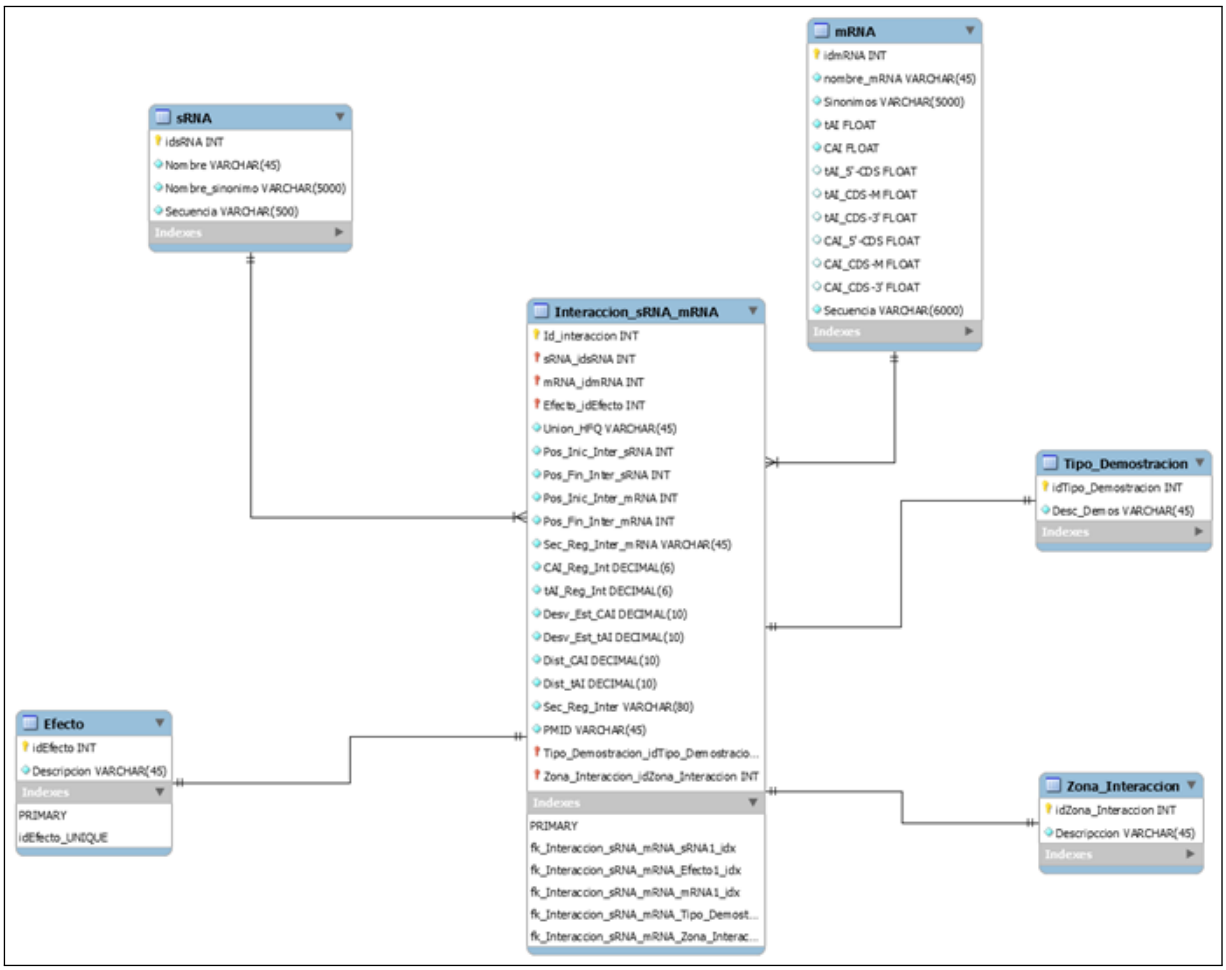


Figura D.1: Modelo de la Base de Datos Creado en Mysql Workbench.