



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Análisis Topológico de la Red de Regulación
Transcripcional de *Bacillus subtilis*

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
BIÓLOGO

PRESENTA:
ALEJANDRA MAYELA MANJARREZ CASAS

DIRECTOR DE TESIS:
DRA. ROSA MARÍA GUTIÉRREZ RÍOS



2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de datos del Jurado

<p>1. Datos del alumno Manjarrez Casas Alejandra Mayela 26 52 22 66 Universidad Nacional Autónoma de México Facultad de Ciencias Biología 405018886</p>
<p>2. Datos del tutor Dra Rosa María Gutiérrez Ríos</p>
<p>3. Datos del sinodal 1 Dr Enrique Merino Pérez</p>
<p>4. Datos del sinodal 2 Dr José Luis Puente García</p>
<p>5. Datos del sinodal 3 Dr Lorenzo Patrick Segovia Forcella</p>
<p>6. Datos del sinodal 4 Dr Julio Augusto Freyre González</p>
<p>7. Datos del trabajo escrito Análisis Topológico de la Red de Regulación Transcripcional de <i>Bacillus subtilis</i> 92 p 2009</p>

A mis padres

AGRADECIMIENTOS

A la Dra. Rosa María Gutiérrez Ríos, por la dirección de esta tesis y por guiarme a lo largo de mis primeros pasos en este complicado camino de la ciencia. Gracias por creer en mí y por todo el apoyo que me brindaste; por tu tiempo y por estar siempre dispuesta a ayudarme. Fue un placer trabajar contigo y espero poder seguir haciéndolo de alguna u otra manera en el futuro.

Al Dr. Enrique Merino Pérez, por aceptarme en tu laboratorio, por ser un magnífico jefe de grupo y por contagiar tu entusiasmo. Gracias por escucharnos. Heidi te echará de menos.

Al Dr. Julio A. Freyre González, por facilitarme la red de regulación de *Bacillus subtilis*; por responder siempre mis dudas de la manera más clara posible. Te admiro y prometo leer todos los artículos que publiques. No pierdas tu genuino interés por hacer ciencia.

Al M.C. Ricardo Ciria, por ayudarme en mis primeros intentos de programar en Perl, aprendí mucho. A la L.I. Nancy Mena por las incontables veces que me apoyó cuando tuve problemas computacionales. A la M.B. María Luisa Tabche, por su invaluable soporte técnico en el laboratorio. A Rosalva, por su apoyo en todo lo referente a asuntos administrativos. A don Pablo y don Lalo, por siempre estar al pendiente de que el lugar en donde trabajamos esté impecable. A todos los integrantes del laboratorio, incluyendo también al grupo de la Dra. Espín, por su amistad, apoyo y risas. ¡Los voy a extrañar!

Al Dr. José Luis Puente por organizar el taller de la Facultad de Ciencias que me trajo hasta aquí. Gracias también por todo el apoyo y tiempo que me brindaste durante las complicadas andanzas administrativas de mi tesis.

A mis sinodales: Rosy, Enrique, José Luis, Lorenzo y Julio, por su veloz pero cuidadosa revisión, sus valiosos comentarios, su apoyo y sus ideas. ¡Qué maravilloso quinteto!

Gracias a Gabriela Román y a Luis Bezares por su ayuda en la elaboración de algunas de las figuras presentadas en esta tesis.

Gracias al Instituto de Biotecnología y a DGAPA por el apoyo económico.

Un agradecimiento muy especial a toda la gente que me apoyó en los días más críticos de este proyecto: a mis sinodales, a Luis, a mi madre, a mis alegres compañeros de banca Raúl y María, y a Nancy.

A la UNAM. Espero algún día poder retribuirle un poco de lo mucho que me ha dado.

A mis dos bacterias facultativas favoritas, por darme lo más valioso que tengo. A mi madre, por ser mi punto de equilibrio, porque a pesar de que estamos destinadas a que los kilómetros nos separen, no sé qué haría sin la certeza de que puedo llamarte a la una de la madrugada para platicar. Sigues siendo (y lo serás siempre) mi mejor amiga. A mi padre, por tu honestidad y tu alegría; por tu ejemplo. Gracias a los dos por su amor y por hacer siempre todo lo posible para que yo esté bien. Los admiro y amo.

A Luis, por ser de otro planeta, por permitir que compartamos nuestra pasión por entender la vida y por contagiarme tu espíritu científico. Gracias por tu amor, por tu apoyo incondicional y por echarme porras durante toda la elaboración de este proyecto. Gracias por ser idealista, evolucionista y loco. ¡Te amo mucho flaquitoide!

A Emma, por tu amor y sabiduría; te extraño. A toda mi familia, por darme siempre la mano, por sus ánimos y su amistad. Gracias por aceptarme como soy, loca y rara; y por perdonarme por todas las veces que he estado ausente.

A todos los maestros que me han inspirado y que, si no fuera por ellos, yo no estaría aquí. Sería imposible nombrarlos a todos.

A mis amigos.

*Not Chaos like together crush'd and bruis'd,
But as the world, harmoniously confus'd:
Where order in variety we see,
And where, tho' all things differ, all agree.*

Fragmento de "Windsor Forest" de Alexander Pope

ÍNDICE GENERAL

Lista de abreviaturas

1. INTRODUCCIÓN	1
1.1. La regulación de la transcripción.....	1
1.1.1. Generalidades de la regulación.....	1
1.1.2. Estudiando la regulación de la transcripción de una manera global.....	6
1.2. Teoría de redes.....	7
1.2.1. Los inicios de la teoría.....	7
1.2.2. Las redes y sus propiedades.....	8
1.2.3. Redes aleatorias.....	12
1.2.4. Redes complejas.....	13
1.2.5. La importancia del concepto de modularidad en Biología.....	15
1.2.6. Motivos topológicos.....	16
1.3. <i>Bacillus subtilis</i> como modelo de estudio.....	18
1.3.1. Acerca de <i>B. subtilis</i>	18
1.3.2. El genoma y su regulación.....	22
2. ANTECEDENTES	23
3. OBJETIVOS, JUSTIFICACIÓN e HIPÓTESIS	26
3.1. Objetivos.....	26
3.2. Justificación.....	26
3.3. Hipótesis.....	27
4. METODOLOGÍA	28
4.1. Reconstrucción de la red de regulación transcripcional de <i>B. subtilis</i>	28
4.2. Propiedades topológicas de la red de regulación transcripcional de <i>B.</i> <i>subtilis</i>	29
4.3. Análisis modular de la red de reguladores transcripcionales de <i>B.</i> <i>subtilis</i>	31
4.4. Análisis de la Biología de los módulos encontrados.....	32

5. RESULTADOS y DISCUSIÓN	33
5.1. Reconstrucción de la red de regulación transcripcional de <i>B. subtilis</i>	33
5.2. Topología de la red de regulación transcripcional de <i>B. subtilis</i>	34
5.2.1. Propiedades topológicas.....	34
5.2.2. La red de regulación transcripcional de <i>B. subtilis</i> presenta una estructura jerárquico-modular.....	37
5.3. Análisis modular de la red de reguladores transcripcionales de <i>B. subtilis</i>	38
5.4. Descripción detallada de la Biología de los módulos encontrados.....	40
5.4.1. Módulo de metabolismo de nitrógeno.....	40
5.4.2. Módulo de síntesis de nucleótidos.....	41
5.4.3. Módulo de respuesta a estrés de secreción.....	42
5.4.4. Módulo de regulación de enzimas degradativas.....	43
5.4.5. Módulo de respiración aerobia y anaerobia.....	44
5.4.6. Módulo de metabolismo de carbono.....	45
5.4.7. Módulo de la cascada de regulación en la célula madre.....	48
5.4.8. Módulo de regulación de la diferenciación.....	50
5.4.9. Módulo de regulación de respuesta a estrés.....	58
5.5. La fisiología de <i>B. subtilis</i> reflejada en la organización de su red de regulación transcripcional.....	63
5.5.1. El papel de los factores sigma en la red.....	63
5.5.2. Diferentes funciones en un mismo módulo.....	65
5.5.3. Funciones presentes en más de un módulo.....	68
5.5.4. Entonces... ¿si hay una relación módulo-función? Reflexiones finales.....	71
6. CONCLUSIONES y PERSPECTIVAS	73
6.1. Conclusiones.....	73
6.2. Perspectivas.....	74
7. BIBLIOGRAFÍA	75
8. APÉNDICES	91
8.1. Apéndice 1. Red de Regulación Transcripcional de <i>Bacillus subtilis</i>	91

8.2.	Apéndice 2. Red de Factores Transcripcionales de <i>Bacillus subtilis</i>	91
8.3.	Apéndice 3. Propiedades Topológicas de los Genes de la Red de Regulación Transcripcional de <i>Bacillus subtilis</i>	91
8.4.	Apéndice 4. Tabla de Motivos Topológicos de Tres Elementos en la Red de Regulación Transcripcional de <i>Bacillus subtilis</i>	91
8.5.	Apéndice 5. Métricas de Similitud.....	92
8.6.	Apéndice 6. La Presencia de Factores σ en el Análisis Modular de la Red de Regulación Transcripcional de <i>Bacillus subtilis</i>	92

ÍNDICE DE FIGURAS y TABLAS

Figuras

Figura 1. Estructura de la RNA polimerasa.....	3
Figura 2. Estructura de un promotor procariótico.....	5
Figura 3. El problema de los puentes de Königsberg.....	9
Figura 4. Ejemplo de una red dirigida.....	10
Figura 5. Representación de la conectividad (k) de un nodo.....	11
Figura 6. Representación gráfica de las interacciones entre nodos vecinos y el coeficiente de agrupamiento de un nodo.....	11
Figura 7. Tres modelos de redes con sus respectivas propiedades topológicas.....	14
Figura 8. Los 13 posibles patrones de conexión con tres nodos en una red dirigida.....	17
Figura 9. Motivo <i>feedforward</i>	17
Figura 10. Primeros dibujos de <i>Bacillus subtilis</i> de los cuales se tiene conocimiento.....	19
Figura 11. Ciclo de esporulación de <i>Bacillus subtilis</i>	20
Figura 12. La diferenciación de diversos tipos celulares en <i>B. subtilis</i>	21
Figura 13. Representación gráfica de la red de regulación transcripcional de <i>B. subtilis</i> ..	33
Figura 14. Distribución de la conectividad [$P(k)$].....	35
Figura 15. Distribución del coeficiente de agrupamiento [$C(k)$].....	36
Figura 16. Distribución de vecinos más cercanos [$T(n)$].....	36
Figura 17. Análisis de agrupamiento de la red de factores transcripcionales de <i>B. subtilis</i>	39

Tablas

Tabla 1. Número de nodos y aristas de la red transcripcional de <i>B. subtilis</i>	34
Tabla 2. Propiedades topológicas de la red de regulación de <i>B. subtilis</i>	35

Lista de abreviaturas

C	Coeficiente de agrupamiento
FF	Motivo estructural <i>feedforward</i>
FT	Factor de transcripción
GR	Gen regulado
Hsp	Proteínas de estrés por calor
k	Conectividad
n	Interacciones entre nodos vecinos
$P(k)$	Distribución de conectividad
$T(n)$	Distribución de vecinos más cercanos
UT	Unidad transcripcional

1. INTRODUCCIÓN

1.1. La regulación de la transcripción

According to the strictly structural concept, the genome is considered as a mosaic of independent molecular blue-prints for the building of individual cellular constituents. In the execution of these plans, however, co-ordination is evidently of absolute survival value.

(François Jacob y Jacques Monod, 1960)

1.1.1. Generalidades de la regulación

Desde que en 1952 se descubrió la estructura de la molécula hereditaria de la vida: el ácido desoxirribonucleico (DNA por sus siglas en inglés)¹⁷, la información que tenemos del funcionamiento molecular y genético de las células ha crecido de manera asombrosa. En las últimas décadas, el genoma de más de mil organismos ha sido secuenciado en su totalidad y se espera que este número crezca significativamente en los próximos años. Sin embargo, el conocimiento de las secuencias que componen cada uno de los genomas no es suficiente para explicar cómo la expresión de dicha información da lugar a bacterias capaces de adaptarse a una variedad de condiciones a través de diversos mecanismos; y mucho menos, para entender como un organismo multicelular, cuyas células comparten el mismo genoma, pueden dar lugar lo mismo a una neurona que a una célula de la piel o del tejido hepático.

La presencia de un gen en la molécula de DNA de un organismo no asegura la expresión del mismo. A pesar de que, en efecto, la instrucción para la elaboración de la molécula de ácido ribonucleico (RNA por sus siglas en inglés) y, en su caso, de la proteína, está codificada en la secuencia de DNA, existen mecanismos que coordinan la producción de éstas para que se lleve a cabo en un momento determinado y en las cantidades adecuadas. Uno de los mecanismos mejor estudiados y, que hasta la fecha, parece ser el más determinante en el control de dicha expresión, es la regulación de la transcripción, propuesto inicialmente por François Jacob y Jaques Monod en 1961⁸.

La transcripción es el proceso a través del cual una de las hebras del DNA se copia a una molécula de RNA, complementaria a la secuencia templado. La enzima encargada de llevar a cabo la transcripción es la RNA polimerasa dependiente de DNA (Fig. 1), que cataliza la síntesis de la cadena de RNA a partir de los sustratos nucleósidos trifosfato. El núcleo de esta enzima está compuesto por las subunidades $\alpha 2\beta\beta'\omega$, pero para que la RNA polimerasa pueda iniciar la transcripción en un promotor específico, requiere de una sexta subunidad denominada σ , formándose así la holoenzima $\alpha 2\beta\beta'\omega\sigma$.

La subunidad σ asegura el reconocimiento del promotor específico y posiciona a la RNA polimerasa en la molécula de DNA, con lo cual se inicia el proceso de transcripción. Una vez que la holoenzima se une al promotor, las dos hebras de DNA se desenrollan alrededor del sitio de inicio de la transcripción para formar una *burbuja de transcripción*, también llamado *complejo abierto*, que es en donde empieza a sintetizarse la nueva cadena de RNA. La RNA polimerasa cataliza la formación de enlaces fosfodiéster, a través de los cuales la cadena de RNA se mantiene unida. Este proceso continúa hasta que la enzima se encuentra con una nueva señal en el DNA: el terminador transcripcional. Es así como concluye la transcripción de un gen y la RNA polimerasa se separa de la molécula del DNA, lista para asociarse a otro factor sigma y unirse a un nuevo promotor.

Existen tres agentes básicos que participan en el control transcripcional de la expresión genética y que son particularmente relevantes para el presente trabajo: las secuencias promotoras, los factores σ y los factores transcripcionales (FTs). A continuación se describe brevemente a cada uno de ellos:

Secuencias promotoras

Un promotor es una secuencia de nucleótidos específica que presenta un gen particular río arriba de la secuencia codificante. Su papel es indicarle el punto de inicio a la RNA polimerasa para que se una el DNA e inicie el proceso de transcripción. A lo largo de este trabajo, no se hará referencia de manera recurrente a las secuencias promotoras. En su lugar se manejarán los nombres de los genes que las presentan y serán denominados *genes regulados* (GRs). La figura 2 muestra la estructura de un promotor bacteriano.

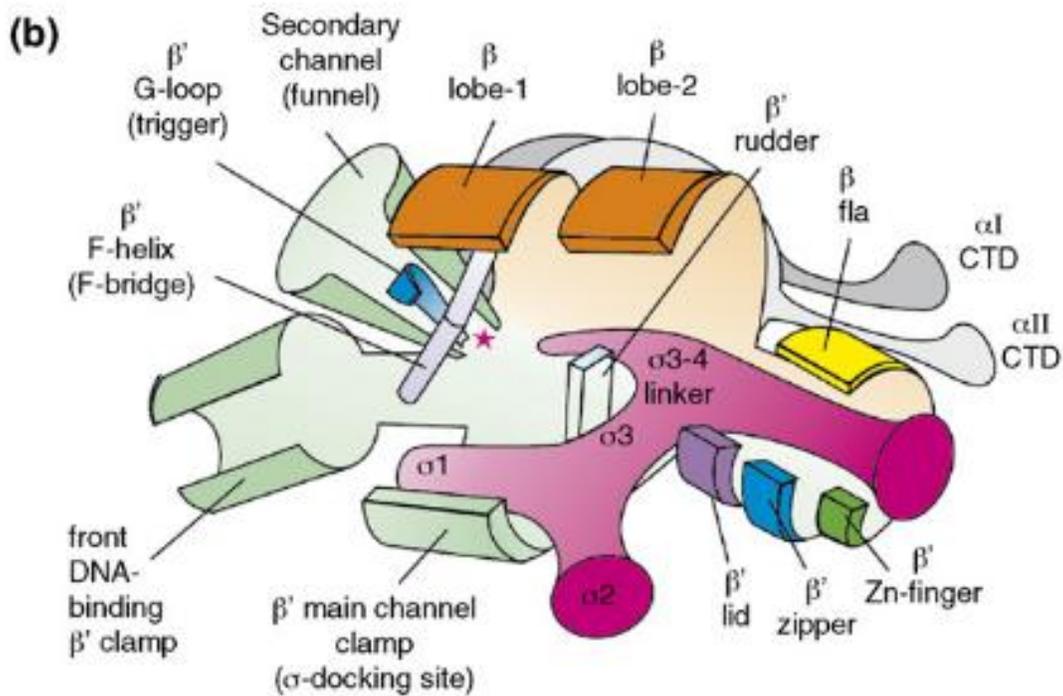
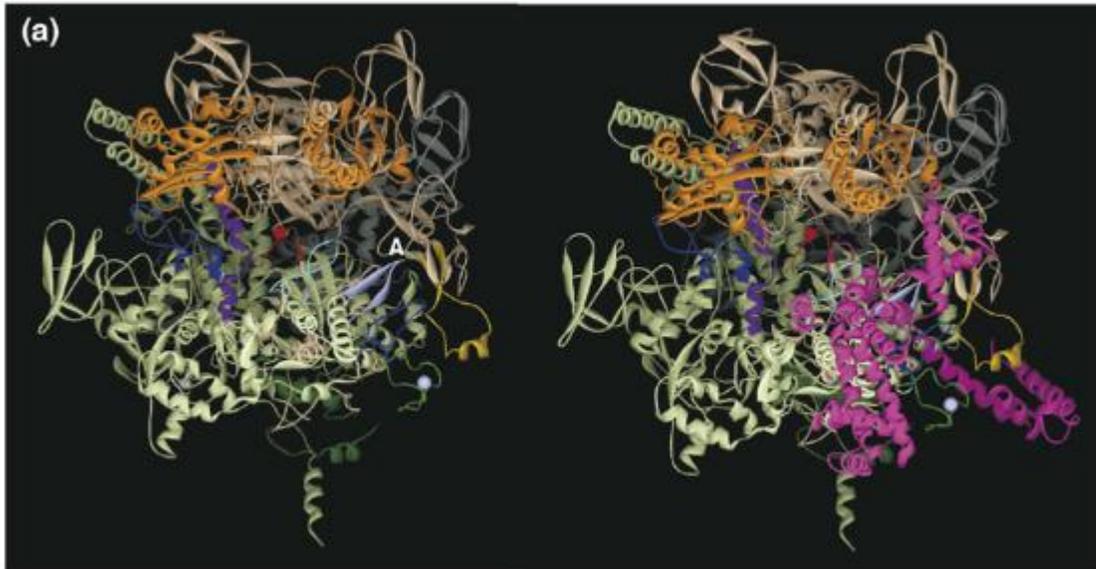


Figura 1. Estructura de la RNA polimerasa. **(a)** Estructuras cristalinas de alta resolución del núcleo de la RNA polimerasa de *Thermus aquaticus* (panel izquierdo) y de la holoenzima de *Thermus thermophilus* (panel derecho). Las subunidades se muestran como listones (α I, gris claro; α II, gris; β , café claro; β' verde claro; ω , verde oscuro; σ , magenta). El sitio catalítico β' de la RNA polimerasa se presenta coloreado de rojo y el ion Mg^{+2} se muestra como una pequeña esfera roja. **(b)** Representación esquemática de la estructura de la RNA polimerasa, con el mismo código de color que (a). El centro catalítico se muestra como una estrella pequeña color magenta. (Imagen tomada de Borukhov y Nudler, 2008.)

Factores σ

Se ha mencionado que la subunidad σ se asocia y disocia a la RNA polimerasa, y que de ella depende en gran medida que la holoenzima se una al promotor de un gen particular. Las bacterias presentan una diversidad de factores sigma, cuya producción también está transcripcionalmente regulada, y a su vez, dependiendo de la presencia de dicho factor, la RNA polimerasa se une o no a un promotor para transcribir un gen. Las células bacterianas regularmente presentan un factor sigma *housekeeping* y un número variable de factores sigma alternativos. El primero se caracteriza por presentar un dominio que da una mayor flexibilidad para el reconocimiento de promotores, mientras que los factores alternativos permiten la inducción de genes que se activan en respuesta a cambios particulares.

Factores transcripcionales

Uno de los tipos de proteína para los cuales codifican los genes son los denominados *factores transcripcionales*. Dichas proteínas participan en la regulación transcripcional de genes a partir de su estructura terciaria, la cual induce (regulación positiva) o reprime (regulación negativa) este proceso. En general, los reguladores negativos se unen al promotor o a una región posterior al mismo, interfiriendo directamente con la RNA polimerasa; mientras que, los reguladores positivos se unen regularmente a una región río arriba del promotor, con el objetivo de reclutar a la RNA polimerasa, para que se una al promotor e inicie la transcripción^{4,10}. Los FTs pueden regular la expresión de uno o muchos genes. Al conjunto de genes regulados por un FT se le denomina *regulón*.

La coactividad de la secuencia del promotor del gen regulado, en conjunto con el FT y la RNA polimerasa con su respectiva subunidad σ , constituye el sistema regulador básico en todos los organismos. Dicho sistema regulador se ejerce sobre una denominada unidad transcripcional (UT). En el caso de las bacterias, una UT está compuesta por las siguientes entidades genéticas: una región regulatoria que incluye al propio promotor, el sitio de inicio de la transcripción, uno o más genes y un sitio de término. Una UT puede presentar uno o más genes, denominada unidad monocistrónica en el primer caso y policistrónica si

presenta más de uno. En las bacterias, es muy común encontrarse UTs policistrónicas, y éstas son también conocidas como operones. Cabe mencionar que en ocasiones los genes de las UTs no codifican para proteínas, si no que lo hacen para RNAs ribosomales (rRNA), RNAs de transferencia (tRNAs), o bien, pequeñas moléculas de RNA reguladoras o microRNAs (miRNAs).

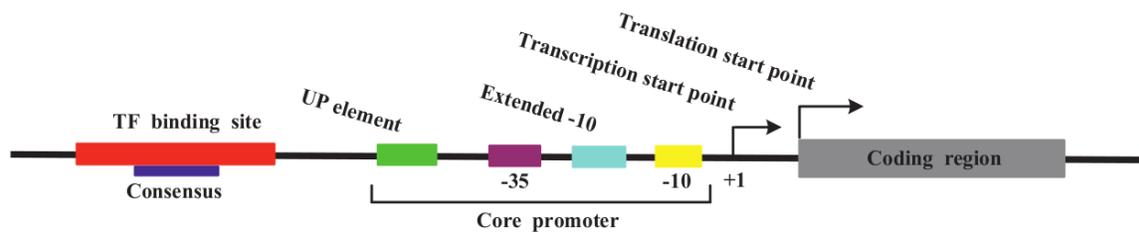


Figura 2. Estructura de un promotor procariótico. Un promotor es una región de DNA en el genoma en la cual la RNA polimerasa y los FTs implicados se unen para iniciar la transcripción. La posición +1 indica el par de base nucleotídica en donde se inicia la transcripción, comúnmente llamada sitio de inicio de la transcripción. Los pares de bases río arriba del sitio de inicio de la transcripción tienen asignados números positivos, mientras aquéllos que se ubican río abajo se muestran con números negativos. El núcleo del promotor consiste en los elementos UP, -10, -35 y -10 extendido. Los elementos -10, -35 y -10 extendido son reconocidos por los dominios 2, 4 y 3 de la subunidad σ de la RNA polimerasa, respectivamente. El elemento UP, localizado río arriba del elemento -35, es reconocido por los dominios C-terminales de las subunidades α de la RNA polimerasa. En ocasiones, el sitio de unión del FT puede traslaparse con el núcleo de la secuencia promotora. Una secuencia consenso es comúnmente localizada en el sitio de unión del FT. (Imagen tomada de Zhou y Yang, 2006.)

1.1.2. Estudiando la regulación de la transcripción de una manera global

A pesar de que han sido muchos los avances que se han logrado estudiando cada gen y entendiendo la manera en que es regulado, fue en la era genómica, debido al crecimiento en la cantidad de información, cuando empezaron a ser necesarias tecnologías que estudiaran la regulación de la transcripción de una manera más global. Diversas técnicas y experimentos a gran escala, tales como los microarreglos y la denominada tecnología ChiP-chip, que combina la inmunoprecipitación de la cromatina con los propios microarreglos, ofrecieron la posibilidad de estudiar la transcripción a grandes escalas.

Como resultado de lo anterior, fue posible empezar a construir modelos de estas interacciones, lo que a la par permitió entender las relaciones de los factores transcripcionales con sus genes regulados como una compleja red. Así, las funciones biológicas dejaron de ser atribuidas a un gen particular, idea que fue sustituida con una visión más holística, en la cual, más que el papel de una sola molécula, lo que interesa es la estructura y la dinámica de las interacciones entre los diversos componentes de la célula, que en conjunto, resultan en la expresión y coordinación de mecanismos y funciones fisiológicas. Diversas redes biológicas, tales como las de interacción proteína-proteína, las redes metabólicas, las redes de señalización, y las propias redes de regulación transcripcional, empezaron a estudiarse a través de la teoría de redes, desenmascarando propiedades y estructuras topológicas^a compartidas entre ellas y representativas de las redes biológicas.

^a La Topología es una disciplina matemática que estudia las propiedades de los cuerpos geométricos que permanecen inalteradas por transformaciones continuas. En una red, la topología se refiere a la arquitectura derivada de la manera en que los elementos que la conforman están conectados.

1.2. Teoría de redes

“... En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el Mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del imperio, toda una Provincia. Con el tiempo, esos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el Tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas.”

*Suárez Miranda: Viajes de varones prudentes,
libro cuarto, cap. XLV, Lérida, 1658*

(tomado de *Del Rigor de la Ciencia* de Jorge Luis Borges, 1954)

Una de las maneras más efectivas para entender mejor un mecanismo o para visualizar la posible solución de un problema es a través de la creación de un modelo. Esto equivale a simplificar y a hacer una representación del objeto o sistema original, que fue tal vez el punto en el que fallaron los cartógrafos del cuento de Borges. En las últimas décadas, la cantidad de información biológica disponible se ha visto incrementada a grandes velocidades y esto hace todavía más necesaria la integración de la información y la simplificación de la misma. La participación de las matemáticas y de las ciencias computacionales ha facilitado la creación de dichos modelos, siendo recientemente la aplicación de la teoría de redes una de las herramientas más útiles para representar y analizar un sistema tan complejo como lo es un ser vivo.

1.2.1. Los inicios de la teoría

La teoría de redes, también conocida como teoría de grafos, tuvo su origen en el siglo XVIII, siendo el matemático suizo Leonhard Euler considerado padre de la misma. Siendo ya un matemático reconocido de la época, a sus oídos llegó un problema que los ciudadanos de la ciudad de Königsberg, en Prusia, se habían estado planteando: dada la

arquitectura de su ciudad (Fig. 3a), a través de la cual fluía el río Pregel, formándose una isla en el centro y separando regiones de tierra a las que sólo podía llegarse cruzando los siete puentes construidos en Königsberg, ¿sería posible encontrar una ruta que permitiera caminar por todas las regiones de la ciudad cruzando una sola vez cada uno de los siete puentes?

Euler, como buen matemático, primero reemplazó el mapa de la ciudad por un diagrama más simplificado (Fig. 3b), que mostraba los elementos principales que se requerían para el planteamiento del problema. Sin embargo, posterior al diseño del diagrama, simplificó más la representación: ahora era un punto lo que equivalía a cada región, y una línea uniendo dos puntos lo que representaba a los puentes (Fig. 3c).

Esta segunda representación elaborada por Euler resulta el primer ejemplo de lo que ahora conocemos como *grafo* o *red*. El matemático publicó sus resultados en el artículo *Solutio problematis ad geometriam situs pertinentis*⁶, en el cual no sólo exponía este modelo que, de manera formal, demostraba que el problema no tenía solución; sino que además presentaba un método general para la representación y solución de problemas semejantes, sentando así muchos de los conceptos básicos de la teoría de grafos que hoy conocemos.

1.2.2. Las redes y sus propiedades

Un grafo o red permite hacer una representación abstracta de una serie de elementos conectados entre sí. La definición formal de grafo G contiene un conjunto de vértices o nodos V y un conjunto de aristas E conectando a estos nodos:

$$G = (V, E)$$

Las aristas que conectan a los nodos pueden o no tener dirección; es decir, en el primer caso se observa una diferencia en la participación de los nodos en la relación que los conecta, en donde uno afecta y el otro está siendo afectado; la ausencia de dirección refleja la participación equivalente de los nodos en la relación. Dependiendo si las aristas en una red presentan o no dirección, el grafo se define como dirigido o no dirigido, respectivamente. Un ejemplo de grafo no dirigido podría ser una red de amigos, de colaboraciones científicas o de localidades conectadas en un mapa como la que representa a

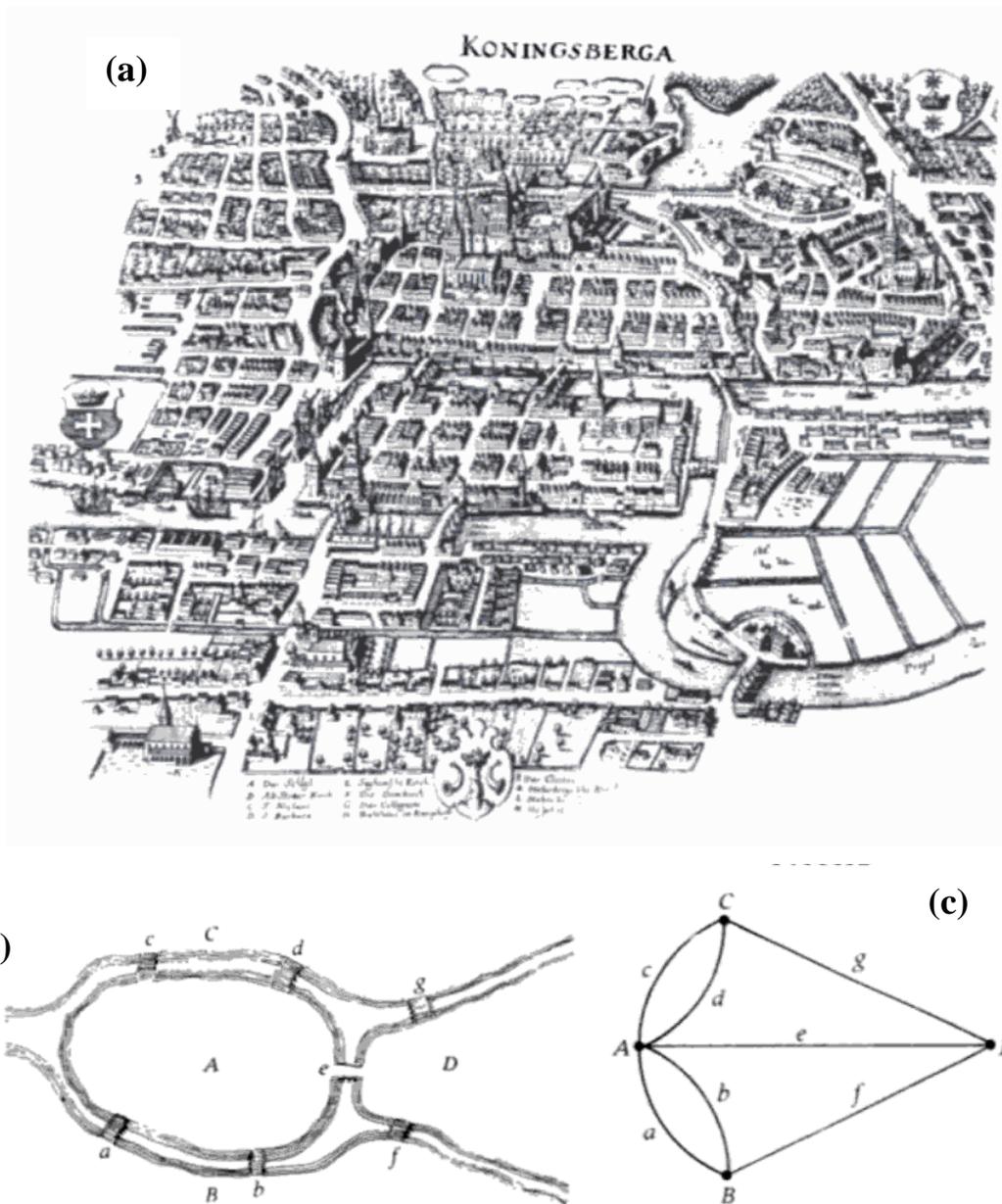


Figura 3. El problema de los puentes de Königsberg. **(a)** Grabado de la vieja ciudad de Königsberg en el este de Prusia, que muestra al río Pregel fluyendo a través de la ciudad. Como puede observarse, este río rodea a una isla, llamada Kneiphof, y en la parte derecha del mapa, se separa en dos vertientes. Para facilitar el movimiento de un lado a otro de la ciudad, se construyeron siete puentes. **(b)** Primer diagrama de la ciudad de Königsberg diseñado por Euler, en donde los espacios con letras mayúsculas representan regiones de tierra y las letras minúsculas denotan la presencia de cada uno de los siete puentes. **(c)** Representación más simplificada de la ciudad, elaborada por Euler y en la cual las letras corresponden al mismo código que en (b). La imagen (a) aparece en Zeiller 1650; (b) y (c) aparecen en Euler 1736. (Tomado de Biggs *et al.*, 1986.)

la ciudad de Königsberg (Fig. 3c). Un ejemplo de grafo dirigido podría ser una red de maestros y alumnos, o de páginas web que presentan vínculos de otras páginas web. La figura 4 muestra un ejemplo de red dirigida.

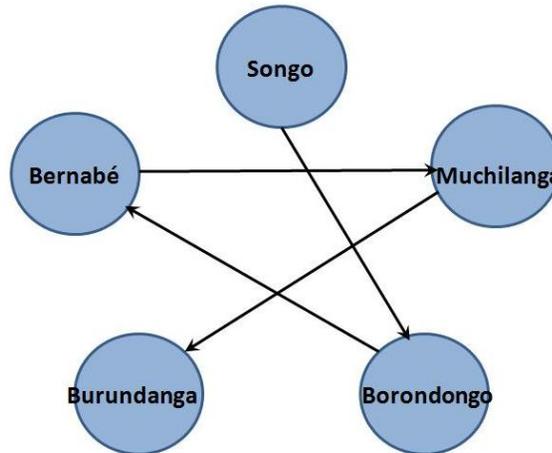


Figura 4. Ejemplo de red dirigida, en la cual dos nodos conectados por la misma arista se ven afectados de manera diferente. Uno lleva a cabo la acción y el otro la recibe. (Ejemplo tomado de Espinosa Soto, 2007.)

Además de estas características, las redes presentan una serie de propiedades que definen su topología. A continuación se resumen los principales parámetros estadísticos que presentan los nodos o vértices de una red.

- **Conectividad (k_i):** En una red, la conectividad de un nodo i se refiere al número de aristas con las que éste se une con otros nodos, sus vecinos, y se expresa con el símbolo k_i (Fig. 5)
- **Interacciones entre nodos vecinos (n_i):** Los vecinos de un nodo i , a su vez, pueden estar conectados entre sí. El número de aristas observadas entre ellos -de un máximo de $k_i(k_i - 1)/2$ para una red no dirigida- se cuantifica con el parámetro n_i (Fig. 6).

- **Coefficiente de agrupamiento (C_i):** Se puede obtener una medida de compactación en la localidad de un nodo i conociendo qué tan conectados entre sí están sus vecinos. Esta medida se obtiene calculando la razón del número de aristas observadas (n_i) entre el número de aristas posibles, y se expresa con el término C_i (Fig. 6).

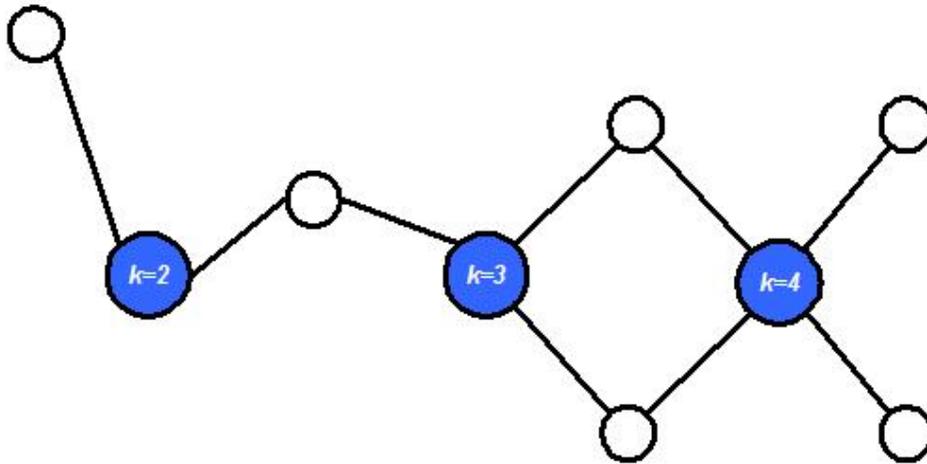


Figura 5. Representación gráfica de la conectividad (k) de un nodo. En esta figura, se señalan la k de los nodos de color azul.

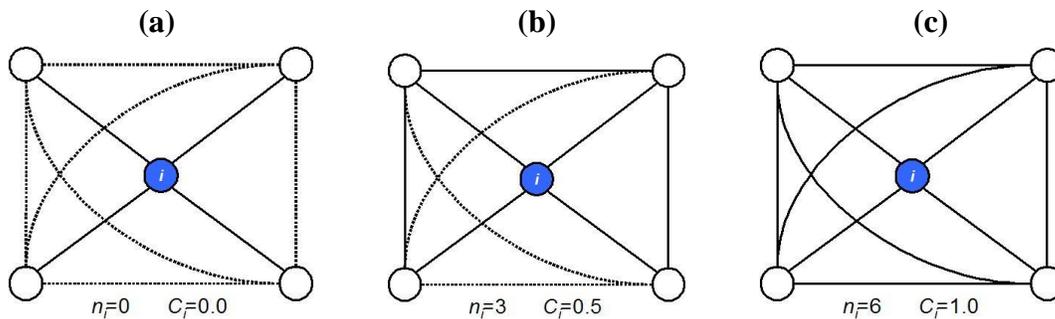


Figura 6. Representación gráfica de las interacciones entre nodos vecinos (n_i) y coeficiente de agrupamiento de un nodo i (C_i). Las líneas punteadas señalan todas las posibles conexiones entre nodos vecinos, mientras que las continuas describen las aristas observadas en la red. (a) Ninguno de los nodos vecinos está conectado con otro nodo vecino. (b) La red presenta la mitad de las conexiones posibles entre vecinos. (c) Todos los nodos vecinos están conectados entre sí, observándose así, todas las aristas posibles.

A partir de estos parámetros, la red como sistema, presenta las siguientes propiedades:

- **Distribución de la conectividad [$P(k)$]**
Probabilidad de que un nodo tenga k aristas.
- **Promedio del coeficiente de agrupamiento**
Promedio del coeficiente de agrupamiento (C) de todos los nodos de la red que tienen un valor $k > 1$. Este parámetro es una medida de la modularidad potencial de la red.
- **Distribución del coeficiente de agrupamiento [$C(k)$]**
Frecuencia de aparición de valores de C dependiendo de k_i .
- **Distribución de vecinos más cercanos [$T(n)$]**
Probabilidad de que los vecinos del nodo i tengan n aristas que los conectan entre sí.

El camino mínimo promedio es otro parámetro relevante para caracterizar una red. Un camino mínimo entre un par de nodos (i, j) es el número mínimo de aristas que se necesitan para llegar de i a j . El promedio involucra los caminos mínimos de todos los pares de la red.

1.2.3. Redes aleatorias

En 1960, los matemáticos Paul Erdős y Alfréd Rényi propusieron el modelo de las redes aleatorias, el cual asume que una red tiene un número determinado de nodos y que éstos están conectados de manera azarosa entre ellos⁵ (Fig. 7Aa). A partir de la construcción de este tipo de redes, se empezaron a definir las propiedades topológicas que todas ellas presentaban en general.

En una red aleatoria, la probabilidad de que un nodo tenga una conectividad (k) es la misma para todos. Así, la [$P(k)$] que presentan estas redes sigue una distribución de Poisson (Fig. 7Ab), lo que indica que un gran número de nodos tiene aproximadamente el mismo valor de conectividad. Respecto al coeficiente de agrupamiento [$C(k)$], la probabilidad de que los vecinos de un nodo estén conectados entre sí equivale a la

probabilidad de que dos nodos al azar estén conectados; por lo tanto los valores de la distribución de los valores de C son independiente de k (Fig. 7Ac).

1.2.4. Redes complejas

En contraste con las redes aleatorias, se ha demostrado que las redes que representan sistemas reales en la naturaleza y en la sociedad muestran características diferentes. Dentro de éstas, se encuentran las redes biológicas, incluyendo las redes de regulación transcripcional, de particular interés para el presente trabajo. A continuación se mencionarán las propiedades topológicas que se han observado en este tipo de redes:

- *Mundo pequeño.* Esta propiedad está ligada con el concepto de camino mínimo, ya descrito. Las redes que la presentan tienen un camino mínimo promedio de bajo valor, lo que resulta en que, escogiendo dos nodos de la red al azar, es probable que sean pocas las aristas necesarias para llegar de uno al otro¹⁸.
- *Libre de escala.*- En las redes complejas, la conectividad no es uniforme como en el caso de las aleatorias. En contraste, la distribución de la conectividad [$P(k)$] sigue una ley de potencia inversa, con exponentes de valor entre el 2 y el 3 (Fig. 7Bb). Esto significa que muchos nodos están conectados con uno o dos vértices, pero pocos son aquellos que presentan una k grande; a estos últimos también se les conoce como *hubs* (Fig. 7Ba)².
- *Modularidad.*- Esta propiedad define a una red que presenta grupos de nodos altamente conectados entre sí, pero con pocas, y en casos nulas, conexiones hacia afuera del grupo. La modularidad es la responsable de que una red presente un promedio de agrupamiento (C) alto¹⁸. Resultados empíricos han demostrado que el promedio de C en redes complejas es significativamente más alto que en redes aleatorias del mismo tamaño y que, en general, este valor es independiente del número de nodos en una red¹.
- *Jerarquía.*- Las redes jerárquicas presentan las dos últimas propiedades descritas. A pesar de que una red modular puede no ser libre de escala, ambas características no son excluyentes. Para explicar que una red sea libre de escala y modular, se asume entonces que los módulos se combinan uno con otro de manera jerárquica. Esto puede observarse en una gráfica que describa la distribución del coeficiente de

agrupamiento de manera dependiente a k . Este tipo de redes se caracteriza por una $C(k) \sim k^{-1}$ (Fig. 7Cc), lo cual implica que aquellos nodos con baja conectividad tienden a ser parte de grupos interconectados, mientras algunos de los denominados *hubs* juegan el papel de conectar dichos módulos (Fig. 7Ca)^{13,14}.

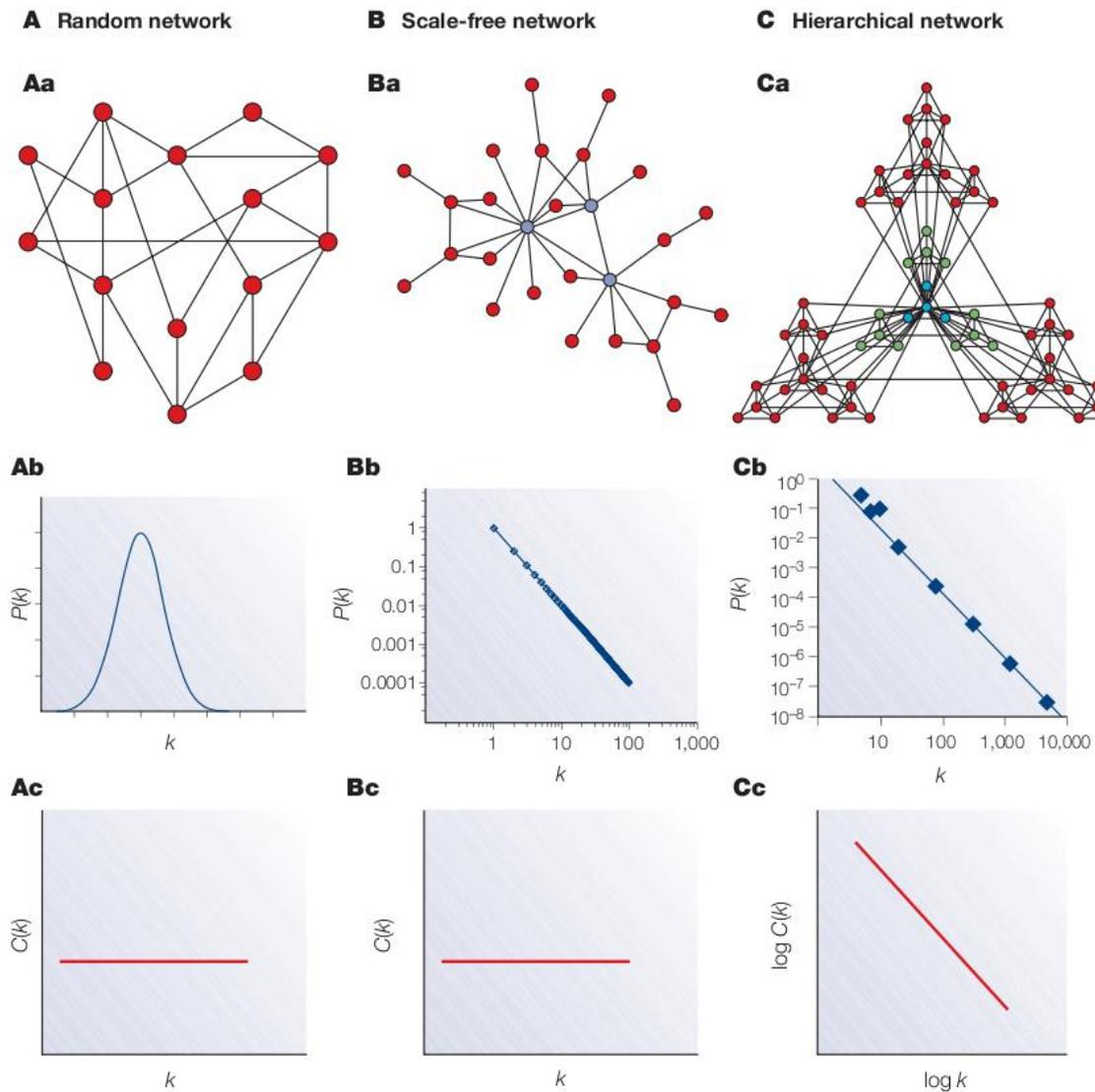


Figura 7. Tres modelos de redes con sus respectivas propiedades topológicas. (A) Redes aleatorias. (B) Redes libres de escala. (C) Redes jerárquicas. Las letras minúsculas asociadas a cada modelo denotan: (a) ejemplo de una representación gráfica del respectivo tipo de red, (b) distribución de conectividad [$P(k)$] y (c) coeficiente de agrupamiento [$C(k)$]. (Tomado de Barabási y Oltvai, 2004.)

1.2.5. La importancia del concepto de modularidad en Biología

En la última década, el concepto de modularidad en los sistemas biológicos ha tomado una gran importancia. Un módulo se define como una entidad discreta, en la cual sus componentes interactúan entre sí y de manera relativamente autónoma e independiente de los componentes de otras entidades, coordinando así una función particular. La modularidad puede darse a muchos niveles y de muchas formas, como lo es la modularidad física, química y espacial⁷.

En Biología, los módulos observados en sistemas vivos, son considerados como agrupamientos primarios de la organización celular, los cuales llevan a cabo una función fisiológica particular. El análisis de estos sistemas a partir de la teoría de redes, ha permitido estudiar esta propiedad utilizando métodos cuantitativos. Los módulos topológicos, siguiendo el concepto de modularidad en general, se caracterizan por contener nodos altamente conectados entre sí, pero relativamente aislados de otros módulos. En la sección anterior se mencionaron algunos de los parámetros topológicos a partir de los cuales se pueden definir módulos discretos en las redes biológicas. Se espera también que estos módulos topológicos desempeñen una función particular dentro del ambiente celular.

Sin embargo, es importante mencionar que los módulos no tienen que ser necesariamente rígidos. La modularidad no equivale a la coexistencia de grupos totalmente independientes de otros nodos, sino que es la formación de pequeños agrupamientos densamente conectados, que se combinan para formar grupos más grandes pero menos cohesivos, y así sucesivamente. De esta manera un módulo comparte información con otros módulos, al igual que una función puede influenciar a otra, como se observa que sucede en los sistemas reales.

Actualmente la idea de que la modularidad existe y de que es biológicamente significativa es ampliamente aceptada. Además, con la visión holística que ha permeado en las últimas décadas en el área del estudio de la vida, uno de los retos de la misma es seguir buscando un entendimiento global del funcionamiento de la célula. Así, el concepto de modularidad en las redes biológicas seguirá siendo de gran importancia para redefinir la noción de función en Biología a través de la relación módulo-función.

1.2.6. Motivos topológicos

El estudio topológico de las redes complejas ha dado como resultado el desarrollo de otro concepto: los motivos topológicos. Éstos son patrones de interconectividad que se presentan en las redes complejas con una frecuencia significativamente más alta que en redes aleatorias^{11,15}.

En las redes dirigidas, estos motivos pueden estar compuestos por uno o más nodos. La figura 8 muestra todos los posibles patrones de conexión posibles con tres nodos. En el caso de las redes de regulación transcripcional, el subgrafo 5 de esta misma figura es el que se ha visto sobrerrepresentado en este tipo de redes, y es también conocido como *feedforward*, también ejemplificado en la figura 9. Otros motivos, implicando un mayor número de nodos, también se han observado en redes de transcripción, tales como *bi-fan*, en donde dos genes corregulan otros dos genes; módulo de un solo input (SIM, por sus siglas en inglés); y regulones densamente traslapados (DOR, por sus siglas en inglés).

El estudio teórico y experimental de los motivos topológicos en las redes de regulación transcripcional, ha permitido entender mejor el papel que juega cada uno de ellos en la organización de la red. En el caso del motivo *feedforward*, uno de los más estudiados, se ha especulado que una de sus posibles funciones es inducir la expresión del gen regulado si la señal del FT input (el que está regulando a los otros dos genes) es persistente, y por el contrario, permitir la rápida desactivación de la misma cuando el input no se esté expresando¹⁵. Adicionalmente, esta estructura provee control temporal de un proceso, ya que la expresión del gen Z depende de la acumulación adecuada de los genes reguladores X y Y (Fig. 9).

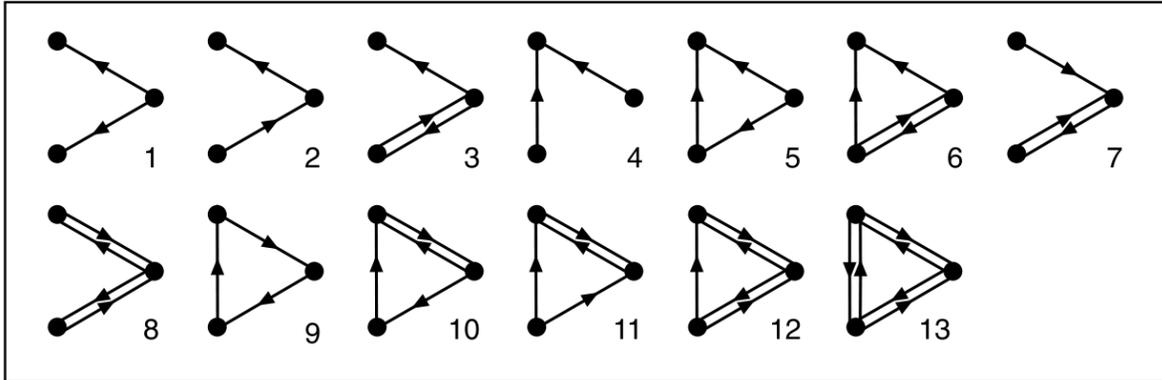


Figura 8. Los 13 posibles patrones de conexión con tres nodos en una red dirigida. (Tomado de Milo *et al.*, 2002.)

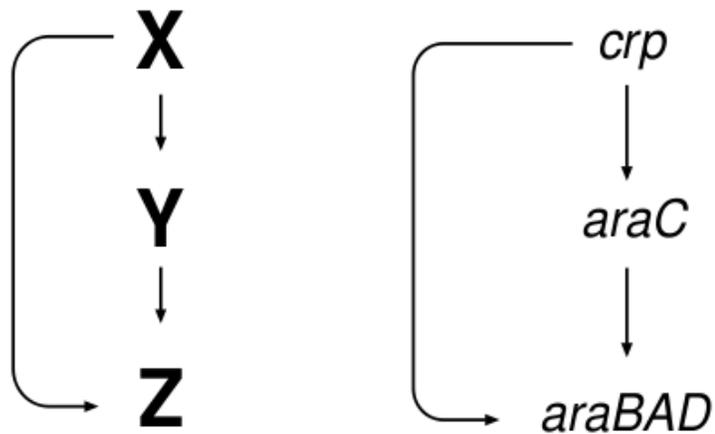


Figura 9. Motivo *feedforward*, que en redes de regulación transcripcional representa un FT X que regula a un segundo FT Y , y ambos regulan la expresión de un gen u operón Z (Tomado de Shen-Orr *et al.*, 2002).

1.3. *Bacillus subtilis* como modelo de estudio

The species name subtilis comes from the Latin for “fine, thin, delicate”. In 14th century French, it morphed into subtil, meaning “clever, dexterous”. We think that the evolution of the word mimics the evolution of the organism. NOMEN EST OMEN.

(Moselio Schaechter & Merry Youle, 2008)

1.3.1. Acerca de *B. subtilis*...

A pesar de que se tiene noción de que fue Ehrenberg el primer científico en describir a este organismo con forma de bacilo que ahora conocemos como *Bacillus subtilis*, entonces designándole el nombre de *Vibrio subtilis*; fue Cohn quien en 1875 renombró a la especie como la llamamos actualmente³. La figura 10 muestra algunos de los dibujos creados por ambos científicos a partir de la observación de este microorganismo. Como puede observarse, ya entonces Cohn señalaba la capacidad de formar esporas que caracteriza a *B. subtilis*, proceso que aún sigue maravillándonos.

Ahora, a más de un siglo después de su descubrimiento, *B. subtilis* es una de las bacterias mejor caracterizadas. Está ubicada en el dominio Eubacteria, y pertenece al *phylum* Firmicutes, grupo formado principalmente por bacterias Gram positivas, las cuales presentan una célula procariótica que carece de membrana externa y, cuya pared celular, está compuesta fundamentalmente por petidoglicano. Algunos de los géneros ubicados en este grupo, incluyendo a *Bacillus*, presentan además un bajo contenido de nucleótidos GC en su genoma.

Esta bacteria Gram positiva inicialmente fue denominada una bacteria de suelo; sin embargo, ahora se sabe que tiene la capacidad de crecer en muchos y muy diversos ambientes, tanto terrestres como acuáticos. Incluso, se conoce que puede también vivir asociada a la superficie de raíz de algunas plantas o en el tracto gastrointestinal de algunos animales. Además, es un microorganismo aerobio facultativo.

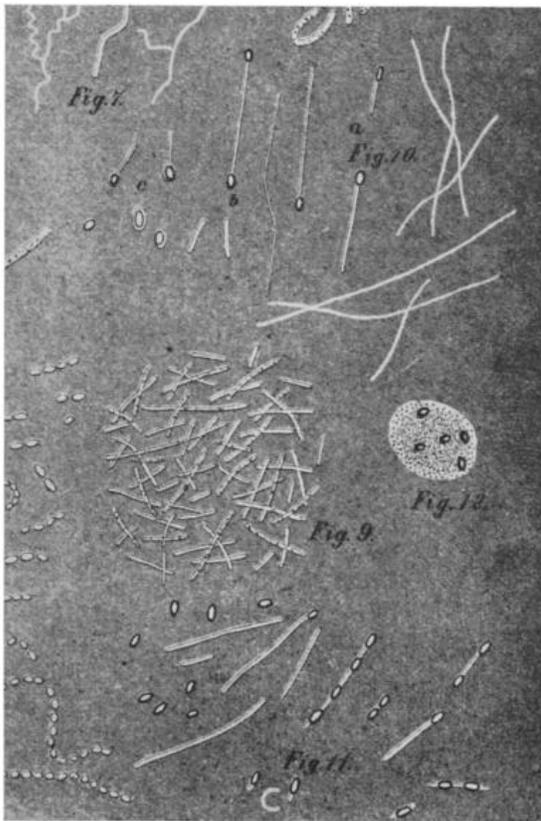
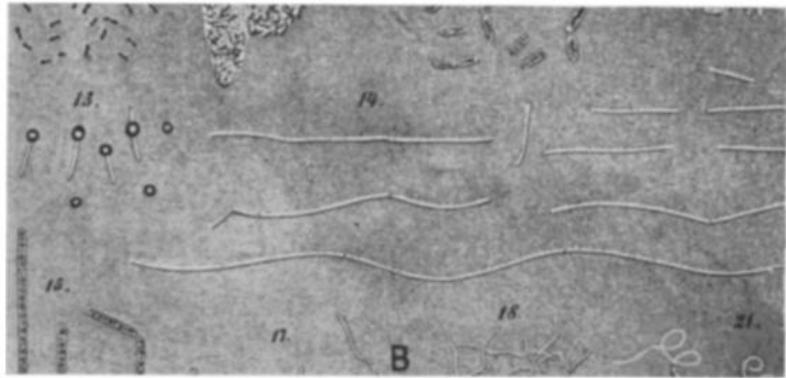
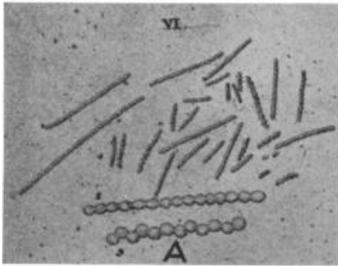


Figura 10. Primeros dibujos de *Bacillus subtilis* de los cuales se tiene conocimiento. (A) *Vibrio subtilis*, realizado por Ehrenberg. (B: 13 y 14) *B. subtilis* mostrando largas hebras y esporas; realizadas por Cohn. (C: 10, 11 y 12) *B. subtilis* mostrando largas hebras con esporas ovaladas; realizadas por Cohn. (D: 5a) El esquema representa la germinación de esporas en *B. subtilis*. (Tomado de Soule, 1932.)

Su capacidad para adaptarse a diversos ambientes se ha atribuido principalmente a la formación de esporas, proceso que lleva a cabo bajo determinadas condiciones de estrés y falta de nutrientes. Éstas son conocidas también como endosporas porque se desarrollan dentro de la célula madre. Son fácilmente dispersadas en el aire, lo cual permite que migren grandes distancias y germinen hasta que encuentren las condiciones adecuadas para hacerlo. La figura 11 ilustra el ciclo de esporulación en *B. subtilis*.

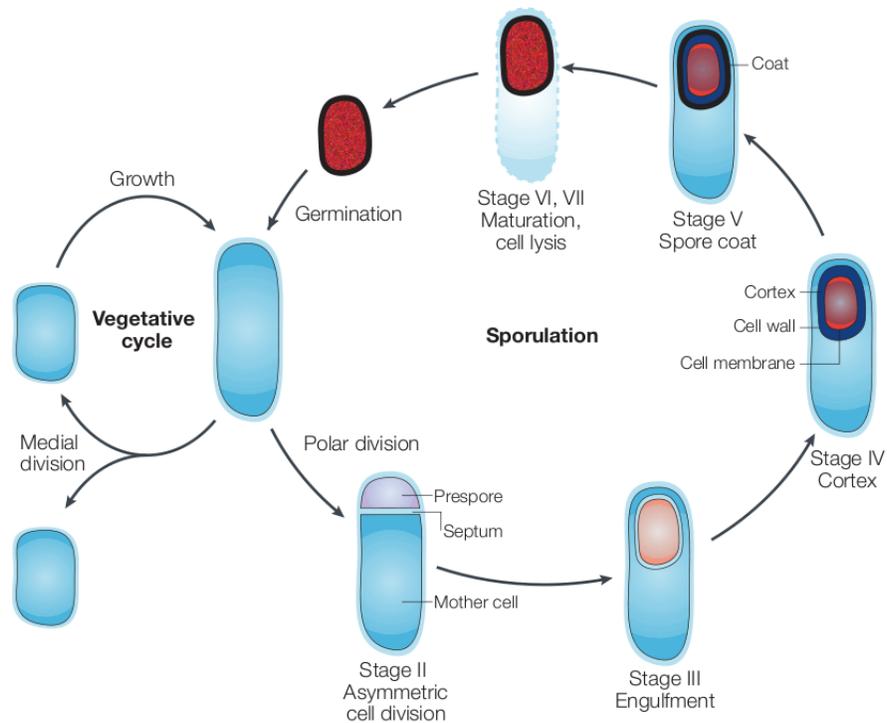


Figura 11. Ciclo de esporulación de *Bacillus subtilis*. Este esquema simplificado muestra sólo las etapas clave de este ciclo. (Tomado de Errington, 2003.)

Sin embargo, a pesar de que el proceso de formación de endosporas ha sido ampliamente estudiado, actualmente se reconoce que *B. subtilis* también tiene la capacidad para diferenciarse en otros tipos celulares, tales como células genéticamente competentes, células móviles, células productoras de matriz extracelular, células productoras de enzimas degradativas y células caníbales (Fig. 12). El concepto de células competentes se refiere a aquéllas que se encuentran en un estado fisiológico en el cual son capaces de tomar DNA

exógeno e integrarlo a su genoma; las células motiles, por su parte desarrollan esta capacidad a través de la formación de un flagelo y la expresión de un sistema de quimiotaxis. *B. subtilis* forma además comunidades microbianas denominadas *biopelículas*, en las cuales las células productoras de matriz extracelular juegan un papel muy importante, al ser las responsables de mantener unida esta comunidad; mientras que las productoras de proteasas, permiten la degradación de proteínas en el medio para que en una comunidad todas se beneficien con ello. El comportamiento caníbal se observa regularmente previo a la esporulación, y las células que lo expresan secretan toxinas que matan a células vecinas, las cuales devoran posteriormente. Muchos de estos tipos celulares conviven juntos en las biopelículas que forma esta especie. Esto hace que *B. subtilis* sea un organismo interesante para estudiar la diferenciación, no sólo por el proceso de esporulación, sino por los otros ya mencionados.

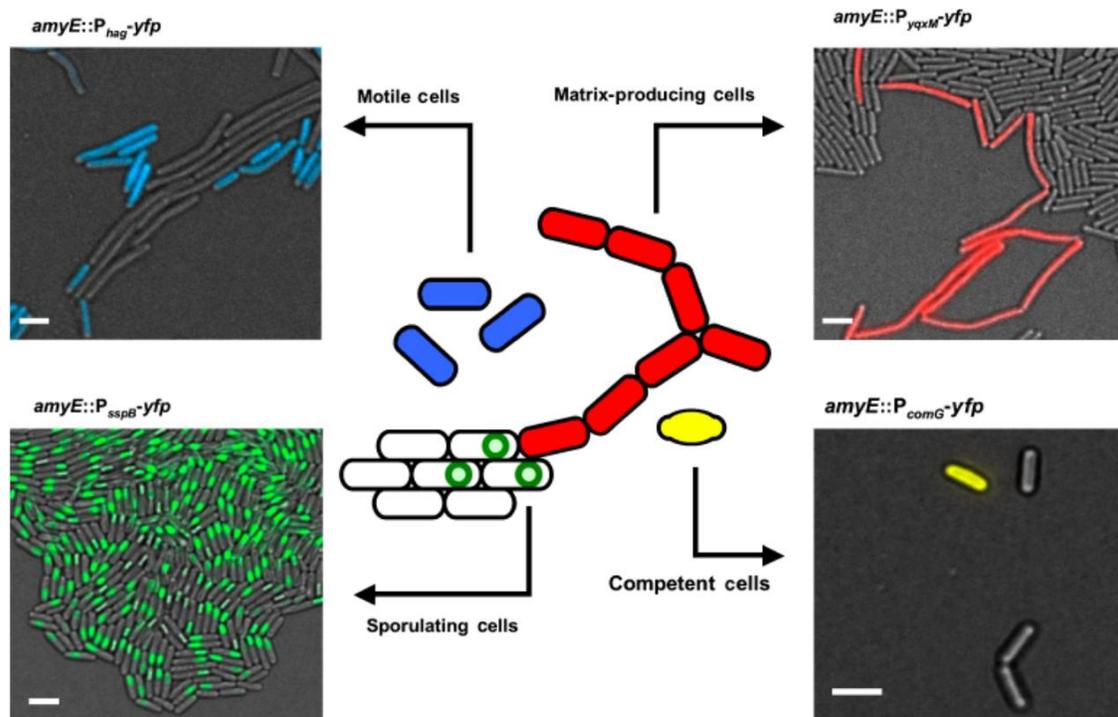


Figura 12. La diferenciación de diversos tipos celulares en *B. subtilis* puede ser monitoreada a través de la fusión con genes reporteros en cada uno de los tipos celulares. En esta imagen el color azul indica células motiles; el rojo, células productoras de matriz; el verde, células en esporulación; y el amarillo corresponde a células competentes. Cada tipo celular presenta un fenotipo característico que la distingue del resto de la población (Tomado de López *et al.*, 2009.)

1.3.2. El genoma y su regulación

Siendo la bacteria Gram positiva mejor caracterizada, el genoma de *B. subtilis* fue el primero de su tipo en ser secuenciado, en 1997⁹. Los resultados de este trabajo arrojaron una gran cantidad de datos: el genoma está compuesto por 4,214,630 pares de bases, codificando para 4105 proteínas, y presentando un 43.52% de contenido GC, consistente con su clasificación en el grupo de las Gram positivas. Sin embargo, más allá de los números, la secuenciación de este genoma ofreció una nueva ventana para el estudio de la Biología de la bacteria, ya que confirmó y predijo algunas de sus características fisiológicas. Se confirmó la ausencia de genes que codifican factores de virulencia, consistente con el hecho de que no se le conoce como patógeno; así también, se encontraron genes que putativamente codificaban para nitrato reductasas, lo que predijo que podría crecer de forma anaeróbica, demostrado experimentalmente poco tiempo después.

A partir de entonces, esta bacteria ha sido entendida aún mejor a través de su genoma. Múltiples trabajos acerca de su regulación han sido publicados, notándose un especial interés por entender el mecanismo de regulación a través del cual esporula y germina. A pesar de que los factores sigma están presentes en todas las bacterias, en *B. subtilis* se ha observado la notable importancia de su papel. El factor sigma *housekeeping* de esta bacteria es llamado σA , ortólogo de $\sigma 70$ de *Escherichia coli*, y hasta ahora se conoce que regula a más de 750 genes. Además de éste, *B. subtilis* cuenta con otros 17 factores sigma alternativos, entre los cuales están: σE , σF , σG , σH y σK , coordinadores de la esporulación; σB , el regulador maestro de respuesta a estrés; y σM , σX y σW , involucrados en la modificación de la composición y propiedades de la envoltura celular.

Parte de este trabajo es analizar la red de regulación transcripcional a la luz del análisis topológico de la misma, por lo que a lo largo del mismo se abundará sobre los mecanismos que presenta *B. subtilis* para adaptarse al medio y cómo éstos se ven reflejados en la organización de su red de regulación.

A más de una década de la secuenciación de su genoma, existen algunas bases de datos de esta bacteria. *Subtilist* integra todos los genes conocidos de *B. subtilis*¹², mientras que DBTBS es una base de datos de la regulación transcripcional de esta bacteria, conteniendo listas de factores transcripcionales, operones, promotores y terminadores¹⁶.

2. ANTECEDENTES

En los últimos años, gracias al desarrollo de técnicas y herramientas que permiten el estudio global de la regulación de la transcripción, ha sido posible generar colecciones importantes de datos que nos describen las relaciones que existen entre los factores transcripcionales y sus genes regulados. Esto ha sido particularmente notorio en organismos modelo, como lo son los procariontes *E. coli* y *B. subtilis*, así como para organismos eucariontes, entre ellos la levadura *Saccharomyces cerevisiae*.

Lo anterior ha sido además útil para incursionar en análisis globales y generar modelos que describan la organización y la relación que existe entre los elementos que componen las redes de interacciones biológicas. La reconstrucción de éstas a partir de la teoría de grafos ha sido particularmente recurrente en la última década. Se han estudiado principalmente redes metabólicas de organismos unicelulares, tales como *Haemophilus influenzae*¹⁶⁸, *Helicobacter pylori*¹⁶⁷, *S. cerevisiae*⁵⁹ y *E. coli*¹⁶⁰. Asimismo, se ha hecho la reconstrucción de las redes de regulación de diversos organismos, siendo particularmente relevante la de *E. coli*, la cual ha generado una gran cantidad de trabajos debido a que su maquinaria de regulación transcripcional es quizá la mejor caracterizada, a la vez que está organizada en bases de datos muy completas y actualizadas constantemente⁶⁸, lo que facilita la construcción de modelos.

Se ha sugerido en múltiples trabajos^{4,77,93}, que las redes de regulación transcripcional presentan una topología libre de escala, lo que implica la presencia de algunos *hubs*, o reguladores globales¹²³ y un mayor número de genes menos conectados. Los estudios topológicos han demostrado también, que las redes biológicas presentan modularidad jerárquica^{158,159}, con lo cual se apoya el nuevo paradigma de que los sistemas biológicos llevan a cabo sus funciones a partir de una organización celular modular⁸³.

Los trabajos de Barabasi A.L. y colaboradores^{5,14,158,159}, que han derivado en la idea de que todas las redes biológicas, incluyendo las de regulación transcripcional, presentan las mismas propiedades topológicas, ha resultado inspiradora para muchos grupos de investigación. Desde entonces, se han hecho un número notable de investigaciones que descomponen a las redes en módulos. Para el presente trabajo, es de particular relevancia

un estudio llevado a cabo por Resendis-Antonio O. *et al.*¹⁶¹, en el cual estudian las propiedades topológicas de la red de regulación de *E. coli*, determinando que es una red libre de escala y que presenta una organización jerárquico-modular. A partir de este resultado, tomaron la red de FTs para llevar a cabo un análisis modular. Los componentes de los módulos obtenidos fueron clasificados tomando como base las funciones descritas en la literatura y con las clases funcionales descritas por Riley M. y colaboradores ubicadas en la base de datos GenProtEC¹⁷². A partir de lo anterior, asociaron una función a cada módulo.

La red utilizada para el trabajo realizado por Resendis-Antonio O. *et al.* contenía sólo FTs y eliminaba a los GRs y factores sigma. Posteriormente Freyre-González J.A., toma la red reportada por Resendis-Antonio O. *et al.* y lleva a cabo de nuevo el análisis, ahora incluyendo a todos los blancos de cada FT. Su resultado muestra, que las clases funcionales encontradas quedan diseminadas a lo largo de todos los módulos, haciendo imposible analizar la red completa a través del mismo método⁶⁰. Por lo tanto, propone un método de descomposición natural, que además de incluir FTs, factores sigma y GRs, se fundamenta en una división basada en las propiedades jerárquico-modulares de las redes⁶¹.

A pesar de que *B. subtilis* es otro modelo importante en microbiología, y de que su red de regulación transcripcional ha sido ampliamente estudiada, una revisión exhaustiva de la literatura nos reveló que prácticamente no hay análisis de este tipo aplicados a la red de esta bacteria. En general sólo se encuentra incluida en análisis multigenómicos, en los cuales la idea básica es la comparación de la propiedades y conservación de la red de *B. subtilis* versus otros microorganismos^{121,125,185}. En la literatura, sólo se encuentra reportado un trabajo cuyo enfoque es únicamente el estudio de la red metabólica y genética de *B. subtilis*. Los autores reconstruyen la red de este microorganismo en un estudio que, aunque involucra a la transcripción, traducción y regulación postraduccional, está orientado al metabolismo central de esta bacteria⁷⁰. Además el método para llevar a cabo el análisis modular sobre la red consiste básicamente en definir la división de los módulos a partir de la presencia de flujos enzimáticos. Por lo tanto, dicho análisis, además de no involucrar toda la red transcripcional conocida hasta ahora, no determina la modularidad a través de un método de agrupamiento que tome las interacciones como base. No hay tampoco en la literatura ningún otro análisis modular de este tipo para *B. subtilis*. Recientemente, se tiene

conocimiento de un trabajo llevado a cabo por Freyre-González J.A. y colaboradores, en el cual se analiza la red de *B. subtilis* con el mismo método de descomposición utilizado para *E. coli*, previamente mencionado⁷. Sin embargo, a pesar de que se estudia la organización modular de la red, dicho análisis utiliza un método y un enfoque diferente al presentado en este estudio.

Así, esta tesis toma como modelo uno de los trabajos antes descritos de *E. coli*¹⁶¹ sobre el análisis topológico de su red de regulación, con el fin de realizar un análisis del mismo tipo acerca de *B. subtilis*. Tomando además en cuenta las observaciones hechas por Freyre-González J.A.⁶¹, se decidió tomar un esquema intermedio que contiene tanto FTs como factores sigma, eliminando a los GRs, estos últimos con relevancia sólo para la obtención de las propiedades topológicas, así como para el análisis de la Biología de los módulos.

3. OBJETIVOS, JUSTIFICACIÓN e HIPÓTESIS

3.1. Objetivos

Objetivo General

Analizar la red de regulación transcripcional de *Bacillus subtilis* desde un enfoque topológico.

Objetivos Particulares

1. Hacer una reconstrucción de la red de regulación transcripcional de *B. subtilis*.
2. Determinar las propiedades topológicas de la red de regulación de *B. subtilis*.
3. Realizar un análisis modular de la red de FTs de *B. subtilis*.
4. Hacer un análisis detallado de la Biología de los módulos encontrados en la red y determinar si se les puede asociar con una función fisiológica.

3.2. Justificación

En los últimos años, la aplicación de la teoría de redes al estudio de los sistemas biológicos, ha demostrado que éstos comparten propiedades topológicas que han empezado a ayudarnos a entender los mecanismos generales que han sido seleccionados a lo largo de la historia evolutiva. Una de estas propiedades es la modularidad, que nos permite descomponer a la red en entidades cohesivas y menos conectadas con el resto de la red. La importancia de esta propiedad y su asociación con posibles roles fisiológicos en los sistemas biológicos son de trascendencia para redefinir el concepto de función a partir de la relación módulo-función. De particular relevancia resulta la aplicación de este tipo de estudios a redes de regulación transcripcional, que son las responsables de controlar la expresión de genes frente a una condición particular. A pesar de que se han hecho análisis topológicos de las redes de diferentes organismos modelo, *Bacillus subtilis* ha estado ausente en ellos. Dado que es la bacteria Gram positiva mejor caracterizada y el organismo modelo de este grupo de bacterias, resulta de gran importancia entender la organización de su red de regulación transcripcional, lo que además facilitará el análisis y entendimiento de los procesos fisiológicos de respuesta que expresa frente a diversas condiciones y señales.

3.3. Hipótesis

Es posible determinar las propiedades estadísticas de la red de regulación de *B. subtilis* que nos permitan definir si la red es modular y jerárquica. De poseer las dos propiedades antes mencionadas, se predice la existencia de módulos topológicos que nos permitan entender las relaciones que existen entre la organización estructural de la red y la función biológica de los componentes encontrados.

4. METODOLOGÍA

4.1. Reconstrucción de la red de regulación transcripcional de *B. subtilis*

La red de regulación transcripcional de *B. subtilis* utilizada para este trabajo se obtuvo a través de la base de datos DBTBS v.5 (2008)¹⁷⁸, facilitada y depurada a través de su evidencia experimental por el Dr. Julio A. Freyre-González del Instituto de Biotecnología de la UNAM.

La tabla que agrupa la información requerida para la elaboración de este trabajo se muestra en el Apéndice 1, que presenta las siguientes tres columnas:

1. Factor transcripcional (FT)

En esta categoría están incluidos los denominados FTs, así como también los factores sigma, por su papel en la regulación transcripcional, como se describió en la introducción.

2. Gen regulado por dicho FT (GR)

3. Tipo de interacción:

- a. Positiva (+).- El FT induce la expresión del gen regulado
- b. Negativa (-).- El FT es represor de la expresión del gen regulado.
- c. Interacción con factor sigma (Transcripción).- El factor sigma permite que la RNA polimerasa se una al promotor del gen regulado.

La representación gráfica de la red se realizó utilizando el programa Cytoscape v.2.6.1¹⁷⁵ con la siguiente correspondencia:

1. Cada nodo representa un gen en la red.
2. Cada arista representa una interacción en la red.
3. La terminación de cada arista representa el tipo de interacción.
 - a. Una flecha corresponde a que el gen señalado está siendo regulado positivamente.
 - b. Una línea perpendicular a la arista corresponde a que el gen está siendo regulado negativamente.

- c. La terminación en punto señala que el gen está siendo regulado por un factor sigma.

La correspondencia respecto al tipo de interacción se utiliza únicamente para representaciones gráficas y para el análisis biológico en este trabajo. Sin embargo, para obtener las propiedades topológicas, así como para llevar a cabo el análisis modular, el grafo es no dirigido, es decir, todas las interacciones son equivalentes. En el caso de un par de nodos que presenten interacciones bidireccionales, en el grafo no dirigido equivaldrán sólo a una interacción.

4.2. Propiedades topológicas de la red de regulación transcripcional de *B. subtilis*

Se caracterizó a la red de regulación transcripcional de *B. subtilis* a través de los siguientes parámetros (definidos en la sección 1.2.2):

1. Número de nodos (genes) y número de aristas (interacciones).
2. Para cada nodo (*i*):
 - a. **Conectividad (k_i)**
 - b. **Interacciones entre nodos vecinos (n_i)**
 - c. **Coficiente de agrupamiento (C_i)**

Se obtuvo a través de la siguiente ecuación:

$$C_i = 2n_i / [k_i(k_i-1)]$$

3. Para toda la red:

- a. **Distribución de la conectividad [$P(k)$]**

Para obtenerla, se determinó la frecuencia con la que aparece un nodo con k aristas. De estos datos se obtuvo la probabilidad para cada k y se graficó cada valor de k contra su probabilidad acumulada, ambas en escala logarítmica.

- b. **Promedio del coeficiente de agrupamiento**

Se obtuvo el promedio del coeficiente de agrupamiento (C_i) de todos los nodos de la red que tienen un valor $k > 1$.

c. Distribución del coeficiente de agrupamiento [$C(k)$]

Se obtuvo a partir de la frecuencia con la que se presentan cada uno de los valores de C_i en la red. Se graficó cada valor de C_i contra su frecuencia, ambas en escala logarítmica.

d. Distribución de vecinos más cercanos [$T(n)$]

Para obtenerla, se determinó la frecuencia con la que aparece un nodo con valor n . De estos datos se obtuvo la probabilidad para cada n y se graficó cada valor de n contra su probabilidad acumulada, ambas en escala logarítmica.

e. Camino mínimo promedio

Para obtener este parámetro se aplicó el algoritmo Floyd-Warshall⁵⁸.

f. Motivos topológicos sobrerrepresentados en la red

Para determinar si la red presenta motivos sobrerrepresentados de tres elementos, se utilizó el programa *mfinder* diseñado en el laboratorio del Dr. Uri Alon en el Weizmann Institute of Science, Rehovot, Israel¹⁰⁴. El algoritmo que utiliza este programa consiste básicamente en buscar todos los patrones de conexión posibles formados por tres elementos, tanto en la red de interés como en cien redes creadas aleatoriamente con el mismo número de nodos y aristas que la original. Son considerados motivos sobrerrepresentados aquellos que aparecen de manera significativamente más frecuente en la red estudiada que en las aleatorias. Ver apéndice 4.

Para la obtención de todos los parámetros se diseñaron e implementaron algoritmos en el lenguaje de programación Perl (a excepción de la búsqueda de motivos topológicos, cuya metodología se explica en el inciso *f* del punto 3 de la presente sección). En el caso del camino mínimo promedio, se utilizó un código prediseñado ubicado en el Módulo Graph-0.91 en el directorio CPAN de Perl.

4.3. Análisis modular de la red de reguladores transcripcionales de *B. subtilis*

El análisis modular se llevó a cabo integrando sólo aquellos genes que codifican para FTs que están conectados a través de mínimo una arista con otro FT.

La única excepción fue *sigA* que codifica para el factor σA , ya que pertenece al grupo 1 de los factores tipo $\sigma 70$, que se caracterizan por una estructura que presenta un domino que da una mayor flexibilidad en el reconocimiento de promotores, a diferencia del resto de los factores sigma en *B. subtilis*, cuya estructura permite inducir la transcripción de regulones que se regulan en respuesta a cambios particulares⁷⁶. Esta diferencia funcional se refleja también en la topología de la red, ya que σA tiene una $k=782$ (se conecta con el 46.5% de los genes de la red), seguido sólo por σE con una $k=169$ (10% de los genes). Por lo tanto, la red que se utilizó para llevar a cabo el análisis modular excluye a *sigA*. En el Apéndice 6 se muestra un análisis modular incluyendo σA , cuyo resultado demuestra la dificultad de incluir en el análisis un factor sigma involucrado en la transcripción de un número tan elevado de genes.

Basándose en la red resultante de factores transcripcionales^b disponible en el Apéndice 2, se calculó el camino más corto (d_{ij}) entre cada par de nodos, a través del algoritmo Floyd-Warshall⁵⁸. A cada valor obtenido, se le calculó la función de asociación ($1/d_{ij}^2$) definida por Rives y Galitski en 2003¹⁶². Esta función permite amplificar aquellas relaciones cercanas y minimizar las remotas. Estos valores asignados a cada una de las relaciones entre todos los pares de nodos se ubican en un rango de 0 a 1. La asociación de un vértice consigo mismo equivale a 1¹⁶².

Se generó una matriz con dichos valores a la cual se le aplicó un algoritmo de agrupamiento jerárquico aglomerativo, utilizando el programa Gene Cluster 3.0⁴⁶. El objetivo de dicho método es agrupar una serie de elementos en un árbol o dendograma. En el árbol que se genera, la longitud de las ramas refleja el grado de similitud (cercanía en la red) entre los genes. El algoritmo funciona de la siguiente manera: se calcula la distancia

^b A pesar de que, formalmente, el análisis modular se aplicó para los genes que codifican para los factores transcripcionales, a lo largo del escrito el concepto de gen y factor transcripcional son equivalentes cuando éstos hacen referencia a los elementos que conforman un módulo.

entre los elementos (nodos) más cercanos en la red, bajo el criterio de una métrica definida, y se les une con una rama. A partir de ahí, ambos nodos ahora equivaldrán a un solo elemento y, de manera cíclica, los siguientes dos elementos más cercanos de la red, se unirán por una rama, cuyo tamaño dependerá de la distancia que los divide.

La métrica es la descripción matemática de similitud que define que tan *cercanos* son dos elementos. La elección de la misma depende notablemente de los datos que serán analizados y por lo tanto fue necesario evaluar los resultados de la aplicación de distintas métricas a la red estudiada. En este trabajo se evaluaron cuatro métricas: la correlación de Pearson no centrada, la tipo Spearman, la tipo Kendall τ y la distancia euclidiana. (Ver apéndice 5). Se eligió el coeficiente de correlación de Spearman por ser la métrica que dio como resultado grupos biológicamente más homogéneos y topológicamente cohesivos.

Para la medición de distancias, se eligió un agrupamiento de distancia promedio, es decir, para agrupar dos elementos en la red (x,y) el parámetro que se toma es la distancia promedio entre todas las que se obtienen a través de la combinación de todos los elementos agrupados en x contra los elementos agrupados en y .

El programa Gene Cluster generó finalmente 3 archivos: .cdt (tabla de los datos agrupados), .gtr (árbol de los elementos), .atr (árbol de los agrupamientos). Los resultados del análisis se visualizaron a través del programa TreeView Classic¹⁶⁵.

La definición de los nueve módulos discretos se obtuvo a partir del árbol generado por el método de agrupamiento. Los cortes se definieron a través de criterios biológicos.

4.4. Análisis de la Biología de los módulos encontrados

A partir de la definición de los módulos resultantes del análisis topológico, se llevó a cabo una descripción biológica detallada de los mismos, básicamente de tipo bibliográfica, integrándola con lo que se observa en la topología intra e intermodular. Asimismo, se sustituyeron los nombres designados inicialmente a FTs sin función definida por los nombres actualmente aceptados; esto como resultado de la falta de actualización de SubtiList desde 2001. Se repitió el análisis modular con los nuevos nombres con el único objetivo de presentar en el resultado, una figura con la designación actualizada de cada uno de los genes.

5. RESULTADOS y DISCUSIÓN

5.1. Reconstrucción de la red de regulación transcripcional de *B. subtilis*

La red de regulación que se reconstruyó para este trabajo se ubica en el Apéndice 1 y la representación gráfica de la misma se observa en la figura 13.

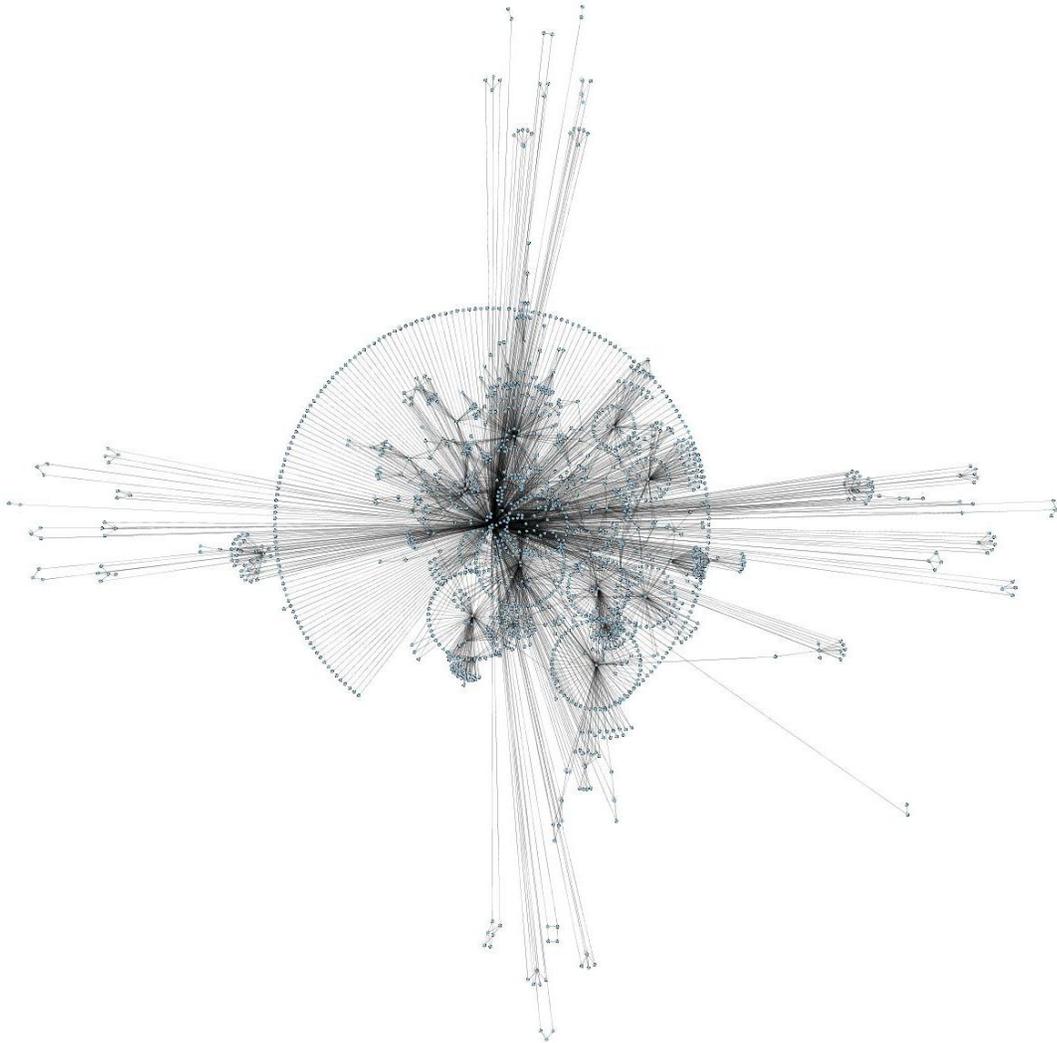


Figura 13. Representación gráfica de la red de regulación transcripcional de *B. subtilis*. Cada nodo representa un gen en la red y cada arista representa una interacción entre dos genes de la red, es decir, que uno de ellos regula la transcripción del otro.

5.2. Topología de la red de regulación transcripcional de *B. subtilis*

5.2.1. Propiedades topológicas

Uno de los primeros pasos necesarios para entender la estructura de la red de regulación de esta bacteria fue caracterizarla. En la tabla 1, se presentan los componentes por los que está formado el grafo reconstruido a partir de los datos de la red y se exponen las propiedades observadas del comportamiento de dicho sistema:

- Para cada nodo i de la red, se presenta su conectividad (k_i), las interacciones entre nodos vecinos (n_i) y su coeficiente de agrupamiento (C_i). La tabla que muestra estos datos se ubica en el Apéndice 3.
- Para toda la red se muestran los parámetros estadísticos agrupados en la tabla 2.

Número de nodos (genes)	Total: 1681 FTs: 154 GRs (incluyendo FTs) : 1632
Número de aristas (interacciones)	Red dirigida: 3096 (incluye 72 autorregulaciones y 8 interacciones bidireccionales)
	Red no dirigida: 3016

Tabla 1. Número de nodos y aristas que presenta la red de regulación transcripcional utilizada para el presente trabajo.

Distribución de la conectividad [$P(k)$]	$\sim k^{-2.11}$ (Fig. 14)
Promedio del coeficiente de agrupamiento	0.538
Distribución del coeficiente de agrupamiento [$C(k)$]	$\sim k^{-1.10}$ (Fig. 15)
Distribución de vecinos más cercanos [$T(n)$]	$\sim n^{-2.03}$ (Fig. 16)
Camino mínimo promedio	3.158
Motivos topológicos de tres elementos sobrerrepresentados	<i>feedforward</i> (FF) (Fig. 9) ^c

Tabla 2. Propiedades topológicas de la red de regulación de *B. subtilis*.

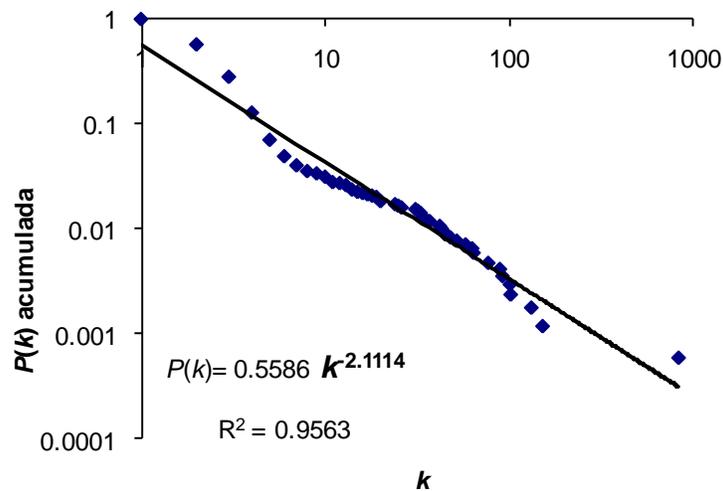


Figura 14. Distribución de la conectividad [$P(k)$].^d

^c Ver Apéndice 4.

^d La gráfica muestra la distribución de $P(k)$ acumulada, lo que dio como resultado la ecuación $P_c(k) = 0.5586k^{-1.114}$. Sin embargo, transformando el valor de $P_c(k)$ a $P(k)$, se obtiene la ecuación presentada en el gráfico.

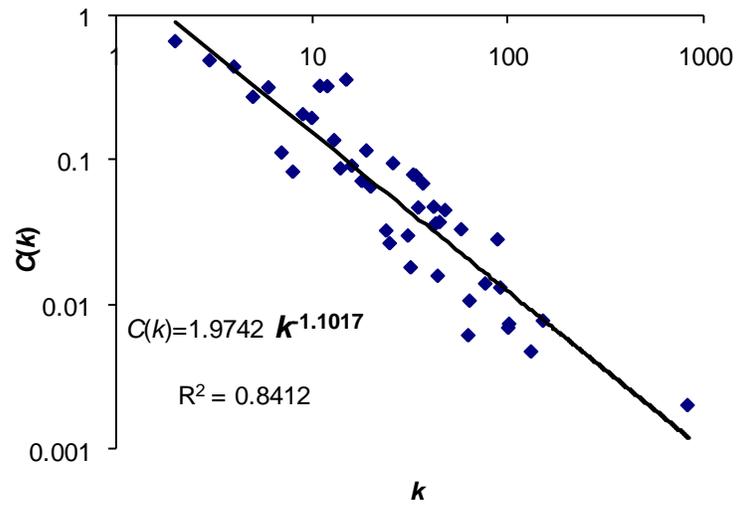


Figura 15. Distribución del coeficiente de agrupamiento [$C(k)$].

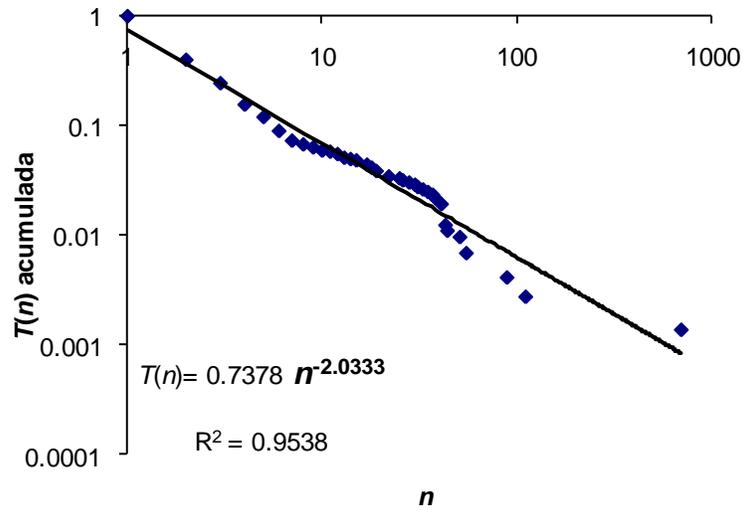


Figura 16. Distribución de vecinos más cercanos [$T(n)$].^e

^e La gráfica muestra la distribución de $T(n)$ acumulada, lo que dio como resultado la ecuación $T_c(n) = 0.7378 n^{-1.0333}$. Sin embargo, transformando el valor de $T_c(n)$ a $T(n)$, se obtiene la ecuación presentada en el gráfico.

5.2.2. La red de regulación transcripcional de *B. subtilis* presenta una estructura jerárquico-modular

Una de las primeras propiedades que se definió en teoría de redes fue la del *mundo pequeño*. La red caracterizada en este trabajo cumple con esta propiedad al presentar un camino mínimo promedio de 3.158, consistente con el hecho de ser una red biológica. A pesar de que esta característica la presentan todas las redes complejas, no son exclusivas de éstas; las redes aleatorias también son de *mundo pequeño*.

Sin embargo, la red transcripcional de *B. subtilis* cumple además con otras dos propiedades de las redes complejas: es libre de escala y presenta una estructura modular. Lo primero se define a partir de su distribución de conectividad [$P(k)$], la cual sigue una ley de potencia inversa con exponente de valor 2.11; esto también nos advierte de la presencia de *hubs* en la red. La modularidad se define a partir del promedio de coeficiente de agrupamiento (C), el cual es de 0.538, significativamente superior al que presentan las redes aleatorias, que suelen tener un promedio con un valor de un orden de magnitud menor.

A pesar de que el promedio de C nos da una medida de la modularidad potencial de la red, se necesitó la evaluación de un tercer parámetro para definir que la red es también jerárquica. Este parámetro es $C(k)$ y su distribución también sigue una ley de potencia inversa con un exponente de valor equivalente a 1.1, lo que nos indica que C es dependiente de k . Esto quiere decir que los nodos ubicados en los pequeños y cohesivos módulos de la red tienen valores de C muy elevados, mientras que los nodos que unen a grupos más grandes pero menos cohesivos, tienen un valor de C menor.

Todas estas características denotan que la red presenta una estructura modular y jerárquica, lo que hace posible un análisis modular.

Adicionalmente, se hizo la búsqueda de motivos topológicos de tres elementos, siendo el motivo *feedforward* (FF) el único que aparece de manera estadísticamente representativa (ver Apéndice 4). Esto confirma lo que otros trabajos previos sobre redes biológicas han demostrado, en los cuales siempre aparece el FF sobrerrepresentado. Como ya se ha mencionado previamente, el papel de este motivo topológico está principalmente ligado a un control temporal de la expresión de un gen, dependiendo de la acumulación adecuada de los dos genes reguladores que completan el triángulo (Fig. 9).

5.3. Análisis modular de la red de reguladores transcripcionales de *B. subtilis*

Después de confirmar que la red de regulación transcripcional de *B. subtilis* presenta una estructura jerárquico-modular, fue posible hacer un análisis que agrupara a los FTs en módulos de acuerdo a la topología de la red. La red de FTs utilizada para dicho análisis (excluyendo a σA) está compuesta por 71 nodos y 81 aristas (ver Apéndice 2). A partir del análisis detallado de la Biología de cada uno de los módulos (presentada en la sección 5.4), se le asignó una función a cada módulo encontrado. La figura 17 presenta el análisis llevado a cabo a partir del algoritmo de agrupamiento jerárquico aglomerativo.

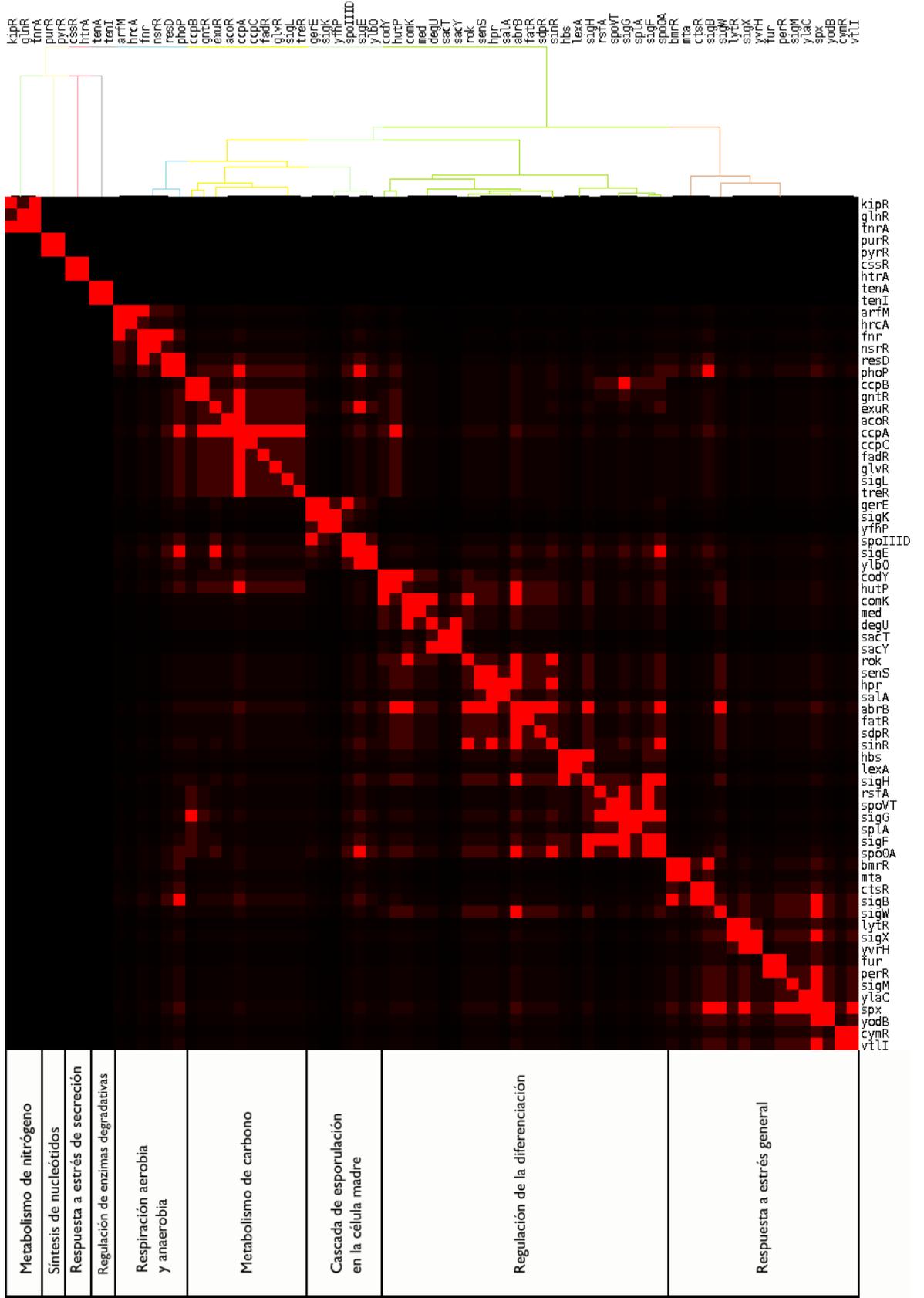


Figura 17 (página anterior). Análisis de agrupamiento de la red de FTs de *B. subtilis*. La intensidad del color rojo de cada punto refleja la cercanía de ambos nodos en la red, siendo 1 el valor máximo y 0 el valor mínimo, en donde 0 está representado por el color negro. El dendograma ubicado en la parte superior de la imagen refleja las agrupaciones de genes obtenidas a partir del análisis; la longitud de las ramas representa la cercanía entre los genes y agrupamientos. Para facilitar la observación del lugar en donde se definió cada corte, las ramas de cada módulo están marcadas con un color diferente. En la parte inferior se muestra la función fisiológica con la cual se asocia cada módulo.

5.4. Descripción detallada de la Biología de los módulos encontrados

5.4.1. Módulo de metabolismo de nitrógeno. (*glnR*, *kipR* y *tnrA*)

El nitrógeno es un elemento esencial para los seres vivos, pero su presencia en el medio puede variar con frecuencia. Para contender con las variaciones de este elemento en su hábitat, *B. subtilis* ha desarrollado mecanismos que le permiten optimizar el aprovechamiento de diferentes fuentes de nitrógeno mediante la modulación de la expresión de muchas enzimas involucradas en la asimilación del nitrógeno en respuesta a la disponibilidad del mismo. Los reguladores TnrA, KipR y GlnR fueron agrupados por nuestro método en un módulo que responde a las fluctuaciones en la cantidad de nitrógeno en el medio.

En *B. subtilis* la asimilación de nitrógeno está controlada principalmente por dos proteínas homólogas: TnrA y GlnR. TnrA es considerado un regulador global del metabolismo de nitrógeno en *B. subtilis* y se activa únicamente frente a condiciones en las cuales este elemento se encuentra limitado²⁰⁰. Esta proteína responde a la presencia de la glutamina, que es la fuente preferencial de nitrógeno de esta bacteria Gram positiva. Cuando se encuentra bajo condiciones en las cuales sólo están presentes otras fuentes de nitrógeno que se metabolizan lentamente, TnrA activa genes y operones que facilitan la utilización y el transporte de amonio, ácido úrico, nitrato, nitrito y oligopéptidos^{180,200,203}; y es represor de otros operones, entre ellos el de *gltAB*, que codifica para la glutamato

sintetasa, una de las enzimas más relevantes para el metabolismo del nitrógeno necesaria para la asimilación de amonio y la biosíntesis de glutamato¹⁶.

GlnR, por su parte, es un represor que participa principalmente cuando la bacteria se encuentra bajo condiciones de exceso de nitrógeno. Regula primordialmente la síntesis de glutamina sintetasa (GlnA) y reprime la expresión de algunos de los genes y operones activados por TnrA, principalmente aquellos que regulan la asimilación de amonio, nitrato y nitrito^{23,55}.

A pesar de que la esporulación se inicia bajo condiciones de crecimiento en donde el nitrógeno es limitado, TnrA y GlnR no se requieren para llevar a cabo dicho proceso⁵⁵. Sin embargo, TnrA inhibe de forma indirecta la esporulación a través de la activación del tercer FT agrupado en este módulo: KipR, que regula al operón *kip*. Dentro de este operón, se localiza también el gen que codifica para KipI, que es un inhibidor de la autofosforilación por ATP de KinA, sensor histidina cinasa involucrada en la iniciación de la esporulación¹⁹⁴. De esta manera, KipI se expresa cuando las células se encuentran en un medio que carece de su fuente preferida de nitrógeno: glutamina, pero no de carbono: glucosa; por lo cual la bacteria aún no se compromete a la esporulación. Se sugiere por lo tanto que, tanto TnrA como GlnR, favorecen la adaptación de la célula durante el crecimiento vegetativo y no durante el proceso de diferenciación⁵⁵.

5.4.2. Módulo de síntesis de nucleótidos (*purR* y *pyrR*)

Las purinas y pirimidinas son esenciales para la célula ya que favorecen la transferencia de energía durante reacciones metabólicas y son las moléculas a partir de las cuales se forman los ácidos nucleicos. En este módulo se agrupan las dos proteínas involucradas en la regulación del metabolismo de bases nitrogenadas: los factores transcripcionales PurR y PyrR.

PurR es represor de diversos genes involucrados en la síntesis, metabolismo y transporte de purinas¹⁷. Se le ha involucrado en la inhibición de la síntesis de cofactores y salvataje de purinas¹⁶⁶. PurR regula además al operon *pyrRPBCAAABKDFE*, en donde se

ubica el gen *pyrR* que codifica para el regulador PyrR. Esta correulación probablemente contribuye a balancear el acervo de purinas y pirimidinas¹⁷⁷.

La proteína PyrR, es una proteína bifuncional, encargada de regular la síntesis y metabolismo de las pirimidinas a través del mecanismo de atenuación transcripcional^{122,191}, además de presentar la función enzimática de una uracil-fosforibosiltransferasa (UMP), que cataliza la reacción para formar uracil monofosfato¹⁹⁰.

PurR y PyrR, así como los genes regulados por ambas, se encuentran separados de los otros FTs analizados, formando un módulo aislado unido sólo al resto de la red transcripcional a través del factor σ A.

5.4.3. Módulo de respuesta a estrés de secreción (*cssR* y *htrA*)

B. subtilis se caracteriza por una notable capacidad para secretar proteínas en el medio. Sin embargo, la secreción de éstas a altos niveles puede traer como consecuencia la acumulación de proteínas mal plegadas, lo que resulta en una situación adversa para la célula⁹⁹. Para combatir dicho estrés, esta bacteria presenta cambios en su expresión genética que están controlados por los factores de transcripción CssR y HtrA, componentes de este módulo. CssR forma parte del sistema de dos componentes CssRS (por sus siglas en inglés, Control secretion stress Regulator and Sensor) requerido en *B. subtilis* para responder al estrés causado por la secreción y acumulación de proteínas mal plegadas ubicadas entre la membrana y la pared celular.

La proteína CssS funciona como sensor y CssR actúa como el regulador de la transcripción de *htrA*⁹⁸, *htrB*³⁹ y del propio sistema *cssRC*⁹⁸. El gen *htrA* codifica para una serina proteasa que está adherida a la membrana y cuya función principal es degradar proteínas mal plegadas y agregadas; mientras que *htrB* codifica para una proteína de la misma clase que HtrA¹⁴⁵. Por su parte, el propio HtrA, que es el segundo FT que conforma este modulo, no ha sido descrito como regulador de genes diferentes a sí mismo.

Por sus características, este módulo y los genes regulados por él forman un regulón complejo unido al resto de la red a través de σ A. Se ha descrito que los genes pertenecientes a este módulo, son inducidos también por choque térmico en mutantes *htrA*, y por ello se les ha clasificado también en una de las clases que agrupan genes inducidos por calor³⁹.

5.4.4. Módulo de regulación de enzimas degradativas (*tenA* y *tenI*)

Durante la transición entre el crecimiento exponencial y la fase estacionaria, *B. subtilis*, al igual que otras bacterias, se caracteriza por secretar enzimas degradativas, tales como las exoproteasas. Éstas tienen la función de actuar sobre productos secretados por la propia célula, así como de degradar productos liberados por células muertas aledañas, con el propósito de romperlos en moléculas más pequeñas (péptidos o mono o disacáridos) para aportar nutrientes a la colonia¹³⁹.

A pesar de que para la producción de este tipo de enzimas se conoce una notable cantidad de proteínas reguladoras, en este módulo se agrupan dos de ellas que pertenecen a un mismo operón *tenAI-goxB-thiSGF-yjbV* y que están implicadas en la regulación de la producción de proteasas extracelulares. TenA estimula la producción a nivel transcripcional de alcalina proteasa, subtilisina, levansucrasa y otras enzimas extracelulares. Mientras que, por el contrario, TenI, al ser un represor de *tenA*, tiene la función de reducir la producción de las enzimas antes mencionadas¹⁴⁹. Se ha sugerido además, que ambas proteínas están involucradas en la biosíntesis o metabolismo de tiamina¹⁸⁷.

TenA y TenI están totalmente desconectados de los demás; es incluso el único de los módulos reportados en este análisis que no está regulado por σ A. Sin embargo, los genes que controlan son también regulados por otras proteínas con funciones similares, tales como DegU y SenS, ubicadas en el módulo de diferenciación, descrito más adelante. Se ha demostrado que TenA y TenI no son esenciales para el crecimiento celular ni tampoco lo son para la producción de enzimas degradativas, sin embargo, su ausencia causa un retraso en la esporulación¹⁴⁹.

5.4.5. Módulo de respiración aerobia y anaerobia. (*arfM*, *hrcA*, *fnr*, *nsrR*, *resD* y *phoP*)

B. subtilis se consideró como un microorganismo aerobio estricto durante mucho tiempo; sin embargo, en las últimas dos décadas se ha determinado que esta bacteria es también capaz de crecer en un medio anaerobio¹⁴². En este módulo se agrupan factores de transcripción que participan en los mecanismos reguladores que permiten modular las capacidades metabólicas en función de la disponibilidad de oxígeno en el medio.

El sistema de transducción de dos componentes ResDE está formado por ResE, que es la cinasa sensor, y ResD, que funciona como el regulador de respuesta. Este sistema regulador es central para esta adaptación, ya que modula la transcripción de numerables genes involucrados tanto en la respiración aerobia como en la anaerobia¹⁴². En un medio en el cual el oxígeno es limitado, ResD inicia una cascada de regulación al inducir la transcripción de *fnr*, cuya proteína codificada, FNR, activa a su vez la transcripción de *arfM*, y éste induce la expresión de HrcA^{37,91,126}. FNR tiene un papel relevante al regular la expresión de los genes que participan en la síntesis de la nitrato reductasa, a partir de la cual *B. subtilis* puede utilizar el nitrato como aceptor de electrones en ausencia de oxígeno¹⁴². FNR induce la transcripción de *arfM*, cuyo producto ArfM, modula la expresión de proteínas involucradas en la respiración anaerobia y en la fermentación¹²⁶. Esta última regula a su vez al operón *dnaK*, en el cual se ubica el gen *hrcA*, cuya proteína codificada, HrcA, es un represor de los genes que codifican para proteínas inducidas en estrés por calor⁹¹.

Por su parte, NsrR es otra de las proteínas presentes en el módulo que tiene una función de adaptación al medio anaerobio. Se ha sugerido que es un sensor de óxido nítrico (NO) y que controla a genes regulados por ResD que responden a este gas¹⁴¹, entre los que se encuentran el operón *narK-fnr*, que incluye al regulador *fnr*; *nasBCDEF* que participa en la asimilación de nitritos; y para el gen *hmp*, que codifica para una flavohemoglobina, la cual está asociada a la respuesta de adaptación frente a estrés nitrosativo¹⁶³.

En este módulo, también fue agrupada la proteína reguladora PhoP, que controla un regulón que responde cuando el fosfato inorgánico se encuentra en bajas concentraciones, y

que constituye la principal respuesta de la célula frente a esta carencia. Los genes regulados por PhoP participan en la tarea de obtener fosfato a partir de sustratos más complejos¹⁸³. PhoP regula y es regulada positivamente por ResD, y ambos corregulan algunos de los genes involucrados en la obtención de fosfato bajo las condiciones mencionadas, entre los cuales se encuentra *resA*, que codifica para una proteína esencial similar a la proteína de biogénesis del citocromo *c*^{18,97}. La topología encontrada, denota que tanto la respiración, como la respuesta a falta de fosfato, son funciones fisiológicas íntimamente relacionadas¹⁸.

5.4.6. Módulo de metabolismo de carbono. (*ccpB*, *gntR*, *exuR*, *ccpA*, *acoR*, *ccpC*, *sigL*, *fadR*, *treR* y *glvR*)

B. subtilis, al ser un organismo heterótrofo, debe obtener carbono del medio, siendo la glucosa la fuente preferida de esta bacteria. Sin embargo, en ausencia de ésta, es capaz de utilizar otros carbohidratos, para lo cual presenta un complejo sistema de regulación. En este módulo, se agrupan los FTs implicados en esta actividad.

La manera en que *B. subtilis* regula transcripcionalmente el metabolismo de carbono, es principalmente a través del fenómeno conocido como represión catabólica (CCR, por sus siglas en inglés). Este evento se caracteriza porque la presencia de una fuente de carbono rápidamente metabolizable en el medio, inhibe la síntesis de enzimas involucradas en la utilización de fuentes de carbono secundarias⁷³. En esta bacteria Gram positiva, el principal FT encargado de controlar este sistema es CcpA, cuya actividad es desencadenada a partir de la fosforilación de la cinasa HPr, que es un sensor intracelular de azúcares¹⁸². Así, la represión catabólica en *B. subtilis* depende de la actividad reguladora de CcpA y de la presencia de diversas fuentes de carbono. El complejo, formado por esta proteína reguladora y su cofactor HPr, se une a secuencias palindrómicas específicas ubicadas en regiones de promotores de operones catabólicos y que son mejor conocidas como sitios *cre*¹³⁵.

Entre los sistemas de utilización de carbohidratos con los que cuenta *B. subtilis*, aparte del de la glucosa, se encuentran los que se encargan de la utilización de gluconato,

hexuronato, acetoína, trehalosa y maltosa. Todos estos sistemas están codificados en operones cuya expresión depende de CcpA. Además, cada una de estas unidades transcripcionales es regulada por otro FT, en ocasiones ubicado en el propio operón. Al ser también controladas por CcpA, estas proteínas reguladoras se agrupan en este módulo y a continuación se hará una descripción de cada una de ellos.

GntR se ubica en el operón *gnt*, responsable del catabolismo de gluconato⁶⁵ y sujeto a represión catabólica mediada por intermedios glicolíticos. GntR funciona como represor de este operón y su función es antagonizada por la presencia de gluconato en el medio⁶⁷. Por su parte, ExuR pertenece al operón *exu*, que participa en la utilización de hexuronato. Para *B. subtilis*, la presencia del sistema codificado en este operón es esencial cuando crece en un medio en el cual la única fuente de carbono es galacturonato. Este compuesto es particularmente importante para esta bacteria, ya que el polimetilgalacturonato, conocido también como pectina, se encuentra en la pared celular de las plantas, y por tanto, se encuentra en el suelo, que es el medio de crecimiento más común para este microorganismo¹³¹.

En fase de crecimiento exponencial, *B. subtilis* excreta diversos compuestos orgánicos, entre los cuales se encuentra la acetoína, que además de permitirle a la bacteria degradar grandes cantidades de glucosa sin una notable acidificación del medio, puede también ser reutilizada cuando otras fuentes de carbono se han terminado⁶. El operón *aco*, regulado positivamente por el factor transcripcional AcoR⁶, codifica para el complejo enzimático de la acetoín deshidrogenasa, responsable del catabolismo de acetoína en esta bacteria Gram positiva⁹⁴.

B. subtilis puede también usar trehalosa como única fuente de carbono y es el operón *tre* el que desempeña esta función, controlado por el represor TreR, contenido en la misma unidad transcripcional¹⁶⁹. Por otro lado, para la utilización de maltosa, el operón *glv* es uno de los sistemas involucrados. GlvR es su regulador positivo, y al igual que todos los operones ya mencionados, es regulado negativamente por CcpA²⁰².

CcpA está también involucrado en el control de la expresión de FadR, que es el regulador transcripcional de la degradación de ácidos grasos. Estos compuestos orgánicos

forman parte de la membrana y son fuentes de energía muy importantes para los organismos⁶⁶. *B. subtilis* puede crecer en medios en donde los ácidos grasos son la única fuente de carbono y es precisamente el regulón de FadR el que participa en la utilización de este compuesto¹²⁹.

Además de controlar proteínas reguladoras, CcpA también es represor del factor σ L. Éste se encuentra involucrado en la regulación del metabolismo de fructosa, levanas⁴¹, arginina⁶⁹ y acetoína⁶, así como en la degradación de leucina y valina^{40,41}. Su participación en el metabolismo de isoleucina, lo implica además en la respuesta adaptativa de la membrana a choque térmico por bajas temperaturas¹⁹⁹. Así, puede observarse que σ L regula principalmente operones relacionados con la degradación de compuestos nitrogenados, notándose una conexión entre el metabolismo de nitrógeno y carbono, que estará presente a lo largo de todo el análisis de esta red de transcripción. Este mecanismo en particular (la regulación de σ L por CcpA) permite el control coordinado de las vías de carbono y nitrógeno en respuesta a la disponibilidad de fuentes de carbono³⁰.

Se ha hecho una descripción general de los FTs regulados por CcpA. Sin embargo, debe tomarse en cuenta que existen otras dos proteínas involucradas en el mecanismo de represión catabólica: CcpB y CcpC, también agrupadas en este módulo. La primera presenta un 30% de similitud con CcpA y, al igual que ésta, ejerce su función de represión dependiendo de la presencia de HPr fosforilada. Paralelamente a CcpA, regula los operones *gnt* y *xyl*, involucrados en el metabolismo de gluconato y xilosa, respectivamente²⁹. CcpC, por su parte, es necesaria para la represión catabólica de genes que codifican para enzimas que participan en el ciclo de Krebs en *B. subtilis*^{102,108}, involucradas principalmente en el metabolismo de carbono, pero implicadas también en el proceso de esporulación¹⁹. CcpC es a su vez regulado por CcpA¹⁰⁸.

5.4.7. Módulo de la cascada de esporulación en la célula madre. (*gerE*, *sigK*, *yfhP*, *spoIIID*, *sigE* y *gerR*)

Cuando los nutrientes se agotan en el medio, *B. subtilis* inicia un proceso denominado esporulación, cuyo objetivo final es formar una estructura en estado latente y muy resistente: la endospora. La esporulación es un proceso de diferenciación muy complejo, en el cual los cambios morfológicos están asociados a diferencias en la expresión de genes en tiempo y espacio. Uno de los primeros cambios que sufre la célula es la formación de un septo que la divide en dos compartimentos asimétricos: el de la célula madre, que es el más largo, y el de la preespora, que es de menor tamaño. Ambos compartimentos presentan diferentes programas de expresión que derivan de cascadas de regulación iniciadas por distintos sigmas. A pesar de que ambos programas están claramente separados en la red transcripcional, son totalmente dependientes uno de otro, ya que se encuentran conectados a través de vías de señalización¹¹⁰. En este módulo, se agrupan los FTs que participan en la cascada de regulación que se lleva a cabo en el compartimento de la célula madre durante este proceso. Esta cascada es de tipo jerárquica, en donde los genes aparecen en el siguiente orden: *sigE* → *spoIIID* y *gerR* → *sigK* → *gerE* y *yfhP*^{44,111}.

Los factores sigma son los responsables de las diferentes cascadas reguladoras que se llevan a cabo en cada compartimento. En la célula madre, el primer factor sigma en aparecer después de la división asimétrica es σE , que es activado por la forma fosforilada de Spo0A, regulador maestro que controla la entrada a la esporulación en *B. subtilis*⁶⁴. Este factor sigma es el responsable de iniciar la transcripción de un regulón que incluye genes con múltiples funciones, entre las cuales se encuentran: promover el engullimiento de la preespora por la célula madre, prevenir una segunda división asimétrica en la célula madre, sintetizar las proteínas involucradas en el ensamblaje de la capa de la espora y mantener un nivel adecuado de actividad metabólica que permita la progresión del proceso de esporulación bajo condiciones de escasez de nutrientes^{45,87,89}. Además, regula genes que codifican para otras proteínas reguladoras presentes en el módulo: SpoIIID¹¹³ y GerR⁵⁴.

Las proteínas SpoIIID y GerR actúan regulando una parte del regulón de σE . SpoIIID regula negativamente aquellos genes responsables de los cambios morfológicos de la célula en las primeras etapas del proceso. Por otro lado, induce la transcripción de otros genes, entre los cuales se ubican aquellos que participan en la producción de la proproteína pro- σK , que como resultado de la red de señalización que comunica a ambos compartimentos, se transformará en la forma madura de dicha proproteína, dando lugar a la aparición del factor σK ⁴⁴. GerR, por su parte, es represor de un número menor de genes en comparación con SpoIIID, compartiendo con esta proteína la regulación del operón *spoIIIA*, que codifica para un complejo necesario para proseguir el proceso de esporulación después del engullimiento^{44,78}. Regula además a los genes *spoIIM* y *spoIIP*, que participan en la eliminación del peptidoglicano presente en el septo que divide a la célula, con el objetivo de llevar a cabo la disolución del mismo, prerequisite para que se desarrolle correctamente el engullimiento^{28,44}.

La activación de σK , cuya precursora es la proproteína pro- σK , requiere de una señal generada en la preespora bajo el control de σG ⁸⁷. Al ser activado, σK controla un regulón, cuyos genes están involucrados en la formación de la capa de la espora y en el proceso de maduración de la misma⁸⁹. Regula también al gen *gerE* que codifica para la proteína GerE³⁸, que al igual que las proteínas reguladoras previamente descritas, funciona como apoyo para la regulación de los genes inducidos por un factor sigma, optimizando su expresión. En este caso, GerE controla una parte del regulón de σK , que está compuesta por genes activados al final de la cascada de expresión llevada a cabo en el compartimento de la célula madre^{44,205}, muchos de ellos son los denominados genes *cot*, que codifican para proteínas de la capa de la espora^{111,205}. Así, la aparición de σK genera un cambio en la expresión de genes respecto a la que se tenía previamente con el regulón controlado por el factor σE ¹¹¹.

El último FT por describir en este módulo es YfhP, cuya función es poco conocida. Hasta la fecha se sabe solamente que es regulado por σK y que es probablemente un represor de la transcripción del operón *yfhQ-fabL-sspE*^{63,201}. La función del primer gen del

operón es desconocida, mientras que *fabL* está involucrado en el metabolismo de lípidos⁸⁴ y *sspE* codifica para una pequeña proteína de espora que es soluble en ácido²⁰¹.

De esta manera, los reguladores clave en esta cascada son los dos factores: σE y σK ; las otras proteínas presentes tienen principalmente la función de activar o reprimir los genes de los regulones de ambos factores sigma. Este prendido y apagado de genes contribuye a llevar a cabo apropiadamente el complejo proceso de morfogénesis que sufre la célula durante la esporulación⁴⁴.

5.4.8. Módulo de regulación de la Diferenciación (Cambios Morfogenéticos). (*codY, hutP, comK, med, degU, sacT, sacY, rok, abrB, senS, hpr, salA, fatR, sdpR, hbs, lexA, sigH, sinR, rsfA, spoVT, sigG, spo0A, spla* y *sigF*)

La diferenciación celular es un proceso biológico fundamental, que resulta de la coordinación de la expresión de diversos genes que involucran cambios morfológicos, así como del establecimiento de distintos programas de expresión genética en los diferentes tipos celulares⁸⁹. *B. subtilis* es un interesante modelo para entender el proceso de diferenciación en procariontes, no sólo por su notable capacidad para producir esporas, sino también por su capacidad para diferenciarse en otros tipos celulares, tales como células móviles, células genéticamente competentes, células productoras de matriz extracelular, células productoras de enzimas degradativas y células secretoras de toxinas, también conocidas como caníbales¹²⁰, todas ellas brevemente descritas en la introducción; lo anterior como respuesta a condiciones fisiológicas adversas. Las diferencias morfológicas que caracterizan a cada tipo celular se originan principalmente a partir de la variación en la expresión de genes en cada una de ellas¹²⁰. En este módulo, se agruparon los principales FTs involucrados en la coordinación de la expresión de dichos genes.

La agrupación en el mismo módulo de los FTs involucrados en distintos destinos celulares señala la notable interacción que hay entre las proteínas que coordinan la expresión de cada tipo celular. Esto es el resultado de la compleja red que gobierna la

diferenciación en *B. subtilis*; y la topología de dicha red refleja, en concordancia con lo observado experimentalmente, que muchos de estos destinos celulares son mutuamente excluyentes¹²⁰. Se observa en el módulo un ciclo con esta característica: Spo0A reprime a AbrB, AbrB reprime al factor σ H, y este último induce la transcripción de Spo0A. Esto significa que las células que expresan Spo0A no expresan AbrB y viceversa¹²⁰. A continuación se describirá el papel de Spo0A y σ H, así como de los otros FTs involucrados en el proceso de esporulación, agrupados también en este módulo. Más adelante se describirá el rol de AbrB en la diferenciación.

El mecanismo de diferenciación mejor estudiado en *B. subtilis* es el de la esporulación. El FT maestro que induce la iniciación de este proceso es Spo0A. Cuando se presenta densidad poblacional y carencia de nutrientes, las encargadas de reconocer estas condiciones son las histidinas cinasas KinA, KinB, KinC, KinD y KinE, siendo las dos primeras las involucradas principalmente en la iniciación de la esporulación. A través de los intermediarios Spo0F y Spo0B, estas cinasas fosforilan a Spo0A²⁴. Una vez fosforilada, esta proteína activa la transcripción de genes involucrados en la formación del filamento axial y en la división asimétrica, así como la de los genes que codifican para los primeros sigmas que aparecen en el proceso: σ F en la preespora y σ E en el compartimento de la célula madre⁸⁹, cuya cascada de regulación fue descrita en el módulo anterior.

El factor σ H es otro de los elementos que juega un papel importante en la iniciación de la esporulación. Regula a varios de los genes que codifican para proteínas involucradas en la fosforilación de Spo0A, entre ellos a *spo0F* y a *kinA*²². Induce también la transcripción de los genes *phr*, inhibidores de proteínas que modulan la entrada a competencia; del operón *ftsAZ*, requerido para la septación asimétrica de la célula; e incluso *spo0A* y el operón *spoIIA*, que codifica para el factor σ F, se transcriben a través de un promotor dependiente de σ H⁸⁹. Además de la regulación de genes que participan en la esporulación y competencia, controla la transcripción de genes que codifican para proteínas involucradas en una gran variedad de funciones, tales como transporte, metabolismo de la pared celular, proteólisis y biogénesis del citocromo²².

La cascada de regulación en la preespora inicia después de la división asimétrica con la expresión del factor σF , regulado positivamente por Spo0A y σH , y cuyo regulón abarca genes involucrados principalmente en los primeros eventos de la esporulación, entre los cuales están *spoIIQ*, requerido para el proceso de engullimiento, y *rsfA*, que codifica para el factor transcripcional RsfA, también agrupado en este módulo^{89,110}. Esta proteína reguladora controla una pequeña parte del regulón de σF , y juntos, regulan a *spoIIR*, que forma parte de la red de señalización que comunica a los dos compartimentos generados en la división; *katX*, que codifica para una catalasa de la espora que es requerida para las resistencia a peróxido de hidrógeno; *gpr*, cuyo producto es una proteasa específica para proteínas protectoras del DNA en la espora; y *spoIIIG*, que codifica para el otro factor sigma involucrado en esta cascada: σG ⁸⁹.

La activación del factor σG , por su parte, dirige la expresión de los genes involucrados en la fase tardía dentro de la preespora y está acoplada a la última etapa del engullimiento. Este factor regula la transcripción de genes, cuyas funciones pueden agruparse en tres principales: esporulación, germinación y protección de la espora frente a eventos que puedan dañar al DNA⁸⁹. Bajo su control están además dos proteínas que a su vez regulan la transcripción de algunos genes del regulón de este factor sigma; éstas son SpoVT¹² y SplA¹⁵⁰. La primera regula un notable número de genes del regulón de σG , entre ellos: *spoIVB*, peptidasa involucrada en la red de señalización que comunica a los dos compartimentos; el operón *gerA*, involucrado en la germinación en respuesta a alanina, al igual que su parálogo, el operón *gerB*; y algunos genes *ssp*, codificadores de proteínas de espora, pequeñas y solubles en ácido (SASPs, por sus siglas en inglés)⁸⁹. SplA, en cambio, corregula con σG únicamente al operón en el cual se ubica el gen que la codifica: *splAB*, cuya expresión resulta en un fotoproducto que le da resistencia a la espora frente a la radiación ultravioleta¹⁴⁴.

El factor σH , además de inducir la transcripción de elementos involucrados en la esporulación, regula al gen *hbs*, que codifica para la proteína HBSu¹³². Ésta es una proteína del tipo similar a las histonas, involucrada en la catálisis y regulación de una diversidad de funciones relacionadas con el cuidado y protección de la estructura del DNA⁵³. Es esencial

para el crecimiento y el desarrollo en *B. subtilis*, lo cual se atribuye a su interacción con la replicación y la transcripción como consecuencia de su facilidad para compactar al DNA y para condensar al cromosoma en estructuras similares a las observadas en el nucleosoma¹³²; mutantes de este gen han resultado en fenotipos con eficiencia reducida durante el proceso de esporulación¹⁰⁹. Se le ha identificado además como componente de una partícula similar a la partícula señal de reconocimiento (SRP, por sus siglas en inglés) en eucariontes, la cual tiene implicaciones en la correcta distinción entre proteínas secretadas y proteínas citoplásmicas, que a su vez es un paso importante para iniciar una vía de secreción¹⁴⁰. HBSu es un represor de su propia expresión⁵³ y hasta la fecha no se conoce que regule la transcripción de otro gen diferente a sí mismo. HBSu es también regulado por LexA¹⁰, que pertenece al sistema SOS en *B. subtilis*. En procariontes, los sistemas SOS son el conjunto de respuestas fisiológicas que responden a agentes dañinos para el DNA. En el caso de esta bacteria Gram positiva, LexA es una de las proteínas que controla dicho sistema, siendo principalmente represor de genes involucrados en la reparación de DNA, recombinación y división celular^{10,72,107}.

Después de describir la parte del módulo que corresponde a los genes involucrados en la esporulación y a otros conectados a los mismos, se retomará el papel del FT pleiotrópico AbrB. Esta proteína está conectada con factores involucrados en la expresión de genes en diversos destinos celulares y controla principalmente genes involucrados en la transición de la fase de crecimiento a la fase estacionaria. Además de la mencionada represión que ejerce sobre el proceso de esporulación, controla la expresión de genes en una gran variedad de procesos metabólicos y fisiológicos, incluyendo producción de enzimas degradativas extracelulares, utilización de nitrógeno, metabolismo de aminoácidos, motilidad, desarrollo de competencia, sistemas de transporte, respuesta a estrés oxidativo y metabolismo de fosfato¹⁵⁶. Los FTs que regula AbrB están incluidos en este módulo y serán expuestos a lo largo de lo que resta de esta descripción.

Dos genes que codifican para proteínas involucradas en el control de enzimas extracelulares, *senS* y *hpr*, son regulados por AbrB; el primero de forma negativa y el segundo de forma positiva^{130,181}. La proteína SenS estimula la expresión de este tipo de enzimas como resultado de que ejerce represión sobre *hpr*^{105,195}. HPr, a su vez, es un

represor de genes que codifican para diversas enzimas degradativas expresadas principalmente durante el estado estacionario, entre ellos *aprE* (subtilisina E), que es una serina proteasa alcalina inducida de manera postexponencial en respuesta a estrés de fase estacionaria; *nprE*, metaloproteasa extracelular neutral¹⁰³; y el operón *oppABCDF* que codifica para un sistema de transporte ABC necesario para la iniciación de la esporulación y desarrollo de competencia¹⁵¹. La transcripción de *hpr* es regulada también por SalA¹⁴⁷, siendo ésta la única regulación que se conoce que ejerce SalA, cuya expresión se ha visto que es inducida por estrés disulfúrico¹¹⁹.

Se ha mencionado que es la presencia de Spo0A fosforilada la que desencadena las cascadas de regulación que resultan en la formación de endosporas. Sin embargo, se ha observado que bajo ciertas condiciones, previo a la esporulación se forman comunidades notablemente estructuradas y asociadas en una matriz extracelular: las denominadas *biopelículas*²¹, en donde algunas de las células presentes en la comunidad, se diferencian en productoras de matriz, lo cual beneficia a toda la población. Ambos tipos celulares, las formadoras de esporas y las productoras de matriz, son regulados a través de la forma fosforilada de Spo0A. Sin embargo, los genes expresados en cada una son diferentes, que es lo que resulta en el fenotipo característico que presentan. Esta diferencia en la expresión genética se origina por la dependencia de la regulación en los niveles de fosforilación de Spo0A. Mientras que las formadoras de esporas requieren un nivel alto de fosforilación de esta proteína, el fenotipo de las productoras de matriz se observa en células que expresan niveles más bajos de Spo0A~P⁶⁴, siendo KinC y KinD – dos de las cinco cinasas encargadas de fosforilar a Spo0A- las ligadas a la formación de biopelículas⁸⁹. A bajos niveles de Spo0A fosforilada, esta proteína activa la transcripción del operón *sinIR*¹⁷⁴, cuya función principal es regular la transición de células independientes y móviles a células productoras de matriz. Esta transición se da principalmente debido a la expresión de dos operones: *epsA-O*¹⁰⁶ y *yqxM-sipW-tasA*³². El primero dirige la síntesis de exopolisacárido, uno de los principales componentes de la matriz, mientras que el segundo, codifica para tres proteínas: TasA, proteína secretada que es el otro de los componentes principales de la matriz; y YqxM, proteína que se sugiere es la responsable de la ubicación de TasA en la matriz; SipW, señal peptidasa requerida para el procesamiento y secreción de las otras dos

proteínas codificadas en dicho operón²⁰. SinR es represor de ambos operones, por lo que su expresión, mantiene prendido el estado de motilidad en *B. subtilis*, mientras que, al ser inhibida, los operones responsables de la producción de matriz son activados¹. La actividad de SinR es inhibida por el otro gen codificado en el operón SinRI: *sinI*. El operón es regulado positivamente por Spo0A y reprimido por AbrB y HPr, con lo cual se observan dos tipos celulares mutuamente excluyentes, derivados del ciclo conformado por AbrB y Spo0A: la presencia de Spo0A desencadena la formación de células productoras de matriz, mientras que, por el contrario, la expresión de AbrB mantiene a las células en su estado mótil.

Otro de los tipos celulares en los que puede diferenciarse *B. subtilis* es aquel en el cual las células presentan el comportamiento denominado *canibalismo*. Este fenómeno suele observarse en colonias en las cuales ya se ha iniciado la vía de la esporulación en algunas células, pero antes de completar el proceso, algunas de éstas retrasan la formación de la endospora a través de la producción de toxinas que liberan para destruir a células vecinas que aún no han entrado al proceso de esporulación, y a partir de las cuales se alimentan para obtener nutrientes y retrasar la formación de las endosporas. Este sistema de toxinas, también conocidas como bacteriocinas, están codificadas en dos operones: *skf* y *sdp*⁷¹. Estos operones están regulados por Spo0A y AbrB. Se ha descrito un tercer operón: *sdpRI*, que es también regulado por AbrB e inducido por *sdpC*, una de las toxinas codificadas en los genes del operón *sdp*^{47,71}, a través de un mecanismo en el cual secuestra al represor SdpR, el primer gen del operón *sdpRI*. El segundo gen de dicho operón, *sdpI* codifica para una proteína que da inmunidad frente al efecto tóxico de la bacteriocina SdpC⁴⁷. AbrB reprime la expresión del operón *sdpRI*, con lo cual se observa una vez más la importancia del circuito en el cual están involucrados Spo0A y AbrB. Las células que presentan a Spo0A expresada, han iniciado el proceso de esporulación, y en ciertos casos, presentan un comportamiento caníbal. Esto significa que *abrB* está siendo reprimido, y por tanto, *sdpRI* expresado, con lo cual la célula caníbal, presenta inmunidad frente a su propia toxina. Mientras que, las células que no expresan a Spo0A aún no se han comprometido al proceso de esporulación y son susceptibles de ser dañadas por las toxinas de los operones mencionados.

AbrB reprime también la transcripción de *fatR*⁸⁰, que codifica para un regulador del operón *fatR-cyp102A3*, que a su vez, codifica para un sistema involucrado en mediar la respuesta adaptativa frente a la toxicidad de ácidos grasos, particularmente frente al ácido linoleico^{79,80}. Sin embargo, hay poca información disponible respecto a este FT, inicialmente denominado YrhI, y renombrado FatR⁸⁰ por un grupo de investigación y BscR¹¹⁷ por otro, este último sinónimo no se encuentra registrado en la base de datos DBTBS.

AbrB está además involucrado en la regulación de la entrada al estado de competencia, otro de los tipos celulares que presenta *B. subtilis*, caracterizado por la asimilación de DNA exógeno en un proceso controlado por la propia célula. AbrB es un represor de la expresión del factor transcripcional ComK⁸¹, que es el que integra las señales que controlan el desarrollo de competencia⁷⁵. Entre los elementos que regula ComK está un sistema de transporte que se une al DNA y lo asocia a la célula; dicho sistema está codificado en los genes *comC* y *nucA*, así como en los operones *comEABC*, *comFABC-yvyF-flgM-yvyG-flgKL* y *comGABCDEFGG-yqzE*. Una vez incorporado el DNA a la célula, la proteína RecA se asocia al DNA y cataliza, en conjunto con la DNA helicasa AddAB, la integración del nuevo fragmento del DNA al genoma. El gen *recA* y el operón *addAB*, que codifican para las dos proteínas mencionadas, son también reguladas por ComK⁸².

Cuando una célula se encuentra en estado de competencia, la replicación del DNA y la división celular se bloquean; ésta podría ser la razón por la cual la expresión de ComK está regulada de manera redundante. ComK, además de ser inhibido por AbrB, es reprimido por CodY¹⁷³, otro regulador transcripcional pleiotrópico acerca del cual se hablará más adelante. La transcripción de *comK* está regulada por otros dos factores transcripcionales: Med¹⁴⁸ y Rok⁹⁰, el primero lo regula positivamente mientras el segundo reprime su expresión. Med ejerce regulación de manera independiente al resto de los factores de transcripción de la red. No se tiene aún conocimiento amplio del papel fisiológico de Med y no se conoce que su transcripción esté regulada por otro factor, tampoco se ha descrito que regule la expresión de otra proteína diferente de ComK. Podría ser una vía adicional para integrar diversas señales que activen al sistema que dirige el estado de competencia. Rok, por el contrario, es un represor que depende de otros factores, está regulado negativamente

por AbrB y SinR⁹⁰ y se ha sugerido que juega también un papel pleiotrópico en la red de regulación transcripcional de *B. subtilis*⁸². Rok inhibe la expresión de genes que codifican para proteínas que codifican para la producción de moléculas extracelulares, incluyendo antibióticos similares a las bacteriocinas. Uno de los operones que regula es *sbo-alb*, que participa en la síntesis y exportación de subtilisina³.

En este módulo y en algunos de los anteriormente descritos, se ha hecho notable la capacidad que tiene *B. subtilis* para producir grandes cantidades de enzimas degradativas extracelulares. Se ha visto también que su regulación resulta redundante y por ello los FTs que regulan su expresión están distribuidos en al menos otro de módulos analizados (regulación de enzimas degradativas). Sin embargo, se sabe que el sistema de dos componentes DegS-DegU es el mecanismo de regulación por excelencia de las enzimas degradativas en esta bacteria, siendo DegS una histidina cinasa encargada de fosforilar a DegU durante la fase estacionaria, con lo cual se inicia la secreción de exoenzimas¹¹⁴.

DegU no sólo tiene un papel relevante para el proceso de diferenciación cuando se encuentra en su estado fosforilado, sino que además regula la expresión de células móviles durante la fase exponencial, en su estado no fosforilado. La capacidad de movimiento en *B. subtilis* está principalmente asociada a la expresión del operón *fla-che*, cuyos genes codifican para proteínas que componen parte de la estructura del flagelo, así como para el sistema de quimiotaxis¹⁹⁸. Mientras que DegU~P reprime este operón⁸, DegU no fosforilado lo activa¹⁸⁹, lo que involucra a este factor de transcripción en ambos tipos celulares, haciéndolos también excluyentes.

La proteína reguladora DegU controla además la expresión del factor transcripcional SacY³⁶, ubicado en el operón *sacXY*. *B. subtilis* tiene la capacidad de utilizar sucrosa a través de dos vías: intracelular y extracelular. En esta última está involucrada una levansucrasa secretada que está codificada en *sacB*, gen regulado precisamente por SacY¹¹. En esta vía participa también SacT, que regula positivamente a SacY. Esta regulación ejercida sobre el operón *sacXY* se lleva a cabo a través del mecanismo de antiterminación¹⁸⁸.

CodY, represor de ComK, es considerado un regulador pleiotrópico involucrado en una diversidad de funciones fisiológicas. Regula genes que codifican para sistemas de transporte, como el operón *dppABCDE*, sistema de transporte para dipéptidos; genes involucrados en la producción de antibióticos, como el operón *srfAA-srfAB-comS-srfAC-srfAD*, que codifica para enzimas que catalizan la producción de sulfactina; genes que participan en esporulación y competencia, como el propio *comK*; así como otros involucrados en el metabolismo de carbono y nitrógeno; así, en general, ayuda a la célula a adaptarse a la poca disponibilidad de nutrientes en el medio^{136,180}. De manera común, la expresión de estos genes es reprimida durante un crecimiento exponencial acelerado¹⁷⁹.

CodY también reprime la expresión del factor transcripcional HutP⁵⁶, el cual regula al operón *hut*¹⁴⁶, que a su vez está involucrado en la utilización de histidina. Los genes ubicados en dicho operón codifican para enzimas que degradan la histidina en amonio y glutamato, fuentes de nitrógeno para *B. subtilis*. La transcripción del gen *hutP* es además inducida por AbrB, lo que sugiere que este último también participa en el fenómeno de represión catabólica cuando las fuentes de carbono se ven limitadas⁵⁷.

5.4.9. Módulo de regulación de respuesta a estrés general (múltiple / no específico / preventivo). (*ctsR, bmrR, mta, sigB, sigW, sigX, lytR, yvrH, fur, perR, sigM, yodB, ylaC, spx, cymR* y *ytlI*)

En este módulo se agrupan principalmente FTs involucrados en la respuesta frente a diversos tipos de estrés general. La esporulación y otras opciones de diferenciación, como se observó en la descripción del módulo pasado, es una forma que tiene *B. subtilis* para enfrentar condiciones adversas en el medio. Sin embargo, la esporulación es el último recurso que tiene esta bacteria y se presenta sólo en una limitada porción de la población en dichas condiciones, ya que es un proceso unidireccional muy costoso para la célula. Muchas células bajo estrés siguen antes algunas de las vías de diferenciación ya señaladas, pero otras se protegen justamente a través de la expresión del regulón de σ B, factor que además está ligado a otras proteínas reguladoras que mantienen a la célula en un estado vegetativo y protegido de una múltiple variedad de tipos de estrés⁸⁵. A diferencia de la

esporulación, la respuesta a estrés general por σ B es reversible, permitiendo así que la célula regrese a un estado de crecimiento normal.

σ B es el regulador maestro de respuesta a estrés general y se activa a través de una vía de señalización muy compleja. Responde a dos tipos de estrés: por falta de energía, ya sea por ausencia de proveedores de carbono, fósforo u oxígeno; y condiciones de estrés ambiental, particularmente exceso de calor, etanol, sales y ácidos¹⁵⁴. Se ha definido que, en general, las funciones de las diversas proteínas que regula este factor sigma confieren resistencia preventiva, múltiple y no específica a diferentes tipos de estrés durante fase estacionaria⁸⁵.

El regulón de σ B incluye una gran diversidad de genes involucrados en prevenir o corregir daños causados por múltiples tipos de estrés; entre ellos *dps*⁹, que se une y protege al cromosoma en situaciones de estrés por calor, por ácido y estrés oxidativo; *opuE*¹⁹³, que responde a estrés osmótico; y otros genes que codifican para proteínas que ofrecen resistencia al estrés por frío. Regula además genes que específicamente previenen o reparan daños en el DNA por estrés oxidativo, como *katE*⁴⁸ y *katX*¹⁵⁵ y genes que codifican para proteínas con un papel de detoxificación en las células: *ydfO*¹⁵⁷, *ykzA*¹⁹² y *yqgZ*¹⁵⁷.

Este factor sigma regula también la expresión de una proteína reguladora de los genes de estrés clase III: CtsR, entre los cuales se encuentran *clpC* y *clpP*, regulados tanto por σ B como por CtsR⁴². El primero codifica para ClpC, una ATPasa, mientras que el segundo lo hace para ClpP, una peptidasa; juntas forman una maquinaria que repara proteínas dañadas por estrés por calor, estrés oxidativo y obstrucción del ribosoma. σ B y CtsR corregulan también al gen *trxA*, que codifica para una tiorredoxina, otra proteína requerida en condiciones de estrés oxidativo¹¹².

Dentro del regulón de σ B, se encuentra el operón *bmrU-bmr-bmrR*¹⁵², el cual está asociado a resistencia a antibióticos. El último gen de este operón codifica para el factor transcripcional BmrR, cuya función es controlar la expresión de genes que ofrecen resistencia a drogas múltiples (MDR, por sus siglas en inglés), los cuales codifican para proteínas de transporte involucradas en la exportación de una gran variedad de compuestos

antimicrobianos⁷⁴. La expresión de *bmrR* por su parte, está sujeta también a la regulación a través de la proteína Mta¹⁵. Entre otros genes que regula Mta está el operón *blt-bltd*², que codifica también para proteínas de transporte MDR. Mta además regula la secreción de bacilibactina, sideróforo que facilita la adquisición de hierro a partir del medio¹³³.

σ B está involucrado además en la regulación de un gen cuya transcripción está controlada de manera múltiple: *spx*¹⁵³. La proteína que codifica dicho gen, Spx, es un regulador transcripcional que coordina la adaptación de *B. subtilis* a estrés disulfúrico, previniendo a la célula del estrés oxidativo que resulta en la formación de enlaces disulfúricos dañinos para la célula¹¹⁶ y activando las operaciones necesarias para revertir los efectos de este daño¹⁴³. Se sabe además que participa en el control sulfato-dependiente de operones que participan en el metabolismo de sulfuro para su correcta utilización^{31,50}. Está involucrado también en la respuesta a otro tipo de estrés oxidativo: aquel que causa daño por modificaciones covalentes de residuos de metionina, generando sulfóxido de metionina (Met-SO), otro compuesto perjudicial para la célula²⁰⁴.

Spx, por su parte, regula la expresión de otros dos factores transcripcionales: YlaC¹²⁸ y YtlI⁵⁰. El primero codifica para un factor sigma de la familia inicialmente denominada de función extracitoplásmica (ECF por sus siglas en inglés)¹²⁸ conocida ahora también como el grupo 4 de factores sigma, el cual se caracteriza por controlar la transcripción de genes involucrados en resistencia a estrés por desnaturalización de proteínas, antibióticos, oxígeno, medios muy alcalinos, calor y frío. Hasta ahora, sólo se conoce que YlaC regula la expresión del propio operón en donde se encuentra el gen que lo codifica: *ylaABCD*¹²⁸; y se ha sugerido que este sigma, en conjunto con *ylaD*, otro de los genes de dicho operón, contribuye a la resistencia por estrés oxidativo, particularmente por exposición a peróxido de hidrógeno¹⁶⁴. La segunda proteína reguladora controlada por Spx, YtlI, regula al operón *ytmIJKLMNO-ytnIJ-ribR-ytnLM*, en respuesta a la disponibilidad de sulfuro³⁵; y *ytlI* a su vez, es regulado por CymR, considerado el regulador maestro de metabolismo de sulfuro⁵². Este FT controla además la transcripción del gen que codifica para CysK, enzima importante para la biosíntesis de cisteína^{52,184}. Así, en este agrupamiento se observa el traslape de regulación por estrés y metabolismo de sulfuro.

Se ha mencionado ya que *Spx* es regulado por varios FTs. Dos de ellos tienen también una relación directa con la resistencia frente a estrés oxidativo: PerR y YodB¹¹⁸. Actualmente se sabe que en muchos sistemas la respuesta frente a estrés oxidativo está íntimamente ligada con sistemas de transporte y almacenamiento de iones metálicos, la llamada metaloregulación. Un ejemplo de esto es precisamente la agrupación de PerR en este módulo, que es el regulador maestro de respuesta a estrés por peróxido, y que a su vez regula la transcripción de *fur*⁶², que codifica para una proteína reguladora de la obtención de hierro del medio. Esto se debe en parte a que la capacidad destructiva de las especies reactivas de oxígeno puede empeorar con la presencia excesiva de hierro intracelular; y particularmente, si este metal reacciona con peróxido de hidrógeno, forma radicales hidroxilo, perjudiciales para la bacteria¹³⁷. Por lo tanto, en esta red encontramos por un lado a PerR, que es capaz de responder al estrés por peróxido. El regulón de este FT incluye a *kata*, que cataliza para una catalasa vegetativa y al operón *ahpCF*, que codifica para una alcalino hidropéroxido reductasa, ambas juegan el papel de remover los compuestos tóxicos formados por este tipo de estrés; incluye también a *mrgA*, que codifica para una proteína que protege del futuro estrés por peróxido; y al operón *hemAXCDBL*, cuyo sistema expresado son enzimas hemo biosintéticas⁸⁸. PerR inhibe además la expresión de *fur*. El regulón de Fur está principalmente integrado por genes involucrados en la biosíntesis y obtención de sideróforos, compuestos que facilitan la obtención de hierro del medio; por ejemplo, regula al operón *dhbACEBF*¹³, que codifica para una maquinaria biosintética para producir el sideróforo corinebactina. Respecto a YodB, sólo se tiene conocimiento de que, junto con PerR, inhibe la expresión de *spx* y dicha represión la ejerce frente a tratamiento de la célula con oxidantes¹¹⁸.

La transcripción de *spx* es regulada también por otros tres factores sigma de la familia ECF o grupo 4: σ_M , σ_W y σ_X . A pesar de que actualmente se reconoce que los factores sigma clasificados en este grupo tienen una variedad de funciones, originalmente se les llamó *de función extracelular* por verse asociados a la modulación de cambios en la superficie de la célula. Estos tres factores coinciden en efecto con este perfil, a la vez que responden a algunos tipos de estrés. Es importante también mencionar que los regulones de estos tres factores sigma se traslapan, regulando en general aspectos relacionados con la

homeostasis de la envoltura celular, así como resistencia a antibióticos¹²⁷. Sin embargo cada regulón tiene sus peculiaridades, las cuales serán descritas a continuación.

La expresión de σM es requerida para crecimiento y supervivencia después de que la célula es sometida a estrés salino y osmótico⁹². Está también involucrado en mantener la integridad de la membrana y la pared celular en respuesta a otros tipos de estrés, como etanol, calor, ácido, superóxido y antibióticos, entre ellos bacitracina, vancomicina y fosfomicina, todos ellos inhibidores de la biosíntesis de la pared celular¹⁸⁶. Favorece la transcripción de operones importantes para la determinación de la forma y división celular, así como para el monitoreo de daños en el DNA y enzimas de detoxificación⁴³.

σW está implicado en la producción y en la detoxificación de compuestos microbianos. Regula muchos genes cuya función aún es desconocida; sin embargo, se ha especulado que en general el regulón de este factor sigma está asociado a la *antibiosis*, incluyendo funciones para la protección de la célula frente a compuestos antimicrobianos, así como para la producción de bacteriocinas¹⁰⁰. Entre los genes cuya transcripción es regulada por σW , están algunos detoxificadores, tales como *fosB*²⁵, que le ofrece a la célula resistencia a la fosfomicina y *pbpE*⁹⁵, que codifica para una proteína capaz de unirse a la penicilina; además del gen *yqeZ*²⁷, que codifica para una proteasa unida a membrana que degrada péptidos hidrofóbicos.

Al igual que σW , la expresión de σX es inducida a partir de la presencia de antibióticos actuando en la pared celular. Los productos codificados en los genes que regula este factor sigma funcionan modificando la composición y las propiedades de la envoltura celular como mecanismo para protegerse de agentes antibacterianos. En este regulón se encuentran dos operones directamente involucrados en dicha función de modificación: *dltABCDE* y *pssA-ybfM-psd*²⁶. El primero codifica para proteínas requeridas para la modificación química de los ácidos teicoicos ubicados en la pared celular, lo que tiene relevancia fisiológica, tanto para la propia resistencia a los compuestos antimicrobianos como para la tasa de autólisis de las células. El segundo codifica para la maquinaria de biosíntesis del lípido neutral de membrana fosfatidilserina decarboxilasa (PE por sus siglas en inglés), el cual es una vía alterna para lidiar con antibióticos catiónicos, ya que PE

modula la carga neta de la superficie celular. El gen *lytR* también es regulado por σX^{96} y codifica para una proteína reguladora que juega un papel atenuador para el operón *lytABC*, que a su vez codifica para un sistema de secreción de autolisinas; por lo tanto, LytR está asociado a la disminución de la expresión de estas enzimas. σX es además controlado por *yvrH*, que codifica para una proteína reguladora que forma parte del sistema de dos componentes YvrGHb, cuya función en general está relacionada con la regulación de funciones homeostáticas de la superficie celular. Este sistema también reprime la expresión del operón *lytABC*, previniendo la autólisis celular¹⁷¹.

Como puede observarse, las proteínas y factores sigma agrupados en este módulo, regulan a genes cuya expresión ofrece protección a una célula que ya no está en crecimiento, resultando fundamental para lidiar con futuros factores de estrés en la supervivencia a largo plazo.

5.5. La fisiología de *B. subtilis* reflejada en la organización de su red de regulación transcripcional

5.5.1. El papel de los factores sigma en la red

A diferencia del trabajo de Resendis-Antonio O. *et al.*¹⁶¹, que hace una descripción de los módulos funcionales derivados del análisis aplicado a la red de regulación de FTs de *E. coli*, excluyendo a los factores sigma; el trabajo presentado en esta tesis, incluye a ambos componentes de la red. Esto se decidió a partir del hecho documentado bibliográficamente de que los factores sigma son determinantes para la respuesta expresada por *B. subtilis* frente a diversas condiciones.

Una de nuestras primeras aproximaciones fue aplicar los mismos análisis descritos en este escrito a la misma red en ausencia de los factores sigma (ver Apéndice 6). Como resultado, obtuvimos que la red se descompone en un número muy elevado de módulos pequeños, asociados a funciones repetitivas en todos ellos. Además de separar a la red, el

quitar todos los factores sigma disminuye el número de FTs incluidos en el análisis (desaparecen aquéllos que sólo están conectados con un factor sigma), lo que en ciertas ocasiones también deriva en la ausencia de funciones observadas en los módulos resultantes.

Por el contrario, la inclusión de factores sigma generó módulos cohesivos que pudieron asociarse a funciones fisiológicas de la célula. Asimismo, los once factores sigma incluidos en el análisis se agrupan en módulos cuya función asignada es también la función a la que el factor sigma se le ha visto asociada. El primero en ser descrito fue σ_L , agrupado en el módulo de metabolismo de carbono de acuerdo a la manera en que está conectado al resto de la red pero coincidiendo con sus funciones como regulador de genes involucrados en el metabolismo de fructosa, levadura, arginina, acetoina, leucina y valina.

Respecto a los factores σ_E , σ_F , σ_G , σ_H y σ_K , se sabe actualmente con certeza que sus respectivas cascadas son las responsables de coordinar el proceso de esporulación en *B. subtilis*. El análisis modular realizado en el presente trabajo agrupó a σ_K y a σ_E en el módulo de la cascada de regulación en la célula madre, mientras que σ_H , σ_G y σ_F fueron ubicados en el módulo de diferenciación. A pesar de que estos cinco factores sigma asociados a una misma función fueron separados en el análisis resultante, situación que se explicará más adelante, ambos módulos se encuentran asociados al proceso de esporulación. El primero describe la expresión de genes en la célula madre durante dicho proceso, mientras que el segundo agrupa a todos los factores involucrados en la generación de distintos tipos celulares en *B. subtilis*, siendo el de la formación de la endospora uno de ellos.

Los restantes cinco factores sigma incluidos en el análisis son σ_B , σ_M , σ_W , σ_X y $YlaC$. El primero es considerado el regulador maestro de respuesta a estrés en esta bacteria Gram positiva, mientras que los otros cuatro pertenecen al grupo 4 de los factores sigma, también llamados familia ECF, caracterizados por responder a diversos factores de estrés, lo cual es consistente con la forma en la que fueron agrupados en el análisis. Los cinco se ubican en el último módulo descrito, cuya función asociada es la respuesta a estrés general: múltiple, no específico y preventivo.

5.5.2. Diferentes funciones en un mismo módulo

Una de las principales razones por la cual se llevó a cabo el estudio detallado de la Biología de cada uno de los módulos fue por el hecho de que la base de datos SubtiList, la cual agrupa a todos los genes de *B. subtilis* y los clasifica dependiendo de su función, no ha sido actualizada desde el año de 2001, con lo cual resultaba imposible la definición de una función discreta para un módulo compuesto por uno o más genes con función desconocida. Aún más complicado resultó el hecho de que en esta clasificación, todos los FTs se agrupan en la misma clase: *Síntesis de RNA: Regulación transcripcional*, por su papel como proteínas reguladoras.

A pesar de que una anotación manual de la clasificación fisiológica de estas proteínas a partir de los datos obtenidos en SubtiList, en conjunto con una breve revisión bibliográfica de los genes sin función definida en dicha base de datos, pudo haber sido suficiente, hubo otra razón que apoyó la idea de llevar a cabo un análisis minucioso de los componentes de cada módulo. Ésta es que, a pesar de que existe un notable número de trabajos que describen la organización topológica de diversas redes de regulación, son pocas aquéllas que analizan la Biología de dicha organización. En el presente trabajo no se deseó omitir este aspecto tan importante de un análisis topológico. Esto permitió analizar, más allá del método utilizado para descomponer a la red en módulos, el porqué de las agrupaciones formadas. Se observan, en pocas ocasiones, más de una función agrupada en un mismo módulo, y por el contrario, más de un módulo presentando una misma función. En la presente sección, así como en la siguiente, se discuten estos casos particulares.

Uno de los módulos que presenta más de una función es el de respiración aerobia y anaerobia. En dicho módulo, cuatro de las seis proteínas (ArfM, Fnr, NsrR y ResD) son parte de una cascada que se desencadena en respuesta a cambios en la disponibilidad de oxígeno en el medio, regulando la expresión de mecanismos que le permitan adaptarse a ambientes anaerobios. Se agrupa además el gen que codifica para PhoP, cuya expresión se induce cuando decrece la presencia de fosfato en el medio. PhoP y el regulador maestro que dirige la cascada mencionada, ResD, se regulan mutuamente de manera positiva. Esta inducción codependiente refleja la relación entre ambas funciones fisiológicas. Una explicación a lo anterior es que la inducción de la expresión de ResD durante escasez de fuentes de fosfato, provee componentes esenciales para el transporte de electrones, a través

de la expresión de una proteína similar a la de biogénesis de citocromo, así como de biosíntesis del grupo hemo A, necesarios para la asimilación de fosfato inorgánico en ATP. Sería relevante entender la importancia de la expresión de PhoP a la par de la expresión de la cadena que se expresa en función de la disponibilidad de oxígeno en el medio; siendo una posibilidad el simple hecho de amplificar los reguladores para enfrentar estas variaciones.

La sexta proteína agrupada en este módulo, HrcA, es un represor de proteínas inducidas por estrés por calor. Más adelante se discute la redundante regulación de este tipo de proteínas en varios módulos; sin embargo la agrupación de HrcA junto con factores involucrados en respiración aerobia y anaerobia se explica a partir del hecho de que *hrcA*, el gen que codifica para dicha proteína, se ubica en el mismo operón que *hemN*, que codifica para una proteína necesaria para el crecimiento anaeróbico, así como para la biosíntesis de grupos hemo, necesarios para la transportación de oxígeno, entre otras funciones.

El módulo de diferenciación es otro caso similar que merece ser discutido. Es el que contiene la mayor cantidad de TFs agrupados, y a pesar de que la mayoría de ellos tienen una implicación directa en el fenotipo expresado en cada tipo celular, hay algunos genes cuya presencia requiere ser explicada. Dos de ellos son los que codifican para Hbsu y LexA, regulados por σ_H , factor sigma involucrado en la esporulación. Estas dos proteínas reguladoras están asociadas a la estructuración y compactación del DNA, así como al cuidado y protección del mismo. Las proteínas que se encargan de organizar, compactar y en general interactuar con el DNA, tal y como lo hace HBSu, homólogo de la proteína HU en *E. coli*, tienen un importante efecto sobre la expresión de genes y el comportamiento de crecimiento; lo que las hace elementales para la sobrevivencia de la célula. De hecho, se sabe que HBSu se expresa no sólo durante la esporulación, sino también durante el crecimiento vegetativo. Respecto a su papel en el proceso de formación de esporas, se ha demostrado una eficacia reducida en la esporulación en mutantes del gen que la codifica. De esta manera, el requerimiento de un sistema que proteja al DNA durante un proceso en el cual esta molécula tiene que llevar a cabo una diversidad de operaciones, así como la participación de una proteína que tiene el papel de compactar el DNA, explica la

agrupación de ambos FTs en este módulo. Sin embargo, debe notarse que su expresión no es exclusiva del proceso de esporulación.

Otra de las proteínas no implicadas de manera directa en la expresión de genes para cada destino celular es FatR, involucrado en la respuesta frente a la toxicidad de ácidos grasos. En la literatura hay poca información de este FT, lo que además resulta difícil para determinar su papel durante la diferenciación. Resulta extraño además no encontrarla regulada por alguno de los FTs agrupados en el módulo de estrés, ya que muchos de ellos controlan la respuesta frente a condiciones tóxicas diversa índole. Sin embargo, se ha reportado que σM regula al gen *yrhJ*¹⁸⁶, ubicado en el mismo operón en donde se encuentra *fatR*. Esto implica que es muy probable que la expresión de FatR esté también regulada por σM , que es precisamente uno de los componentes del módulo de estrés. Esta situación reagruparía a este FT, asignándolo probablemente en dicho módulo. En cualquier caso, es necesario un estudio más detallado de la función y la regulación de FatR.

En este mismo módulo de diferenciación se agrupa también HutP, regulado tanto por CodY y AbrB a nivel intramodular, como por CcpA, que pertenece a un módulo externo (metabolismo de carbono). A pesar de que HutP está directamente implicado en la utilización de fuentes alternativas de carbono y nitrógeno, su agrupación en este módulo se debe a que su expresión es regulada por CodY y AbrB ubicados en el módulo de diferenciación, ambos descritos en la literatura como factores globales, implicados limitadamente, también en el metabolismo de carbono. Sin embargo, en este caso hace falta analizar la relevancia de HutP en el módulo de diferenciación, ya que es el único FT de metabolismo de carbono reportado corregulado por estos FTs. No obstante, en el mismo módulo se observa un ejemplo similar: la presencia de SacT y SacY. A pesar de que ambas están involucradas en la utilización de sucrosa, se agrupan con estos FTs ya que la regulación de *sacY* está mediada por CodY. La relación radica en que a través de la regulación del metabolismo de sucrosa, se regula una proteína secretada: la levansucrasa codificada en *sacB*. Esto tiene implicaciones en el fenotipo del tipo celular asociado a la regulación de enzimas en *B. subtilis*. Así, a la par de estar involucradas en el metabolismo de sucrosa, SacT y SacY también participan en la expresión de genes particulares de las células degradativas durante el proceso de diferenciación.

En el módulo de estrés se observa un caso similar: además de incluir los FTs involucrados en la respuesta a diversos tipos de estrés, entre ellos el disulfúrico, se agrupan dos genes asociados al metabolismo de sulfuro: *ytlI* y *cymR*. Sin embargo, esto tiene sentido a partir del hecho de que la proteína Spx cumple una doble función: regula la respuesta a estrés disulfúrico e inhibe los operones involucrados en el metabolismo de sulfuro cuando éste está presente en el medio. Es esta proteína la responsable de que *ytlI* y *cymR* hayan sido agrupados en dicho módulo.

5.5.3. Funciones presentes en más de un módulo

El análisis modular arrojó dos funciones que aparecen reguladas de manera redundante y a través de módulos independientes en la red. Por un lado se observan varios módulos asociados a la regulación de proteínas de estrés por calor y, por otro, dos módulos totalmente desconectados entre sí en la red, que regulan la producción de enzimas degradativas. A continuación se discutirán ambas situaciones.

Una de las adversidades a las que tienen que enfrentarse las bacterias es el cambio de temperatura en el ambiente, para lo cual han desarrollado complejos programas de regulación genética. La respuesta a estrés por calor es uno de los sistemas mejores conocidos e involucra a las denominadas proteínas de estrés por calor (Hsp por sus siglas en inglés). En *B. subtilis* se han descrito ya una diversidad de proteínas de este tipo, reguladas a través de distintos mecanismos. En el análisis que se llevó a cabo en este trabajo, esta variedad en los mecanismos de regulación se hace evidente al encontrar más de un módulo involucrado en el control de la expresión de estas proteínas.

Tradicionalmente, las proteínas Hsp de *B. subtilis* han sido clasificadas en tres clases⁸⁶ recientemente incrementadas a seis¹⁷⁰. Dos de ellas son reguladas por represores transcripcionales, una clase es regulada a través de un factor sigma, una cuarta clase es regulada por un FT desconocido, otra más es controlada por un sistema de dos componentes y, la última, incluye una notable cantidad de proteínas Hsp cuyo mecanismo de regulación es aún desconocido. Sin embargo, a pesar de estar todas ellas clasificadas como proteínas de estrés por calor, muchas de ellas también responden a otra variedad de estímulos de estrés, como lo es la clase II, que se hace referencia al regulón de σ^B .

Asimismo, la clase III, regulada por el factor de transcripción CtsR, además de estrés por calor, está involucrada a la respuesta a estrés oxidativo y por obstrucción del ribosoma. Se sabe también de la importancia de las proteínas de la clase V, regulada por el sistema C_{ss}RS, para responder frente a estrés de secreción. De las restantes cuya regulación es conocida sólo la clase I, en donde se ubican las proteínas reguladas por HrcA, está asociada únicamente a la respuesta de estrés por calor.

Estas cuatro clases se reparten en tres de los nueve módulos definidos en el presente trabajo. La clase I se ubica en el módulo de regulación aerobia y anaerobia, lo cual como ya se explicó en la sección 5.5.2, se debe a que se ubica en el mismo operón que una proteína indispensable en condiciones de anaerobiosis. Sin embargo, la propia importancia de HrcA en la cascada de regulación observada en la respiración anaerobia, no ha sido estudiada ni discutida. Es importante también mencionar que, sin la regulación ejercida por el FT del módulo de respiración (ArfM), HrcA es capaz de responder al estrés por calor de manera independiente a dicha proteína; el operón *groESL*, que codifica para chaperoninas de la clase I de proteínas Hsp, es regulado por HrcA, pero no tiene sitios de unión con ArfM. No obstante, encontrar a HrcA desconectada en la red del resto de los mecanismos de regulación de proteínas Hsp, apoya la idea de que es un sistema aislado. Respecto a la clase II y clase III, reguladas por σ B y CtsR respectivamente, se agruparon en el mismo módulo: el de estrés general. Esto sucedió debido a que la expresión de la proteína reguladora es controlada también por σ B, lo que sugiere que una fracción de la clase III está contenida en la clase II. La clase V de este tipo de proteínas fue agrupada en otro módulo aislado: el de respuesta a estrés de secreción, en donde el factor transcripcional C_{ss}R, como parte del sistema de dos componentes C_{ss}RS, es el regulador responsable de la expresión de proteínas de este tipo.

Lo anterior hace evidente, a partir del análisis modular de FTs de *B. subtilis*, la clasificación de los mecanismos reguladores de las proteínas de estrés por calor. Este tipo de análisis resultaría muy útil tomando redes de regulación de otros organismos y observando el módulo o módulos que contienen a los reguladores de estas proteínas. Por ejemplo, a *E. coli* sólo se le conoce un regulador; mientras que otras bacterias Gram positivas presentan varios mecanismos, al igual que *B. subtilis*, con la diferencia de que el

traslape entre las clases difiere del que se presenta en este estudio. Por lo tanto, valdría la pena hacer comparaciones de la topología de los mecanismos que regulan a las proteínas Hsp. Respecto a los observados en *B. subtilis*, se ha llegado incluso a la conclusión de que el sistema que tiene esta bacteria para regular a sus proteínas de estrés por calor es el más complicado que se ha descrito en bacterias. Esta red tan sofisticada puede ser explicada a partir del hecho de que esta bacteria, dados los ambientes en los que se desarrolla, está sujeta a una gran y compleja variedad de tipos de estrés. Sin embargo, aún queda en el aire la misma pregunta planteada en el final de uno de los más recientes artículos de revisión de este tema: “¿Cuál es la relevancia biológica de dividir los genes Hsp en tantos regulones diferentes?”¹⁷⁰.

La presencia de una misma función en varios módulos también se observa en la regulación de enzimas degradativas, que aparece en el propio módulo de regulación de enzimas degradativas, así como en el módulo de diferenciación. En el primero, la producción de dichas enzimas es regulada por TenA, mientras que en el segundo la regulación la ejercen DegU y SenS. La importancia de este tipo de proteínas radica en que *B. subtilis* las utiliza para degradar compuestos complejos y utilizarlos como fuentes alternativas de energía. Como se describió en el módulo de diferenciación, DegU es el regulador maestro que coordina la expresión de estas enzimas en situaciones en las cuales hay deficiencia de nutrientes en el medio. Así en dicho módulo vemos el papel de las enzimas como responsables del fenotipo de uno de los tipos celulares de *B. subtilis* en una biopelícula. No obstante, es poco lo que puede decirse de TenA en el otro módulo. Existe muy poca información de este mecanismo de regulación y mutantes del gen que codifica para esta proteína reguladora, señalan que ésta no es esencial para el crecimiento ni para la propia producción de las enzimas. Se ha especulado que el papel de TenA es compensar la función de *degU* o *senS* cuando se presenta alguna mutación de estos genes. Esto podría explicar la redundancia de la regulación de las enzimas; sin embargo, existe también la posibilidad de que ambos grupos de FTs estén conectados a través de alguna proteína reguladora aún no descrita.

La esporulación es otro de los casos evidentes en los cuales una función aparece en más de un módulo. En el análisis, hay una agrupación asociada a la cascada de regulación que se da en la célula madre durante el proceso de esporulación y, a la par, se observa otra

serie de genes involucrados en la esporulación que se agrupan en el módulo de diferenciación. Esto últimos consisten básicamente en Spo0A y σ H, ambos involucrados en la iniciación de dicho proceso, así como los genes expresados en la cascada de regulación que se da en la preespora. La separación de ambos grupos de genes en dos módulos tiene su origen en la propia compartimentalización de la expresión genética durante la formación de la endospora. A partir de que la proteína Spo0A fosforilada desencadena el proceso, la célula madre y la preespora expresan una cadena de genes totalmente independiente una de la otra a nivel de regulación transcripcional. Sin embargo, sí existe comunicación entre ambos compartimentos, pero ésta es sólo a través de una compleja vía de señalización. Incluyendo este tipo de interacciones, la red asociaría probablemente a ambos grupos en un mismo módulo. Sin embargo, utilizando sólo la red de regulación de los FTs, los resultados agrupan de manera adecuada a la cascada de la madre independiente de la cascada de la preespora, esta última agrupada con el resto de FTs involucrados en las distintas vías de diferenciación que presenta *B. subtilis*, todas ellas juntas por las razones que ya se explicaron en dicho módulo: muchas de ellas son mutuamente excluyentes.

5.5.4. Entonces... ¿sí hay una relación módulo-función? Reflexiones finales

A través del análisis de la Biología de cada uno de los componentes de todos los módulos, fue posible asociar una función fisiológica a cada uno de los agrupamientos. Sin embargo, es necesario aclarar que, como se esperaba, no se encontraron módulos totalmente independientes ni completamente homogéneos. Respecto a lo primero, se observan genes pertenecientes a un módulo conectados con genes de otros módulos. Esto es el reflejo de la manera en que los componentes de la célula están conectados y de cómo una función tiene influencia sobre otra. Es particularmente evidente como el módulo de metabolismo de carbono está conectado con cuatro de los ocho módulos restantes, lo cual se traduce en la importancia de la utilización de este elemento para poder llevar a cabo cualquier otra función.

Refiriéndonos ahora a la homogeneidad de los componentes de los módulos, las secciones 5.5.2 y 5.5.3 explicaron los pocos casos en donde esta propiedad no se cumple.

En la mayoría de éstos, el análisis de la Biología de los componentes ofrece una razón coherente que explica este agrupamiento, con lo que además se entiende que la relación módulo-función no es totalmente rígida. Sin embargo, también se observan casos en los cuales es difícil explicar el agrupamiento de un componente en un módulo, lo cual puede deberse a las herramientas utilizadas para el análisis, como es la base de datos o la falta de información de ciertos genes, así como el propio método. A pesar de esto, el análisis bibliográfico ayudó a definir de manera concreta una función para cada agrupamiento.

Otra idea que se quisiera rescatar es que, a pesar de que la definición reciente de modularidad en Biología hace referencia al concepto de función como tal, en donde función es una tarea particular que lleva a cabo la célula; en este trabajo se observa una relación módulo-respuesta, lo que significa que los genes agrupados en entidades cohesivas comparten la característica de responder frente a una condición determinada. Esto podría parecer evidente, pero refleja que el sistema está organizado en entidades con un papel determinado dada una serie de estímulos. Aunque podrían parecer dos conceptos idénticos, función y respuesta, la diferencia radica en que un ejemplo de dos funciones independientes podrían ser producción de antibióticos y biosíntesis de lípidos, mientras que un módulo de respuesta a una condición particular, podría incluir uno o varios genes de las dos funciones mencionadas. Sin embargo, a lo largo de este escrito se manejó principalmente el concepto de función como tal, dejando como sugerencia la reflexión de esta última observación para futuros análisis de redes de regulación transcripcional.

6. CONCLUSIONES y PERSPECTIVAS

6.1. Conclusiones

El trabajo presentado en esta tesis analiza la red de regulación transcripcional de *Bacillus subtilis* desde un enfoque topológico. Para ello, se llevó a cabo la reconstrucción de la red con el objetivo de determinar sus propiedades estadísticas. La red estudiada presenta las características de una libre de escala y tiene una organización jerárquico-modular. La determinación de estas propiedades topológicas permitió realizar un análisis de la red de FTs de *B. subtilis* y descomponerla en nueve módulos discretos, los cuales están altamente conectados de manera intramodular y combinados entre ellos jerárquicamente, formando unidades más grandes y menos cohesivas.

El análisis de la Biología de los módulos encontrados permitió asociar una función fisiológica a cada entidad. A pesar de que los módulos no son totalmente independientes ni homogéneos, los componentes de cada uno de ellos comparten la característica de responder frente a ciertas condiciones o estímulos. Se observan relaciones intermodulares establecidas en su mayor parte por FTs globales o pleiotrópicos; sin embargo no necesariamente todos los FTs globales están altamente conectados en la fracción de red analizada modularmente.

Por otro lado, la inclusión de los factores sigma fue una aportación importante en relación al trabajo realizado por Resendis-Antonio y colaboradores¹⁶¹ en *E. coli*. Esto permitió incluir en el análisis un notable número de funciones reguladas por cascadas σ , entre ellas el proceso de esporulación, ampliamente estudiado y caracterizado experimentalmente en las últimas décadas. Existe una gran cantidad de datos acerca de la regulación de este proceso y es importante hacer notar que este sesgo en la información disponible se refleja en los resultados de este trabajo.

Finalmente, se hace énfasis en la organización del sistema estudiado en esta investigación, que confirma las propiedades esperadas de una red biológica. La

descomposición en módulos discretos, por su parte, ofrecen una visión global de los mecanismos de respuesta que tiene que enfrentar *B. subtilis* en el medio en el que vive.

6.2. Perspectivas

A partir de la elaboración de este trabajo y de lo concluido anteriormente, se sugieren los siguientes puntos para futuras investigaciones:

- Llevar a cabo la curación de la información agrupada en las bases de datos disponibles, la cual requiere ser corregida y actualizada. El presente trabajo puede ser una aportación para ello.
- Hacer un análisis más profundo de los motivos de regulación, los cuales finalmente son otro nivel de organización modular.
- Realizar una comparación de éste con otros métodos, incluyendo el de la descomposición en orígenes y el de descomposición natural propuesto por Freyre-González J.A. *et al.*, con el fin de entender la organización de la red desde distintos enfoques, así como para evaluar los diferentes métodos disponibles que existen actualmente.
- Repetir el análisis de este trabajo, incluyendo ahora otros elementos reguladores de la expresión genética. Otra alternativa es la integración de distintos tipos de redes biológicas, tal como las redes de interacción proteína-proteína.
- Generar modelos dinámicos de algunos de los módulos encontrados en este trabajo. El de diferenciación sería particularmente relevante de analizar desde este enfoque.
- Comparar los resultados de esta red con la de otras redes de regulación transcripcional.

7. BIBLIOGRAFÍA

1. Aguilar,C., Vlamakis,H., Losick,R. & Kolter,R. Thinking about *Bacillus subtilis* as a multicellular organism. *Curr. Opin. Microbiol.* **10**, 638-643 (2007).
2. Ahmed,M. *et al.* Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *J. Bacteriol.* **177**, 3904-3910 (1995).
3. Albano,M. *et al.* The Rok protein of *Bacillus subtilis* represses genes for cell surface and extracellular functions. *J. Bacteriol.* **187**, 2010-2019 (2005).
4. Albert,R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947-4957 (2005).
5. Albert,R. & Barabasi,A.L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47-97 (2002).
6. Ali,N.O., Bignon,J., Rapoport,G. & Debarbouille,M. Regulation of the acetoin catabolic pathway is controlled by sigma L in *Bacillus subtilis*. *J. Bacteriol.* **183**, 2497-2504 (2001).
7. Alonso-Pavon,J.A. Caracterización y análisis de la red de regulación transcripcional de *Bacillus subtilis*. 2008. Tesis de licenciatura. Licenciatura en Ciencias Genómicas, Universidad Nacional Autónoma de México. Cuernavaca, Morelos.
8. Amati,G., Bisicchia,P. & Galizzi,A. DegU-P represses expression of the motility fla-che operon in *Bacillus subtilis*. *J. Bacteriol.* **186**, 6003-6014 (2004).
9. Antelmann,H. *et al.* Expression of a stress- and starvation-induced *dps/pexB*-homologous gene is controlled by the alternative sigma factor sigmaB in *Bacillus subtilis*. *J. Bacteriol.* **179**, 7251-7256 (1997).
10. Au,N. *et al.* Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.* **187**, 7655-7666 (2005).
11. Aymerich,S. & Steinmetz,M. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc. Natl. Acad. Sci. U. S. A* **89**, 10410-10414 (1992).
12. Bagyan,I., Hobot,J. & Cutting,S. A compartmentalized regulator of developmental gene expression in *Bacillus subtilis*. *J. Bacteriol.* **178**, 4500-4507 (1996).
13. Baichoo,N., Wang,T., Ye,R. & Helmann,J.D. Global analysis of the *Bacillus subtilis* Fur regulon and the iron starvation stimulon. *Mol Microbiol.* **45**, 1613-1629 (2002).

14. Barabasi,A.L. & Albert,R. Emergence of scaling in random networks. *Science* **286**, 509-512 (1999).
15. Baranova,N.N., Danchin,A. & Neyfakh,A.A. Mta, a global MerR-type regulator of the Bacillus subtilis multidrug-efflux transporters. *Mol Microbiol.* **31**, 1549-1559 (1999).
16. Belitsky,B.R., Wray,L.V., Jr., Fisher,S.H., Bohannon,D.E. & Sonenshein,A.L. Role of TnrA in nitrogen source-dependent repression of Bacillus subtilis glutamate synthase gene expression. *J. Bacteriol.* **182**, 5939-5947 (2000).
17. Bera,A.K., Zhu,J., Zalkin,H. & Smith,J.L. Functional dissection of the Bacillus subtilis pur operator site. *J. Bacteriol.* **185**, 4099-4109 (2003).
18. Birkey,S.M., Liu,W., Zhang,X., Duggan,M.F. & Hulett,F.M. Pho signal transduction network reveals direct transcriptional regulation of one two-component system by another two-component regulator: Bacillus subtilis PhoP directly regulates production of ResD. *Mol Microbiol.* **30**, 943-953 (1998).
19. Blencke,H.M. *et al.* Regulation of citB expression in Bacillus subtilis: integration of multiple metabolic signals in the citrate pool and by the general nitrogen regulatory system. *Arch. Microbiol.* **185**, 136-146 (2006).
20. Branda,S.S., Chu,F., Kearns,D.B., Losick,R. & Kolter,R. A major protein component of the Bacillus subtilis biofilm matrix. *Mol Microbiol.* **59**, 1229-1238 (2006).
21. Branda,S.S., Gonzalez-Pastor,J.E., Ben Yehuda,S., Losick,R. & Kolter,R. Fruiting body formation by Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A* **98**, 11621-11626 (2001).
22. Britton,R.A. *et al.* Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of Bacillus subtilis. *J. Bacteriol.* **184**, 4881-4890 (2002).
23. Brown,S.W. & Sonenshein,A.L. Autogenous regulation of the Bacillus subtilis glnRA operon. *J. Bacteriol.* **178**, 2450-2454 (1996).
24. Burbulys,D., Trach,K.A. & Hoch,J.A. Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. *Cell* **64**, 545-552 (1991).
25. Cao,M., Bernat,B.A., Wang,Z., Armstrong,R.N. & Helmann,J.D. FosB, a cysteine-dependent fosfomycin resistance protein under the control of sigma(W), an extracytoplasmic-function sigma factor in Bacillus subtilis. *J. Bacteriol.* **183**, 2380-2383 (2001).

26. Cao,M. & Helmann,J.D. The *Bacillus subtilis* extracytoplasmic-function sigmaX factor regulates modification of the cell envelope and resistance to cationic antimicrobial peptides. *J. Bacteriol.* **186**, 1136-1146 (2004).
27. Cao,M. *et al.* Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/microarray analysis (ROMA), and transcriptional profiling approaches. *J. Mol Biol.* **316**, 443-457 (2002).
28. Chastanet,A. & Losick,R. Engulfment during sporulation in *Bacillus subtilis* is governed by a multi-protein complex containing tandemly acting autolysins. *Mol Microbiol.* **64**, 139-152 (2007).
29. Chauvaux,S., Paulsen,I.T. & Saier,M.H., Jr. CcpB, a novel transcription factor implicated in catabolite repression in *Bacillus subtilis*. *J. Bacteriol.* **180**, 491-497 (1998).
30. Choi,S.K. & Saier,M.H., Jr. Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis*: potential connection between carbon and nitrogen metabolism. *J. Bacteriol.* **187**, 6856-6861 (2005).
31. Choi,S.Y., Reyes,D., Leelakriangsak,M. & Zuber,P. The global regulator Spx functions in the control of organosulfur metabolism in *Bacillus subtilis*. *J. Bacteriol.* **188**, 5741-5751 (2006).
32. Chu,F., Kearns,D.B., Branda,S.S., Kolter,R. & Losick,R. Targets of the master regulator of biofilm formation in *Bacillus subtilis*. *Mol Microbiol.* **59**, 1216-1228 (2006).
33. Cohn,F. Untersuchungen über *Bakterien*. *Beiträge zur Biologie der Pflanzen* **1**, 127-224 (1872).
34. Collado-Vides,J., Magasanik,B. & Gralla,J.D. Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**, 371-394 (1991).
35. Coppee,J.Y. *et al.* Sulfur-limitation-regulated proteins in *Bacillus subtilis*: a two-dimensional gel electrophoresis study. *Microbiology* **147**, 1631-1640 (2001).
36. Crutz,A.M. & Steinmetz,M. Transcription of the *Bacillus subtilis* *sacX* and *sacY* genes, encoding regulators of sucrose metabolism, is both inducible by sucrose and controlled by the DegS-DegU signalling system. *J. Bacteriol.* **174**, 6087-6095 (1992).
37. Cruz,R.H. *et al.* Fermentative metabolism of *Bacillus subtilis*: physiology and regulation of gene expression. *J. Bacteriol.* **182**, 3072-3080 (2000).

38. Cutting,S., Panzer,S. & Losick,R. Regulatory studies on the promoter for a gene governing synthesis and assembly of the spore coat in *Bacillus subtilis*. *J. Mol Biol.* **207**, 393-404 (1989).
39. Darmon,E. *et al.* A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of *Bacillus subtilis*. *J. Bacteriol.* **184**, 5661-5671 (2002).
40. Debarbouille,M., Gardan,R., Arnaud,M. & Rapoport,G. Role of bkdR, a transcriptional activator of the sigL-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*. *J. Bacteriol.* **181**, 2059-2066 (1999).
41. Debarbouille,M., Martin-Verstraete,I., Kunst,F. & Rapoport,G. The *Bacillus subtilis* sigL gene encodes an equivalent of sigma 54 from gram-negative bacteria. *Proc. Natl. Acad. Sci. U. S. A* **88**, 9092-9096 (1991).
42. Derre,I., Rapoport,G. & Msadek,T. CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol Microbiol.* **31**, 117-131 (1999).
43. Eiamphungporn,W. & Helmann,J.D. The *Bacillus subtilis* sigma(M) regulon and its contribution to cell envelope stress responses. *Mol Microbiol.* **67**, 830-848 (2008).
44. Eichenberger,P. *et al.* The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS. Biol.* **2**, e328 (2004).
45. Eichenberger,P. *et al.* The sigmaE regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *J. Mol Biol.* **327**, 945-972 (2003).
46. Eisen,M.B., Spellman,P.T., Brown,P.O. & Botstein,D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A* **95**, 14863-14868 (1998).
47. Ellermeier,C.D., Hobbs,E.C., Gonzalez-Pastor,J.E. & Losick,R. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell* **124**, 549-559 (2006).
48. Engelmann,S., Lindner,C. & Hecker,M. Cloning, nucleotide sequence, and regulation of katE encoding a sigma B-dependent catalase in *Bacillus subtilis*. *J. Bacteriol.* **177**, 5598-5605 (1995).
49. Erdos,P. & Renyi,A. On the Evolution of Random Graphs. *Bulletin of the International Statistical Institute* **38**, 343-347 (1960).
50. Erwin,K.N., Nakano,S. & Zuber,P. Sulfate-dependent repression of genes that function in organosulfur metabolism in *Bacillus subtilis* requires Spx. *J. Bacteriol.* **187**, 4042-4049 (2005).

51. Euler,L. *Solutio problematis ad geometriam situs pertinentis. Commentarii Academiae Scientiarum Imperialis Petropolitanae* **8**, 128-140 (1741).
52. Even,S. *et al.* Global control of cysteine metabolism by CymR in *Bacillus subtilis*. *J. Bacteriol.* **188**, 2184-2197 (2006).
53. Fernandez,S. & Alonso,J.C. *Bacillus subtilis* sequence-independent DNA-binding and DNA-bending protein Hbsu negatively controls its own synthesis. *Gene* **231**, 187-193 (1999).
54. Feucht,A., Evans,L. & Errington,J. Identification of sporulation genes by genome-wide analysis of the sigmaE regulon of *Bacillus subtilis*. *Microbiology* **149**, 3023-3034 (2003).
55. Fisher,S.H. Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la difference! *Mol Microbiol.* **32**, 223-232 (1999).
56. Fisher,S.H., Rohrer,K. & Ferson,A.E. Role of CodY in regulation of the *Bacillus subtilis* hut operon. *J. Bacteriol.* **178**, 3779-3784 (1996).
57. Fisher,S.H., Strauch,M.A., Atkinson,M.R. & Wray,L.V., Jr. Modulation of *Bacillus subtilis* catabolite repression by transition state regulatory protein AbrB. *J. Bacteriol.* **176**, 1903-1912 (1994).
58. Floyd,W.F. Algorithm 97: Shortest path. *Communications of the ACM* **5**, 345 (1962).
59. Forster,J., Famili,I., Fu,P., Palsson,B.O. & Nielsen,J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244-253 (2003).
60. Freyre-Gonzalez,J.A. Disección de la arquitectura funcional de la red de regulación transcripcional de *Escherichia coli*: Un enfoque de descomposición natural. 2009. Tesis Doctoral. Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México.
61. Freyre-Gonzalez,J.A., Alonso-Pavon,J.A., Trevino-Quintanilla,L.G. & Collado-Vides,J. Functional architecture of *Escherichia coli*: new insights provided by a natural decomposition approach. *Genome Biol.* **9**, R154 (2008).
62. Fuangthong,M., Herbig,A.F., Bsat,N. & Helmann,J.D. Regulation of the *Bacillus subtilis* fur and perR genes by PerR: not all members of the PerR regulon are peroxide inducible. *J. Bacteriol.* **184**, 3276-3286 (2002).
63. Fujita,M. Identification of new sigma K-dependent promoters using an in vitro transcription system derived from *Bacillus subtilis*. *Gene* **237**, 45-52 (1999).

64. Fujita,M., Gonzalez-Pastor,J.E. & Losick,R. High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. *J. Bacteriol.* **187**, 1357-1368 (2005).
65. Fujita,Y., Fujita,T., Miwa,Y., Nihashi,J. & Aratani,Y. Organization and transcription of the gluconate operon, gnt, of Bacillus subtilis. *J. Biol. Chem.* **261**, 13744-13753 (1986).
66. Fujita,Y., Matsuoka,H. & Hirooka,K. Regulation of fatty acid metabolism in bacteria. *Mol Microbiol.* **66**, 829-839 (2007).
67. Fujita,Y. & Miwa,Y. Catabolite repression of the Bacillus subtilis gnt operon mediated by the CcpA protein. *J. Bacteriol.* **176**, 511-513 (1994).
68. Gama-Castro,S. *et al.* RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* **36**, D120-D124 (2008).
69. Gardan,R., Rapoport,G. & Debarbouille,M. Role of the transcriptional activator RocR in the arginine-degradation pathway of Bacillus subtilis. *Mol Microbiol.* **24**, 825-837 (1997).
70. Goelzer,A. *et al.* Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of Bacillus subtilis. *BMC Syst. Biol.* **2**, 20 (2008).
71. Gonzalez-Pastor,J.E., Hobbs,E.C. & Losick,R. Cannibalism by sporulating bacteria. *Science* **301**, 510-513 (2003).
72. Goranov,A.I., Kuester-Schoeck,E., Wang,J.D. & Grossman,A.D. Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in Bacillus subtilis. *J. Bacteriol.* **188**, 5595-5605 (2006).
73. Gorke,B. & Stulke,J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613-624 (2008).
74. Grkovic,S., Brown,M.H. & Skurray,R.A. Regulation of bacterial drug export systems. *Microbiol. Mol Biol. Rev.* **66**, 671-701, table (2002).
75. Grossman,A.D. Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis. *Annu. Rev. Genet.* **29**, 477-508 (1995).
76. Gruber,T.M. & Gross,C.A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441-466 (2003).
77. Gu,X. An evolutionary model for the origin of modularity in a complex gene network. *J. Exp. Zool. B Mol Dev. Evol.* **312**, 75-82 (2009).

78. Guillot,C. & Moran,C.P., Jr. Essential internal promoter in the spoIIIA locus of *Bacillus subtilis*. *J. Bacteriol.* **189**, 7181-7189 (2007).
79. Gustafsson,M.C., Palmer,C.N., Wolf,C.R. & von Wachenfeldt,C. Fatty-acid-displaced transcriptional repressor, a conserved regulator of cytochrome P450 102 transcription in *Bacillus* species. *Arch. Microbiol.* **176**, 459-464 (2001).
80. Gustafsson,M.C. & von Wachenfeldt,C. A novel diffusible substance can overcome the apparent AbrB repression of the *Bacillus subtilis* fatR promoter. *FEMS Microbiol. Lett.* **199**, 197-202 (2001).
81. Hamoen,L.W. *et al.* The *Bacillus subtilis* transition state regulator AbrB binds to the -35 promoter region of comK. *FEMS Microbiol. Lett.* **218**, 299-304 (2003).
82. Hamoen,L.W., Venema,G. & Kuipers,O.P. Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* **149**, 9-17 (2003).
83. Hartwell,L.H., Hopfield,J.J., Leibler,S. & Murray,A.W. From molecular to modular cell biology. *Nature* **402**, C47-C52 (1999).
84. Heath,R.J., Su,N., Murphy,C.K. & Rock,C.O. The enoyl-[acyl-carrier-protein] reductases FabI and FabL from *Bacillus subtilis*. *J. Biol. Chem.* **275**, 40128-40133 (2000).
85. Hecker,M., Pane-Farre,J. & Volker,U. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu. Rev. Microbiol.* **61**, 215-236 (2007).
86. Hecker,M., Schumann,W. & Volker,U. Heat-shock and general stress response in *Bacillus subtilis*. *Mol Microbiol.* **19**, 417-428 (1996).
87. Henriques,A.O. & Moran,C.P., Jr. Structure, assembly, and function of the spore surface layers. *Annu. Rev. Microbiol.* **61**, 555-588 (2007).
88. Herbig,A.F. & Helmann,J.D. Roles of metal ions and hydrogen peroxide in modulating the interaction of the *Bacillus subtilis* PerR peroxide regulon repressor with operator DNA. *Mol Microbiol.* **41**, 849-859 (2001).
89. Hilbert,D.W. & Piggot,P.J. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol Biol. Rev.* **68**, 234-262 (2004).
90. Hoa,T.T., Tortosa,P., Albano,M. & Dubnau,D. Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing comK. *Mol Microbiol.* **43**, 15-26 (2002).
91. Homuth,G., Rompf,A., Schumann,W. & Jahn,D. Transcriptional control of *Bacillus subtilis* hemN and hemZ. *J. Bacteriol.* **181**, 5922-5929 (1999).

92. Horsburgh,M.J. & Moir,A. Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt. *Mol Microbiol.* **32**, 41-50 (1999).
93. Hu,X. & Wu,F.X. Mining and state-space modeling and verification of sub-networks from large-scale biomolecular networks. *BMC Bioinformatics* **8**, 324 (2007).
94. Huang,M., Oppermann-Sanio,F.B. & Steinbuchel,A. Biochemical and molecular characterization of the *Bacillus subtilis* acetoin catabolic pathway. *J. Bacteriol.* **181**, 3837-3841 (1999).
95. Huang,X., Gaballa,A., Cao,M. & Helmann,J.D. Identification of target promoters for the *Bacillus subtilis* extracytoplasmic function sigma factor, sigma W. *Mol Microbiol.* **31**, 361-371 (1999).
96. Huang,X. & Helmann,J.D. Identification of target promoters for the *Bacillus subtilis* sigma X factor using a consensus-directed search. *J. Mol Biol.* **279**, 165-173 (1998).
97. Hulett,F.M. The signal-transduction network for Pho regulation in *Bacillus subtilis*. *Mol Microbiol.* **19**, 933-939 (1996).
98. Hyyrylainen,H.L. *et al.* A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress. *Mol Microbiol.* **41**, 1159-1172 (2001).
99. Hyyrylainen,H.L., Sarvas,M. & Kontinen,V.P. Transcriptome analysis of the secretion stress response of *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **67**, 389-396 (2005).
100. Illing,N. *Bacillus subtilis* and its closest relatives: From genes to cells. *Nature* **415**, 263-264 (2002).
101. JACOB,F. & MONOD,J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol Biol.* **3**, 318-356 (1961).
102. Jourlin-Castelli,C., Mani,N., Nakano,M.M. & Sonenshein,A.L. CcpC, a novel regulator of the LysR family required for glucose repression of the *citB* gene in *Bacillus subtilis*. *J. Mol Biol.* **295**, 865-878 (2000).
103. Kallio,P.T., Fagelson,J.E., Hoch,J.A. & Strauch,M.A. The transition state regulator Hpr of *Bacillus subtilis* is a DNA-binding protein. *J. Biol. Chem.* **266**, 13411-13417 (1991).
104. Kashtan,N., Itzkovitz,S., Milo,R. & Alon,U. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics.* **20**, 1746-1758 (2004).

105. Kawachi,E., Abe,S. & Tanaka,T. Inhibition of *Bacillus subtilis* scoC expression by multicopy senS. *J. Bacteriol.* **187**, 8526-8530 (2005).
106. Kearns,D.B., Chu,F., Branda,S.S., Kolter,R. & Losick,R. A master regulator for biofilm formation by *Bacillus subtilis*. *Mol Microbiol.* **55**, 739-749 (2005).
107. Kelley,W.L. Lex marks the spot: the virulent side of SOS and a closer look at the LexA regulon. *Mol Microbiol.* **62**, 1228-1238 (2006).
108. Kim,H.J., Jourlin-Castelli,C., Kim,S.I. & Sonenshein,A.L. Regulation of the *Bacillus subtilis* ccpC gene by ccpA and ccpC. *Mol Microbiol.* **43**, 399-410 (2002).
109. Kohler,P. & Marahiel,M.A. Mutational analysis of the nucleoid-associated protein HBSu of *Bacillus subtilis*. *Mol Gen. Genet.* **260**, 487-491 (1998).
110. Kroos,L. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. *Annu. Rev. Genet.* **41**, 13-39 (2007).
111. Kroos,L., Zhang,B., Ichikawa,H. & Yu,Y.T. Control of sigma factor activity during *Bacillus subtilis* sporulation. *Mol Microbiol.* **31**, 1285-1294 (1999).
112. Kruger,E. & Hecker,M. The first gene of the *Bacillus subtilis* clpC operon, ctsR, encodes a negative regulator of its own operon and other class III heat shock genes. *J. Bacteriol.* **180**, 6681-6688 (1998).
113. Kunkel,B., Kroos,L., Poth,H., Youngman,P. & Losick,R. Temporal and spatial control of the mother-cell regulatory gene spoIIID of *Bacillus subtilis*. *Genes Dev.* **3**, 1735-1744 (1989).
114. Kunst,F., Msadek,T., Bignon,J. & Rapoport,G. The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in *Bacillus subtilis*. *Res. Microbiol.* **145**, 393-402 (1994).
115. Kunst,F. *et al.* The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249-256 (1997).
116. Larsson,J.T., Rogstam,A. & von Wachenfeldt,C. YjbH is a novel negative effector of the disulphide stress regulator, Spx, in *Bacillus subtilis*. *Mol Microbiol.* **66**, 669-684 (2007).
117. Lee,T.R., Hsu,H.P. & Shaw,G.C. Transcriptional regulation of the *Bacillus subtilis* bscR-CYP102A3 operon by the BscR repressor and differential induction of cytochrome CYP102A3 expression by oleic acid and palmitate. *J. Biochem.* **130**, 569-574 (2001).

118. Leelakriangsak,M., Kobayashi,K. & Zuber,P. Dual negative control of spx transcription initiation from the P3 promoter by repressors PerR and YodB in *Bacillus subtilis*. *J. Bacteriol.* **189**, 1736-1744 (2007).
119. Leichert,L.I., Scharf,C. & Hecker,M. Global characterization of disulfide stress in *Bacillus subtilis*. *J. Bacteriol.* **185**, 1967-1975 (2003).
120. Lopez,D., Vlamakis,H. & Kolter,R. Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiol. Rev.* **33**, 152-163 (2009).
121. Lozada-Chavez,I., Janga,S.C. & Collado-Vides,J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* **34**, 3434-3445 (2006).
122. Lu,Y., Turner,R.J. & Switzer,R.L. Roles of the three transcriptional attenuators of the *Bacillus subtilis* pyrimidine biosynthetic operon in the regulation of its expression. *J. Bacteriol.* **177**, 1315-1325 (1995).
123. Madan,B.M., Balaji,S. & Aravind,L. General trends in the evolution of prokaryotic transcriptional regulatory networks. *Genome Dyn.* **3**, 66-80 (2007).
124. Madan,B.M. & Teichmann,S.A. Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. *Trends Genet.* **19**, 75-79 (2003).
125. Madan,B.M., Teichmann,S.A. & Aravind,L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol Biol.* **358**, 614-633 (2006).
126. Marino,M., Ramos,H.C., Hoffmann,T., Glaser,P. & Jahn,D. Modulation of anaerobic energy metabolism of *Bacillus subtilis* by arfM (ywiD). *J. Bacteriol.* **183**, 6815-6821 (2001).
127. Mascher,T., Hachmann,A.B. & Helmann,J.D. Regulatory overlap and functional redundancy among *Bacillus subtilis* extracytoplasmic function sigma factors. *J. Bacteriol.* **189**, 6919-6927 (2007).
128. Matsumoto,T., Nakanishi,K., Asai,K. & Sadaie,Y. Transcriptional analysis of the ylaABCD operon of *Bacillus subtilis* encoding a sigma factor of extracytoplasmic function family. *Genes Genet. Syst.* **80**, 385-393 (2005).
129. Matsuoka,H., Hirooka,K. & Fujita,Y. Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *J. Biol. Chem.* **282**, 5180-5194 (2007).
130. McCready,P., Takagi,M. & Doi,R.H. Regulation of *Bacillus subtilis* senS by homologous regulatory regions of senS and the inducible cat gene. *Biochem. Biophys. Res. Commun.* **193**, 1110-1115 (1993).

131. Mekjian,K.R., Bryan,E.M., Beall,B.W. & Moran,C.P., Jr. Regulation of hexuronate utilization in *Bacillus subtilis*. *J. Bacteriol.* **181**, 426-433 (1999).
132. Micka,B. & Marahiel,M.A. The DNA-binding protein HBSu is essential for normal growth and development in *Bacillus subtilis*. *Biochimie* **74**, 641-650 (1992).
133. Miethke,M., Schmidt,S. & Marahiel,M.A. The major facilitator superfamily-type transporter YmfE and the multidrug-efflux activator Mta mediate bacillibactin secretion in *Bacillus subtilis*. *J. Bacteriol.* **190**, 5143-5152 (2008).
134. Milo,R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).
135. Miwa,Y., Nakata,A., Ogiwara,A., Yamamoto,M. & Fujita,Y. Evaluation and characterization of catabolite-responsive elements (cre) of *Bacillus subtilis*. *Nucleic Acids Res.* **28**, 1206-1210 (2000).
136. Molle,V. *et al.* Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.* **185**, 1911-1922 (2003).
137. Moore,C.M. & Helmann,J.D. Metal ion homeostasis in *Bacillus subtilis*. *Curr. Opin. Microbiol.* **8**, 188-195 (2005).
138. Moszer,I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Lett.* **430**, 28-36 (1998).
139. Msadek,T. When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiol.* **7**, 201-207 (1999).
140. Nakamura,K., Yahagi,S., Yamazaki,T. & Yamane,K. *Bacillus subtilis* histone-like protein, HBSu, is an integral component of a SRP-like particle that can bind the Alu domain of small cytoplasmic RNA. *J. Biol. Chem.* **274**, 13569-13576 (1999).
141. Nakano,M.M., Geng,H., Nakano,S. & Kobayashi,K. The nitric oxide-responsive regulator NsrR controls ResDE-dependent gene expression. *J. Bacteriol.* **188**, 5878-5887 (2006).
142. Nakano,M.M. & Zuber,P. Anaerobic growth of a "strict aerobe" (*Bacillus subtilis*). *Annu. Rev. Microbiol.* **52**, 165-190 (1998).
143. Nakano,S., Kuster-Schock,E., Grossman,A.D. & Zuber,P. Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A* **100**, 13603-13608 (2003).
144. Nicholson,W.L., Chooback,L. & Fajardo-Cavazos,P. Analysis of spore photoproduct lyase operon (splAB) function using targeted deletion-insertion

- mutations spanning the *Bacillus subtilis* operons ptsHI and splAB. *Mol Gen. Genet.* **255**, 587-594 (1997).
145. Noone,D., Howell,A. & Devine,K.M. Expression of ykdA, encoding a *Bacillus subtilis* homologue of HtrA, is heat shock inducible and negatively autoregulated. *J. Bacteriol.* **182**, 1592-1599 (2000).
 146. Oda,M., Kobayashi,N., Ito,A., Kurusu,Y. & Taira,K. cis-acting regulatory sequences for antitermination in the transcript of the *Bacillus subtilis* hut operon and histidine-dependent binding of HutP to the transcript containing the regulatory sequences. *Mol Microbiol.* **35**, 1244-1254 (2000).
 147. Ogura,M., Matsuzawa,A., Yoshikawa,H. & Tanaka,T. *Bacillus subtilis* SalA (YbaL) negatively regulates expression of scoC, which encodes the repressor for the alkaline exoprotease gene, aprE. *J. Bacteriol.* **186**, 3056-3064 (2004).
 148. Ogura,M., Ohshiro,Y., Hirao,S. & Tanaka,T. A new *Bacillus subtilis* gene, med, encodes a positive regulator of comK. *J. Bacteriol.* **179**, 6244-6253 (1997).
 149. Pang,A.S., Nathoo,S. & Wong,S.L. Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in *Bacillus subtilis*. *J. Bacteriol.* **173**, 46-54 (1991).
 150. Pedraza-Reyes,M., Gutierrez-Corona,F. & Nicholson,W.L. Spore photoproduct lyase operon (splAB) regulation during *Bacillus subtilis* sporulation: modulation of splB-lacZ fusion expression by P1 promoter mutations and by an in-frame deletion of splA. *Curr. Microbiol.* **34**, 133-137 (1997).
 151. Perego,M., Higgins,C.F., Pearce,S.R., Gallagher,M.P. & Hoch,J.A. The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol Microbiol.* **5**, 173-185 (1991).
 152. Petersohn,A., Antelmann,H., Gerth,U. & Hecker,M. Identification and transcriptional analysis of new members of the sigmaB regulon in *Bacillus subtilis*. *Microbiology* **145** (Pt 4), 869-880 (1999).
 153. Petersohn,A. *et al.* Identification of sigma(B)-dependent genes in *Bacillus subtilis* using a promoter consensus-directed search and oligonucleotide hybridization. *J. Bacteriol.* **181**, 5718-5724 (1999).
 154. Petersohn,A. *et al.* Global analysis of the general stress response of *Bacillus subtilis*. *J. Bacteriol.* **183**, 5617-5631 (2001).
 155. Petersohn,A., Engelmann,S., Setlow,P. & Hecker,M. The katX gene of *Bacillus subtilis* is under dual control of sigmaB and sigmaF. *Mol Gen. Genet.* **262**, 173-179 (1999).

156. Phillips,Z.E. & Strauch,M.A. Bacillus subtilis sporulation and stationary phase gene expression. *Cell Mol Life Sci.* **59**, 392-402 (2002).
157. Price,C.W. *et al.* Genome-wide analysis of the general stress response in Bacillus subtilis. *Mol Microbiol.* **41**, 757-774 (2001).
158. Ravasz,E. & Barabasi,A.L. Hierarchical organization in complex networks. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **67**, 026112 (2003).
159. Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N. & Barabasi,A.L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555 (2002).
160. Reed,J.L., Vo,T.D., Schilling,C.H. & Palsson,B.O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
161. Resendis-Antonio,O. *et al.* Modular analysis of the transcriptional regulatory network of E. coli. *Trends Genet.* **21**, 16-20 (2005).
162. Rives,A.W. & Galitski,T. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U. S. A* **100**, 1128-1133 (2003).
163. Rogstam,A., Larsson,J.T., Kjelgaard,P. & von Wachenfeldt,C. Mechanisms of adaptation to nitrosative stress in Bacillus subtilis. *J. Bacteriol.* **189**, 3063-3071 (2007).
164. Ryu,H.B., Shin,I., Yim,H.S. & Kang,S.O. YlaC is an extracytoplasmic function (ECF) sigma factor contributing to hydrogen peroxide resistance in Bacillus subtilis. *J. Microbiol.* **44**, 206-216 (2006).
165. Saldanha,A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics.* **20**, 3246-3248 (2004).
166. Saxild,H.H., Brunstedt,K., Nielsen,K.I., Jarmer,H. & Nygaard,P. Definition of the Bacillus subtilis PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with glyA, guaC, pbuG, xpt-pbuX, yqhZ-fold, and pbuO. *J. Bacteriol.* **183**, 6175-6183 (2001).
167. Schilling,C.H. *et al.* Genome-scale metabolic model of Helicobacter pylori 26695. *J. Bacteriol.* **184**, 4582-4593 (2002).
168. Schilling,C.H. & Palsson,B.O. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. *J. Theor. Biol.* **203**, 249-283 (2000).
169. Schock,F. & Dahl,M.K. Expression of the tre operon of Bacillus subtilis 168 is regulated by the repressor TreR. *J. Bacteriol.* **178**, 4576-4581 (1996).

170. Schumann,W. The Bacillus subtilis heat shock stimulon. *Cell Stress. Chaperones.* **8**, 207-217 (2003).
171. Serizawa,M. *et al.* Functional analysis of the YvrGHb two-component system of Bacillus subtilis: identification of the regulated genes by DNA microarray and northern blot analyses. *Biosci. Biotechnol. Biochem.* **69**, 2155-2169 (2005).
172. Serres,M.H., Goswami,S. & Riley,M. GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins. *Nucleic Acids Res.* **32**, D300-D302 (2004).
173. Serron,P. & Sonenshein,A.L. CodY is required for nutritional repression of Bacillus subtilis genetic competence. *J. Bacteriol.* **178**, 5910-5915 (1996).
174. Shafikhani,S.H., Mandic-Mulec,I., Strauch,M.A., Smith,I. & Leighton,T. Postexponential regulation of sin operon expression in Bacillus subtilis. *J. Bacteriol.* **184**, 564-571 (2002).
175. Shannon,P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504 (2003).
176. Shen-Orr,S.S., Milo,R., Mangan,S. & Alon,U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64-68 (2002).
177. Shin,B.S., Stein,A. & Zalkin,H. Interaction of Bacillus subtilis purine repressor with DNA. *J. Bacteriol.* **179**, 7394-7402 (1997).
178. Sierro,N., Makita,Y., de Hoon,M. & Nakai,K. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res.* **36**, D93-D96 (2008).
179. Sonenshein,A.L. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Curr. Opin. Microbiol.* **8**, 203-207 (2005).
180. Sonenshein,A.L. Control of key metabolic intersections in Bacillus subtilis. *Nat. Rev. Microbiol.* **5**, 917-927 (2007).
181. Strauch,M.A. *et al.* The transition state transcription regulator abrB of Bacillus subtilis is a DNA binding protein. *EMBO J.* **8**, 1615-1621 (1989).
182. Stulke,J. & Hillen,W. Regulation of carbon catabolism in Bacillus species. *Annu. Rev. Microbiol.* **54**, 849-880 (2000).
183. Sun,G., Birkey,S.M. & Hulett,F.M. Three two-component signal-transduction systems interact for Pho regulation in Bacillus subtilis. *Mol Microbiol.* **19**, 941-948 (1996).

184. Tanous,C. *et al.* The CymR regulator in complex with the enzyme CysK controls cysteine metabolism in *Bacillus subtilis*. *J. Biol. Chem.* **283**, 35551-35560 (2008).
185. Teichmann,S.A. & Babu,M.M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492-496 (2004).
186. Thackray,P.D. & Moir,A. SigM, an extracytoplasmic function sigma factor of *Bacillus subtilis*, is activated in response to cell wall antibiotics, ethanol, heat, acid, and superoxide stress. *J. Bacteriol.* **185**, 3491-3498 (2003).
187. Toms,A.V., Haas,A.L., Park,J.H., Begley,T.P. & Ealick,S.E. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry* **44**, 2319-2329 (2005).
188. Tortosa,P. & Le Coq,D. A ribonucleic antiterminator sequence (RAT) and a distant palindrome are both involved in sucrose induction of the *Bacillus subtilis* sacXY regulatory operon. *Microbiology* **141** (Pt 11), 2921-2927 (1995).
189. Tsukahara,K. & Ogura,M. Promoter selectivity of the *Bacillus subtilis* response regulator DegU, a positive regulator of the *fla/che* operon and *sacB*. *BMC Microbiol.* **8**, 8 (2008).
190. Turner,R.J., Bonner,E.R., Grabner,G.K. & Switzer,R.L. Purification and characterization of *Bacillus subtilis* PyrR, a bifunctional pyr mRNA-binding attenuation protein/uracil phosphoribosyltransferase. *J. Biol. Chem.* **273**, 5932-5938 (1998).
191. Turner,R.J., Lu,Y. & Switzer,R.L. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (*pyr*) gene cluster by an autogenous transcriptional attenuation mechanism. *J. Bacteriol.* **176**, 3708-3722 (1994).
192. Volker,U., Andersen,K.K., Antelmann,H., Devine,K.M. & Hecker,M. One of two *osmC* homologs in *Bacillus subtilis* is part of the sigmaB-dependent general stress regulon. *J. Bacteriol.* **180**, 4212-4218 (1998).
193. von Blohn,C., Kempf,B., Kappes,R.M. & Bremer,E. Osmostress response in *Bacillus subtilis*: characterization of a proline uptake system (OpuE) regulated by high osmolarity and the alternative transcription factor sigma B. *Mol Microbiol.* **25**, 175-187 (1997).
194. Wang,L., Grau,R., Perego,M. & Hoch,J.A. A novel histidine kinase inhibitor regulating development in *Bacillus subtilis*. *Genes Dev.* **11**, 2569-2579 (1997).
195. Wang,L.F. & Doi,R.H. Complex character of *senS*, a novel gene regulating expression of extracellular-protein genes of *Bacillus subtilis*. *J. Bacteriol.* **172**, 1939-1947 (1990).

196. WATSON,J.D. & CRICK,F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
197. Watts,D.J. & Strogatz,S.H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442 (1998).
198. Werhane,H. *et al.* The last gene of the fla/che operon in *Bacillus subtilis*, ylxL, is required for maximal sigmaD function. *J. Bacteriol.* **186**, 4025-4029 (2004).
199. Wiegeshoff,F., Beckering,C.L., Debarbouille,M. & Marahiel,M.A. Sigma L is important for cold shock adaptation of *Bacillus subtilis*. *J. Bacteriol.* **188**, 3130-3133 (2006).
200. Wray,L.V., Jr., Ferson,A.E., Rohrer,K. & Fisher,S.H. TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A* **93**, 8841-8845 (1996).
201. Yamamoto,H., Mori,M. & Sekiguchi,J. Transcription of genes near the sspE locus of the *Bacillus subtilis* genome. *Microbiology* **145** (Pt 8), 2171-2180 (1999).
202. Yamamoto,H., Serizawa,M., Thompson,J. & Sekiguchi,J. Regulation of the glv operon in *Bacillus subtilis*: YfiA (GlvR) is a positive regulator of the operon that is repressed through CcpA and cre. *J. Bacteriol.* **183**, 5110-5121 (2001).
203. Yoshida,K. *et al.* Identification of additional TnrA-regulated genes of *Bacillus subtilis* associated with a TnrA box. *Mol Microbiol.* **49**, 157-165 (2003).
204. You,C. *et al.* Spx mediates oxidative stress regulation of the methionine sulfoxide reductases operon in *Bacillus subtilis*. *BMC Microbiol.* **8**, 128 (2008).
205. Zheng,L. *et al.* Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes. *J. Mol Biol.* **226**, 1037-1050 (1992).

8. APÉNDICES

8.1. Apéndice 1

Red de Regulación Transcripcional de *Bacillus subtilis*

La red se encuentra disponible en la pestaña Ap1 del archivo de Excel Ap1_Ap2_Ap3_tesis_AManjarrez.xls, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm

8.2. Apéndice 2

Red de Factores Transcripcionales de *Bacillus subtilis*

La red se encuentra disponible en la pestaña Ap2 del archivo de Excel Ap1_Ap2_Ap3_tesis_AManjarrez.xls, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm

8.3. Apéndice 3

Propiedades Topológicas de los Genes de la Red de Regulación Transcripcional de *Bacillus subtilis*

La tabla se encuentra disponible en la pestaña Ap3 del archivo de Excel Ap1_Ap2_Ap3_tesis_AManjarrez.xls, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm

8.4. Apéndice 4

Tabla de Motivos Topológicos de Tres Elementos en la Red de Regulación Transcripcional de *Bacillus subtilis*

La tabla se encuentra disponible en el archivo de Word Ap4.doc, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm

8.5. Apéndice 5

Métricas de Similitud

Las métricas de similitud utilizadas, así como los resultados obtenidos con la aplicación de cada una, se muestran en el archivo de Word Ap5.doc, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm

8.6. Apéndice 6

La Presencia de Factores σ en el Análisis Modular de la Red de Regulación Transcripcional de *Bacillus subtilis*

Los resultados obtenidos incluyendo sA en el análisis y los obtenidos excluyendo a todos los sigmas son presentados en el archivo de Word Ap6.doc, ubicado en la página web http://www.ibt.unam.mx/biocomputo/manjarrez_casas.htm