

11281



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

INSTITUTO DE CIENCIAS BIOMÉDICAS
DIVISIÓN DE ESTUDIOS DE POSGRADO
CENTRO DE INVESTIGACIÓN SOBRE
FIJACIÓN DE NITRÓGENO

REPRESENTACIÓN DE ASPECTOS DINÁMICOS DE
LA REGULACIÓN GENÉTICA

EVALUACIÓN DE LA RED DE REGULACIÓN DE
E. coli A TRAVÉS DE DATOS DE MICROARREGLOS

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE
DOCTOR EN CIENCIAS
PRESENTA:
ROSA MARÍA GUTIÉRREZ RÍOS

ASESOR:
P. JULIO COLLADO VIDES

CUERNAVACA. MORELOS.

2008/



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

ESTA TESIS NO SALE
DE LA BIBLIOTECA

Este trabajo está dedicado a Toño, Bernardo y a mi madre

Agradecimientos

Agradezco al Dr. Julio Collado por la dirección de esta tesis. Así mismo le agradezco de manera especial a los Dres. Alejandro Garcíaarubio y Miguel Ángel Cevallos por las largas sesiones de discusión que fructificaron en este bello trabajo. Igualmente agradezco a Dr. Gabriel Moreno por las sugerencias a este trabajo y a los revisores de esta tesis por los comentarios para mejorar la versión final de esta tesis.

A Emma, Josefina y a Gabriela por todo su amor.

A mis compañeros y amigos del laboratorio por cada momento.

Agradezco a DEGP-UNAM por la beca otorgada para la realización de este trabajo. Igualmente agradezco al CONACYT y al NIH a través de los donativos 0028 y GM62205-02 respectivamente por el apoyo económico brindado en la fase final de este tesis.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Rosa Maria Gutierrez
Rios

FECHA: 5 de Enero 2004

FIRMA: RMR

Resumen

La regulación del inicio de la transcripción es un proceso que involucra diversas interacciones entre los elementos *cis* en el DNA y los elementos reguladores en *trans*. La información disponible sobre estas interacciones despertó el interés de científicos teóricos que se dieron a la tarea de desarrollar modelos generales de la regulación utilizando herramientas formales. Sin embargo, la comprobación de algunos de estos modelos se dejó a un lado durante algún tiempo, dada la carencia de datos masivos.

Los métodos de secuenciación masiva motivaron a algunos grupos a desarrollar herramientas que permitieron la cuantificación de la expresión de forma global, así como la creación de bases de datos bien organizadas. Esto despertó nuevamente en la comunidad el interés por probar y generar modelos generales de regulación de la expresión.

En este trabajo proponemos una serie de reglas generales aplicables a la predicción de los estados activos e inactivos de los factores de transcripción basados en la literatura de una gran variedad de sistemas de regulación del inicio de la transcripción en *E. coli*.

E. coli, ha servido como un organismo modelo de sistemas procarióticos, de modo que la información disponible en la literatura es vasta. Esto permitió elegir entre todos los sistemas de regulación descritos en la literatura, a aquellos que proporcionarían el marco necesario para obtener las reglas. Estas reglas desarrolladas para sistemas regulados por uno y hasta n factores, actuando sobre la misma región, fueron estructuradas de manera lógica.

La aplicación de las reglas para sistemas regulados por un factor y el uso de experimentos de expresión global, dio como resultado la predicción de efectores transcripcionales que nos permitieron inferir el estado activo o inactivo del factor transcripcional. Estas predicciones se corroboran, midiendo el nivel de consistencia entre los microarreglos y las reglas para sistemas con regiones reguladas por más de un factor transcripcional.

El resultado de esta comparación, arrojó un nivel de consistencia por arriba del 70% en la red de regulación para los cuatro experimentos probados. Esto nos permite afirmar, entre otras conclusiones, que nuestras reglas pueden ser incorporadas a la modelación de redes genéticas en *E. coli*, y organismos con mecanismos de regulación semejantes.

Abstract

Regulation of transcription initiation is a process that involve different interactions between cis acting elements in the DNA molecule and the regulatory elements in trans. The available information related with these interactions made theoretical scientists to get interested in developing general regulatory models using formal tools. Whole genome sequencing, motivated some groups to generate experimental tools for quantifying gene expression in a massive way and build up well organized data bases. These advances induced again in the community the interest for develop and prove theoretical models for studying gene expression.

The purpose of this work is to find in the transcriptional regulatory systems described in the literature for the bacteria *E. coli*, general properties that facilitate us to construct a rule based approach that help us to predict the states of transcriptional factors as active or inactive.

E. coli, has been used as a model organism. Which means that we can find in literature detailed information related with transcriptional regulatory systems. Within this information we found the necessary elements for develop our rules. The rules describing the behavior of one, or more than one, transcriptional factors regulating the same transcription unit, were organized following a logical model.

Using the rules developed for systems regulated by one transcription factor and DNA microarrays, we predicted the presence or absence of transcriptional effectors that let us to assign the active and inactive state of the transcription factors. We corroborated our predictions measuring the level of consistency between microarrays and the rules for systems regulated by more than one transcription factor.

The result of this comparison, shows a level of consistency by up to 70% when rules were applied in the network. Which indicates that these rules can be incorporated for modeling genetic networks in *E. coli* and perhaps in other organisms with similar regulatory systems.

Organización del texto

La explosión de las Ciencias Genómicas ha traído como consecuencia la generación de datos masivos. La obtención de estos datos, ha despertado nuevamente el interés por probar modelos que ayuden a explicar distintos fenómenos biológicos en forma global.

Tanto los modelos existentes como los que se desarrollen requieren de un nivel importante de comprensión del fenómeno biológico modelado. Dado que uno de los objetivos del proyecto fue encontrar reglas que permitan modelar el fenómeno denominado inicio de la transcripción en la bacteria *Escherichia coli* (*E. coli*), hemos decidido describir en este manuscrito, de manera muy precisa, algunos de los mecanismos que consideramos más representativos del fenómeno de regulación del inicio de la transcripción mediado por proteínas reguladoras. De modo que el lector encontrará en la parte introductoria (sección 1.1 a la sección 1.3), aspectos generales de la expresión genética en bacterias, así como una descripción de un promotor clásico en *E. coli*, y de los distintos factores sigma que dan lugar a parte de la respuesta celular derivada de las condiciones ambientales y de crecimiento. En la sección 1.4 revisaremos con detalle los aspectos relacionados con la regulación mediada por proteínas reguladoras, (también conocidas como factores de transcripción), para unidades de transcripción (UTs) que presentan sitios de pegado al DNA para un solo regulador. En seguida, la sección 1.5 presenta al lector una revisión relacionada con UTs con sitios de pegado para más de un factor de transcripción en la región reguladora.

La sección 1.6 de la introducción nos proporciona un acercamiento a los diferentes métodos de modelación de redes o circuitos genéticos. Los cuales, como se describe en la sección, pueden ser utilizados de acuerdo al nivel de información disponible. La generación de las reglas dio lugar a la predicción de los estados de expresión de los genes regulados por los factores de transcripción, lo cual permitió corroborar los resultados de dichas predicciones con los estados de expresión obtenidos en experimentos de expresión global (también llamados microarreglos). Por lo tanto, en la sección 1.7 incluimos los fundamentos y una descripción metodológica de la técnica de microarreglos, así como, variantes de la misma desarrolladas en distintos laboratorios. Finalmente las secciones 1.8 y 1.9 de nuestra introducción, presentan algunos de los métodos más utilizados para el análisis de los datos generados por esta metodología.

Después de la presentación de los objetivos (capítulo 2), el lector encontrará en el capítulo 3 los resultados y discusión del proyecto de investigación. En la sección 3.1 (manuscrito del artículo, se muestra la construcción de las

reglas de carácter lógico, utilizando la información experimental recabada de la literatura, relacionada con los sistemas presentados en las secciones 1.4 y 1.5 de este manuscrito. Así mismo la sección 3.1, explica el desarrollo de un método que permite obtener valores discretos de encendido y apagado a partir de los datos de microarreglos en diferentes condiciones experimentales. Como se describe en la sección 3.4, los estados discretos y las reglas de interacción para un factor de transcripción sobre la UT dieron lugar a la predicción de la presencia y ausencia de los metabolitos que favorecen o desfavorecen el pegado al DNA de los reguladores transcripcionales. En la sección 3.6 y 3.7, se discuten aspectos relacionados con la expansión del modelo a otras bacterias y eucariotes unicelulares. Finalmente, el capítulo 4 presenta las conclusiones y perspectivas de este trabajo.

Índice general

1. Introducción.	9
1.1. Generalidades de la expresión genética.	9
1.2. Estructura de la RNAP	10
1.3. Estructura de promotores en <i>E. coli</i>	12
1.4. Regulación de la expresión	15
1.5. Mecanismos de regulación	19
1.6. Circuitos genéticos.	23
1.7. Metodología de los Microarreglos.	26
1.8. Análisis del arreglo.	27
1.9. Métodos para el análisis de la expresión.	31
2. Antecedentes directos y objetivos.	33
3. Resultados y discusión	35
3.1. Manuscrito del artículo.	35
3.2. Consideraciones para el análisis de microarreglos.	36
3.3. Microarreglos y regulación	36
3.4. Aspectos adicionales	38
3.5. Estados de los reguladores	41
3.6. Otros niveles de regulación	45
3.7. Expansión a Eucariotes	46
4. Conclusiones y perspectivas.	48
A. Ligas útiles en Internet	50
B. Tablas.	51

Índice de figuras

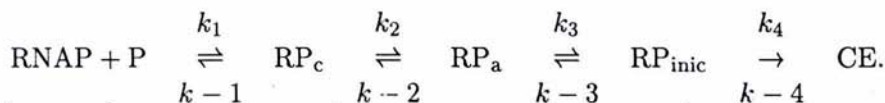
1.1. Estructura de la RNA...	11
1.2. Factores sigma en...	13
1.3. Estructura de la ...	14
1.4. Organización de la...	18
1.5. Organización de los...	21
1.6. Ejemplos de los sistemas duales...	24
1.7. Metodología para la...	28
1.8. Placa revelada de...	29
3.1. Número de efectores...	42
3.2. Porcentaje de predicción...	44

Capítulo 1

Introducción.

1.1. Generalidades de la expresión genética.

La transcripción es el proceso por medio del cual la información codificada en una de las hebras de la molécula de DNA se copia en una molécula de RNA. Este proceso lo media la RNA polimerasa dependiente de DNA (RNAP) que reconoce una serie de nucleótidos en el DNA llamada secuencia promotora y que de manera inicial forma con esta secuencia un complejo cerrado binario (RP_c). Subsecuentemente el DNA se desnaturaliza formando un complejo abierto (RP_a), que en presencia de los cuatro nucleótidos trifosfatados genera un complejo de iniciación (RP_{inic}) que puede comprometerse con una serie de inicios abortivos que generan pequeños mensajeros. Los ciclos abortivos terminan cuando finalmente la RNAP se libera del promotor, pierde el factor sigma y escapa a un proceso de elongación productiva (Record 1996). La eficiencia de cada una de estas etapas depende de la constante de equilibrio en cada uno de los pasos de la reacción siendo estos procesos reversibles hasta la etapa de formación del complejo de iniciación, como se describe en la siguiente reacción (Record 1996):



Existen elementos reguladores que pueden actuar en alguna de las distintas etapas. Se ha visto principalmente que muchos reguladores actúan en el paso que propicia la transición de complejo cerrado a complejo abierto (Hochschild and Dove 1998). Estos reguladores pueden clasificarse como positivos o de activación y negativos o de represión (Hochschild and Dove

1998). Los mecanismos de activación tienen como objetivo eliminar dos de estos pasos limitantes. Uno de ellos es el propio pegado de la polimerasa al promotor y el otro se relaciona con eludir la barrera cinética que promueva la mejor estabilidad de un complejo cerrado a diferencia de un complejo abierto (Roy et al. 1998). Mientras que los represores limitarían el acceso de la polimerasa al promotor (Roy et al. 1998). Aunque existen estos esquemas generales de regulación mediada por proteínas represoras y activadoras, encontramos modelos que describen con mayor precisión la interacción directa de los reguladores con el promotor y en algunos casos como los mecanismos de acción positivos también su acción directa sobre la RNAP .

1.2. Estructura de la RNA polimerasa y modulación a través de los factores sigma.

La holoenzima RNAP, formada de un núcleo enzimático conformado por cuatro subunidades proteicas ($\alpha_2\beta\beta'$) y un factor σ (figura 1) dirige el pegado de la RNAP a alguna clase de promotor específico (Hochschild and Dove 1998). El número total de moléculas de RNAP en una cepa de *E. coli* en crecimiento exponencial es aproximadamente de 2000 (Ishihama 1988). Juntas estas dos observaciones acentúan la importancia de la RNAP para escoger que genes va a transcribir (Ishihama 1988).

En *E. coli*, el factor sigma 70 (σ^{70}) es la subunidad más abundante que reconoce al promotor. Adicionalmente, existen seis factores sigma alternativos de acuerdo a las condiciones sobre las cuales se ha observado su actividad. Muchos de los genes expresados durante la fase de crecimiento son transcritos por la acción del factor σ^{70} , mientras que el factor σ^s es esencial en la transcripción de algunos genes durante la fase llamada estacionaria y en algunas condiciones de estrés (Hengge-Aronis 1993a; Hengge-Aronis 1993b; Ishihama 2000). El factor σ^{54} se utiliza para la expresión de genes cuando la bacteria se enfrenta a una deficiencia de nitrógeno (Merrick 1993) y otras condiciones de estrés (Shingler 1996). El factor σ^{28} se requiere para la expresión de los elementos flagelares y de los genes involucrados en los procesos de quimiotaxis (Helmann 1991). Mientras que el factor σ^{32} transcribe a los genes necesarios para la respuesta adaptativa a estrés por calor (Gross 1996); el factor σ^{24} modula la expresión de los genes involucrados en la producción de proteínas encargadas del plegamiento de otras proteínas que fueron desnaturalizadas en condiciones de estrés como por ejemplo choque térmico (Erickson and Gross 1989; Erickson et al. 1987; Mecsas et al. 1993). Finalmente, el producto del gene *fecI* -identificado originalmente como un

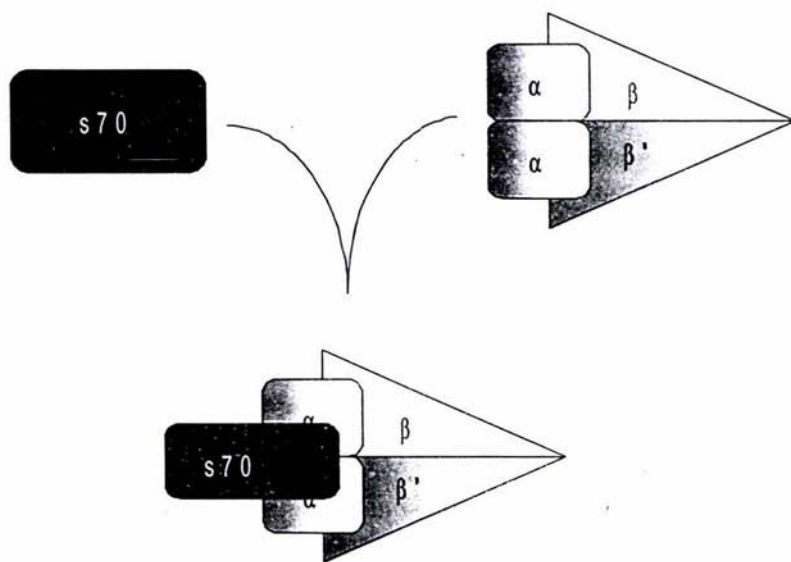


Figura 1.1: Estructura de la RNA polimerasa de *E. coli* propuesta por Ishihama.

regulador positivo- (Pressler et al. 1988), se ha reclasificado como un miembro de la familia de factores sigma que activa la transcripción de los genes *fecA* y *fecI* (Enz et al. 1995). Cada uno de estos factores participa en la transcripción de un conjunto de genes específicos.

1.3. Estructura de promotores en *E. coli*.

La RNAP reconoce a dos elementos hexaméricos en la secuencia promotora, posicionados a 7 pares de bases del inicio de la transcripción, conocidos como las cajas -10 y -35 y cuyas secuencias consenso son (TATAAT) y (TTGACA) respectivamente (Figura 3a). Cuando estas regiones son ocupadas por el factor σ^{70} y la holoenzima, cubren una distancia de 77 pares de bases (pb) (Dombroski et al. 1992). Las regiones -10 y -35 , están separadas por una distancia variable que oscila entre 15 y 21 pb llamada espaciador. En algunos promotores, existe una estructura rica en AT (conocida como elemento UP) que descansa aproximadamente a 20 residuos río arriba de la región -35 y que reconoce a la subunidad α de la RNAP (Ross et al. 1993). La fuerza de interacción entre el promotor y la RNAP depende directamente de variantes en el consenso, así como de las modificaciones en el espacio que las separan (Figura 3b y c), (Dombroski et al. 1992). Como puede esperarse son justo estas variaciones en las secuencias promotoras las responsables de crear las barreras cinéticas ya antes descritas. De manera más clara, existen dos factores que ayudan a entender el concepto de fuerza del promotor. El primero es un componente de equilibrio, relacionado con el grado de ocupación del promotor a partir de las concentraciones de σ -RNAP en el medio. El segundo componente es de origen cinético, y está relacionado con la velocidad de isomerización para pasar de un complejo cerrado a un complejo abierto.

Estos dos componentes se han evaluado de manera experimental por medio de ensayos de iniciación abortiva (Hawley and McClure 1983; McClure et al. 1983). La manera de facilitar la iniciación de la transcripción, está dada por las proteína activadoras, que en el caso de bacterias reconocen sitios específicos en el DNA y también a algunas regiones de la RNAP promoviendo con ello la expresión genética.

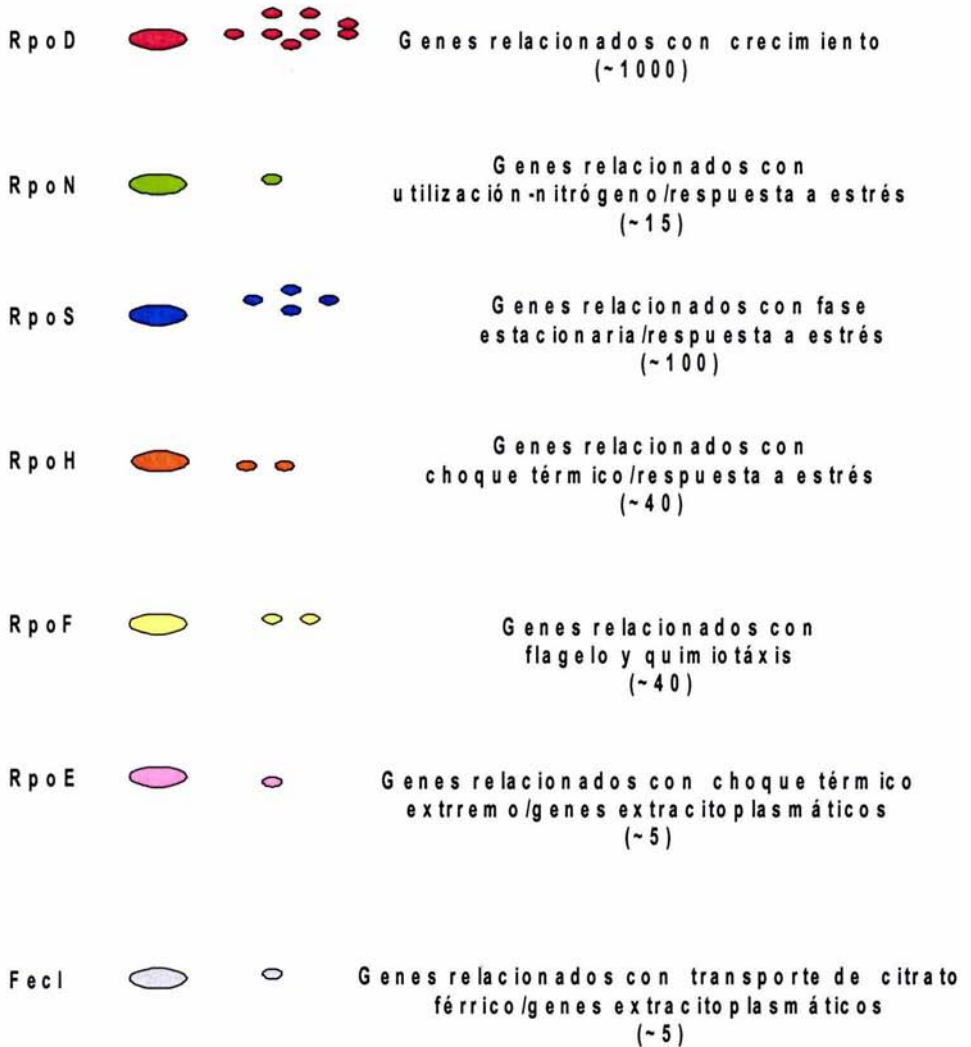


Figura 1.2: Factores sigma en *E. coli* y su concentración por célula (Ishihama, 2000).

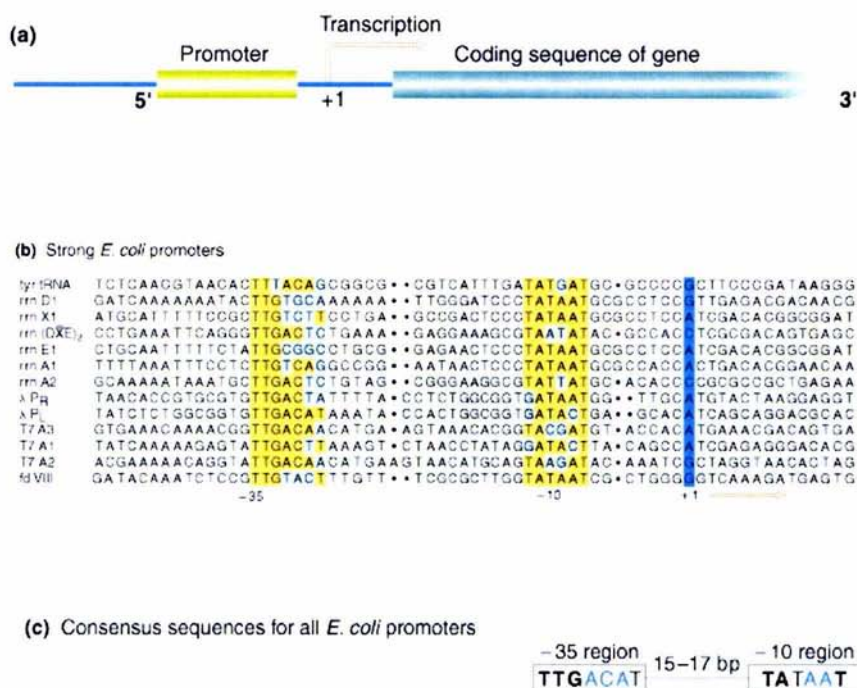


Figura 1.3: Estructura de la región promotora en *E. coli* para promotores sigma 70 y su consenso.

1.4. Regulación de la expresión genética al inicio de la transcripción.

El mecanismo básico de cómo las bacterias controlan su transcripción fue formulado por Jacob y Monod en 1961. Ellos propusieron que la expresión de los genes está regulada por las interacciones específicas que ocurren entre secuencias *cis* localizadas en el DNA con los productos de secuencias codificadas en regiones *trans*. Los elementos en *trans* son usualmente proteínas llamadas factores transcripcionales o proteínas reguladoras, las cuales se sintetizan en respuesta a estímulos ambientales como temperatura, osmolaridad, pH, o condiciones nocivas que evocan respuestas generales de protección, como agentes que causan daño al DNA, así mismo, la presencia de minerales, fuentes de energía, aceptores de electrones y metabolitos específicos constituyen una segunda clase de estímulos que modifican a la expresión genética.

Como consecuencia de las interacciones entre el organismo y el medio, las bacterias han desarrollado una gran variedad de arreglos regulatorios y una organización de los genes en grupos con función relacionada. Cuando un conjunto de genes comparte la misma unidad de transcripción, se le conoce como operón. De la misma forma, un regulón se describe como un grupo de genes regulados por la misma proteína en respuesta a un estímulo común (Maas 1964).

La expresión genética es modulada en cada etapa del inicio de la transcripción. De esta forma, los genes pueden encenderse o apagarse de distintos modos. Por ejemplo, a través de modificaciones en la RNAP, rearrreglos en la estructura del DNA que conectan a los genes con promotores particulares dada la acción de las proteínas reguladoras; pequeñas moléculas de RNA que activan o inhiben la transcripción o finalmente modulando la elongación y terminación de la transcripción en sitios específicos. La mayoría de los estímulos los detecta la célula a través de componentes solubles en el citoplasma ya sea porque el estímulo logró penetrar en la célula, o bien porque el estímulo afectó a algunos componentes celulares o del metabolismo que ahora pueden funcionar directamente como pequeñas moléculas inductoras o co-represoras de las proteínas reguladoras del inicio de la transcripción; logrando con ello, mediar una gran parte de la respuesta celular.

Activadores. Los activadores transcripcionales son proteínas que estimulan la expresión de genes. Muchos activadores transcripcionales en *E. coli* se unen específicamente al DNA en sitios localizados río arriba del promotor. Como veremos a continuación, la mayoría de los activadores examinados

hasta el momento parecen interactuar con la subunidad σ^{70} (en particular con extremo C-terminal del dominio CTD) o la subunidad σ^{70} (Hochschild and Dove 1998) de la RNAP (figura 4).

Los activadores que se unen al DNA y que contactan a la subunidad $\alpha - CTD$ funcionan, porque estabilizan la unión de la RNAP al promotor, permitiendo así la formación del complejo cerrado (Hochschild and Dove 1998). El ejemplo clásico, lo provee la proteína receptora de AMPc (CRP), que regula al promotor lac (*plac*). Aquí CRP se une a un sitio centrado en promedio a -61.5 pb del inicio de la transcripción e interactúa específicamente con la subunidad α -CTD aparentemente estabilizando su unión con el DNA entre la región -35 y el sitio de unión de CRP (Ebright and Busby 1995).

Existen activadores que contactan a la subunidad sigma 70 (Kim et al. 1995). El ejemplo más estudiado se encuentra en el bacteriófago λ y su activador cI que se une al sitio que traslapa a la región -35 del promotor λP_{RM} . La superficie expuesta, cargada negativamente de la región de activación de la proteína cI hace contacto directo con la subunidad σ^{70} de la RNAP (Kim and Hu 1997). El blanco para esta región se ha identificado como la región 4 de σ^{70} la cual incluye un dominio hélice-vuelta-hélice (HTH) que se cree contacta al hexámero -35 en la secuencia de DNA. Hay que resaltar que, a diferencia de los activadores que entran en contacto con la subunidad α -CTD, aquellos que contactan a la subunidad σ^{70} siempre se unen en la proximidad la región -35 de la secuencia promotora.

Se han reportado también activadores que hacen contacto con otras superficies de la RNAP, por ejemplo, se sabe que CRP contacta un dominio en la región amino-terminal (N-terminal) de la subunidad α de la RNAP en promotores en donde CRP traslapa la región -35 de la secuencia promotora (Busby and Ebright 1997; Niu et al. 1996). Otro ejemplo lo encontramos en la proteína DnaA que activa al promotor λP_R contactando a la subunidad β de la RNAP (Miller et al. 1997). Otro mecanismo implica el contacto de dos regiones del activador con la RNAP; en este caso la región -35 la ocupa el activador de modo que la región α -CTD queda imposibilitada a unirse al DNA en esta posición (Rhodius and Busby 1998). En el caso de promotores que contienen un elemento UP, en algunos casos, la ocupación de la región -35 provoca un desplazamiento a una nueva posición en la que la región $\alpha - CTD$ se une río-arriba de la región -35 (Rhodius and Busby 1998). Un ejemplo de esto lo encontramos en la proteína dimérica FNR (activadora de la expresión de la fumarato y de la enzima nitrato reductasa) que activa la transcripción de muchos promotores uniéndose a sitios que traslapan a la región -35 . La región de activación río- arriba del dímero de FNR

parece contactar a la región 4 de la subunidad sigma de la RNAP, mientras una segunda región de activación río-arriba del dímero hace contacto con la subunidad α -CTD de la RNAP (Bell and Busby 1994; Busby and Ebright 1997). Como podemos notar, para este mecanismo FNR hace contacto con dos subunidades de la RNAP promoviendo así la transcripción. Finalmente, se ha descrito un mecanismo en el que al parecer la conformación del DNA se altera en la región promotora. La proteína MerR, codificada en el transposón Tn501, une un dímero en la región -35 y -10 del promotor merP (Ansari et al. 1995). El ion mercurio sirve como señal que provoca un cambio conformacional en la proteína MerR, que ayuda a que el DNA se rearregle realineando de esta forma las regiones del -10 y -35 del promotor (Ansari et al. 1995). La proteína SoxR miembro de la familia de reguladores a la cual pertenece MerR, actúa de forma similar (Hidalgo and Demple 1997).

Represores. Los mecanismos por medio de los cuales se inhibe la transcripción son esencialmente de dos tipos: El primero de ellos se reduce a impedir que el complejo de iniciación llegue a un evento de elongación, lo cual puede lograrse con un represor posicionado en la región río abajo. Lo antes mencionado puede describirse más precisamente de la siguiente forma: en una primera etapa el represor puede bloquear la transición de complejo cerrado a complejo abierto. En este caso, tanto el represor como la RNAP, ocupan el DNA de manera simultánea (in vitro). Estos forman un complejo ternario de tal forma que la RNAP es incapaz de abrir al DNA en la región 10 para dar lugar a la formación del complejo abierto (Rojo 1999). Un ejemplo lo tenemos para la proteína GalR -represor de galactosa-. La región de pegado de GalR, regula a los promotores P_1 y P_2 del operón *gal*. Estos promotores se encuentran separados entre sí por 5bp. Así mismo, el represor GalR presenta dos sitios operadores O_E y O_I , localizados a -65.5 y $+53.5$ pb con respecto al inicio de la transcripción. Se ha observado que GalR en presencia de la proteína HU -proteína asociada a la cromatina- se une a ambos operadores y genera una horquilla en el DNA que inhibe la transcripción a partir de ambos promotores (Aki and Adhya 1997). Cuando la horquilla no se forma debido a la ausencia de HU o porque el operador O_E no esté presente, GalR no inhibe el pegado de la RNAP a los promotores P_1 y P_2 , aunque sí inhibe la transcripción a partir del promotor P_1 , desapareciendo la transición de complejo cerrado a abierto (Choy et al. 1997). Debe notarse que en este caso, GalR se une a la posición -65.5 lo que sugiere que probablemente ésta unión no inhibe el pegado de la RNAP (Rojo 1999).

Los represores pueden actuar también inhibiendo el abandono de la RNAP del promotor. Después de la formación del RP_{init} , la RNAP debe de abandonar al promotor. La región promotora puede ser alcanzada fácil-

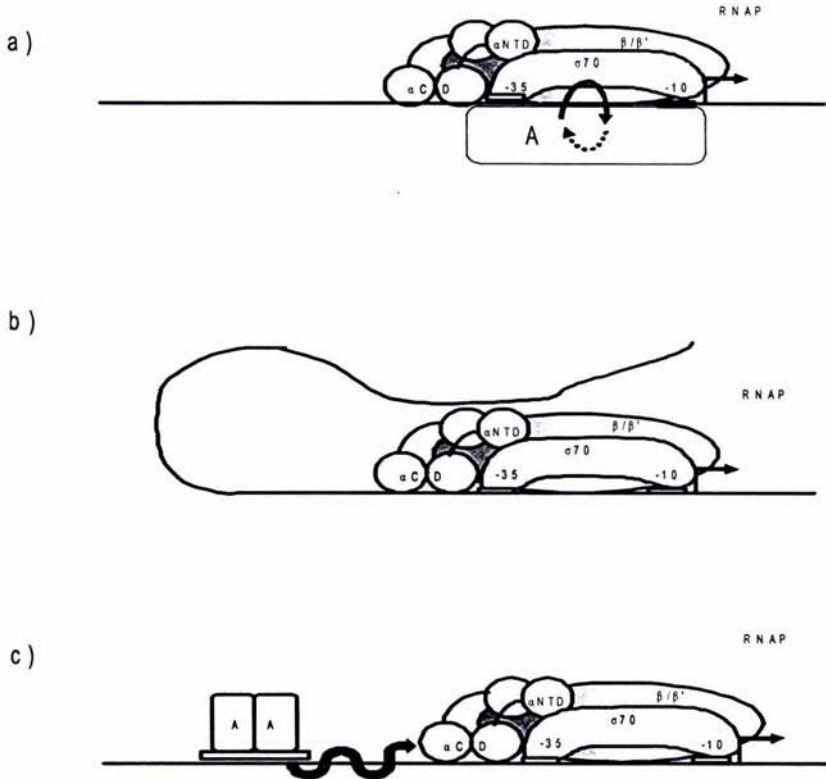


Figura 1.4: Organización de la RNAP y activadores. a) El activador A altera la yuxtaposición de las secuencias -10 y -35 en el promotor de modo que la RNAP puede unirse y alterar la transcripción. b) El activador A colocado río- arriba del promotor provoca el doblado de la secuencia facilitando la interacción de la RNAP. c) El activador A localizado en la región río-arriba provoca un cambio conformacional en el promotor induciendo así la transcripción.

mente por represores inhibiendo así la transcripción. Se ha visto que el represor LacI se une río abajo en la posición +13 o +15 con respecto al +1 de la transcripción, del promotor tardío del fago T7 (Lopez et al. 1998), impidiendo la transcripción a partir de este promotor. Cuando el operador se mueve a la posición -47, la transcripción prosigue formando un complejo estable de elongación probablemente debido a que la RNAP se ha liberado (Rojo 1999).

La represión también puede generarse cuando un represor inhibe el acceso de la RNAP al promotor. Muchos son los represores que se sabe trabajan de esta manera, por ejemplo, el represor λ CI cuando se une al operador O_{RI} del promotor viral pR (Hawley et al. 1985); el represor LexA cuando se une al promotor *uvrA* en *E. coli* (Bertrand-Burggraf et al. 1987); o la proteína LacI sobre el operador O_1 en el promotor *lac* (Matthews 1996). Finalmente, el último de los mecanismos conocidos está relacionado con proteínas de baja especificidad por sus sitios. Estos reguladores cubren vastas regiones de la secuencia nucleotídica cercanas al sitio de nucleación que eventualmente incluyen a la región del promotor (Serrano et al. 1990).

1.5. Mecanismos de regulación mediados por más de una proteína

Hemos visto hasta el momento la relación entre un activador o un represor con la RNAP. Sin embargo, existen varios ejemplos en los que más de un activador interactúa con la RNAP para de este modo inducir el fenómeno de activación. Alguno de estos modelos cuando describen la interacción de los reguladores y su acción sobre la secuencia, suponen que la función de dos o más activadores se acopla co-dependientemente a la actividad del promotor en función de dos señales fisiológicas distintas (Rhodius and Busby 1998). Cuatro modelos se han descrito que dan cuenta de este fenómeno (figura 5). En el primer modelo, los activadores hacen contactos independientes con la RNAP sin requerir interacción directa entre ambos activadores. Este modelo genera diferentes arquitecturas dentro del promotor.

La primera se genera cuando uno de los activadores se pega a la región -35 y hace contacto con la región 4 de subunidad sigma de la RNAP o la región α NTD. El segundo activador se une río arriba con respecto al -35, haciendo contacto con la región α CTD (Hochschild and Dove 1998). Por ejemplo, en *E. coli*, las proteínas CRP y FNR actúan co-dependientemente regulando al promotor *ansB*. En este caso CRP y FNR interactúan con dos de las subunidades α de la RNAP (Scott et al. 1995).

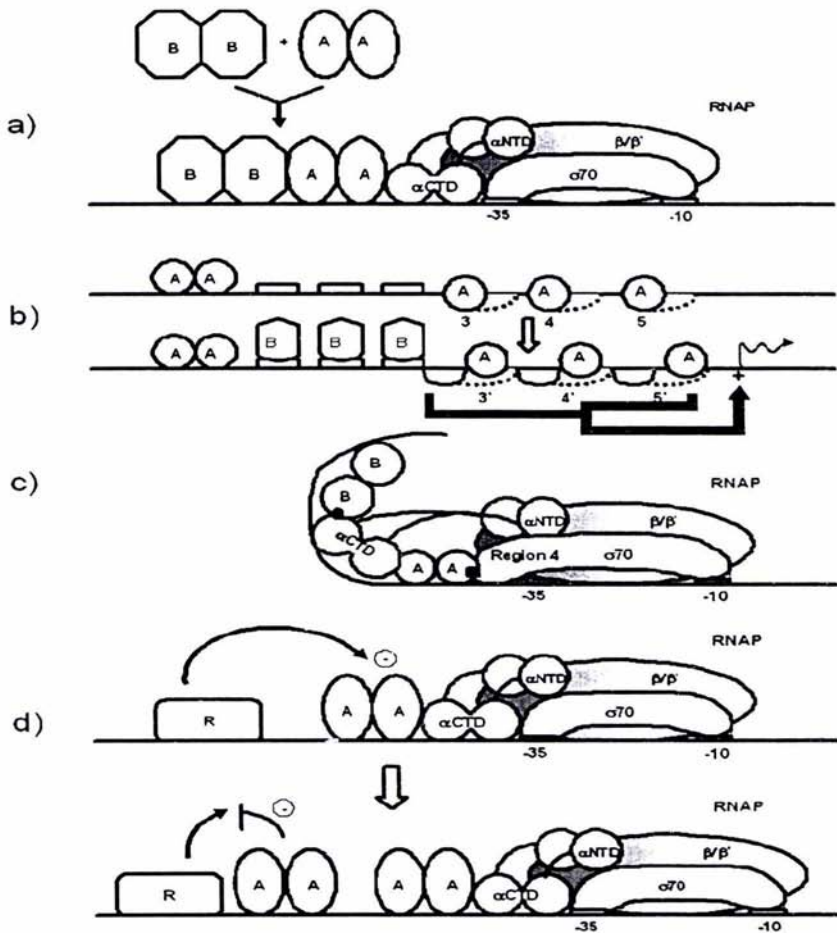
En el segundo modelo, un primer activador ayuda a que un segundo activador alcance una posición en la que es capaz de activar la transcripción. El ejemplo clásico de este modelo se encuentra en el promotor del gene *malK*, para el cual CRP se une río arriba de este promotor disparando el reposicionamiento de tres moléculas de la proteína MalT en el DNA, de tal suerte que la transcripción puede ser activada (Richet et al. 1991). En el tercer modelo, un activador dobla el DNA para facilitar la función del segundo activador que puede interactuar directamente con la RNAP. Por ejemplo, la proteína IHF (Factor de Integración al Huésped) funciona con un segundo activador, doblando el DNA en promotores del tipo σ^{54} (Goosen and van de Putte 1995).

El último modelo, describe un fenómeno nunca antes visto en bacterias hasta la descripción de este sistema. Este modelo sugiere que dos activadores co-dependientes pueden actuar cooperativamente. Así, la presencia de uno es esencial para el pegado del segundo (Wade et al. 2001). El activador MelR miembro de la familia de activadores AraC/XylR se une cooperativamente a sitios adyacentes en la secuencia de DNA del promotor *melAB* (Gallegos et al. 1997), en presencia del azúcar melibiosa y CRP con su efector AMPc (Kolb et al. 1993). Esta cooperatividad trae como consecuencia la formación de un complejo nucleoprotéico que cubre una región en el DNA mayor a 100 pb. La presencia de CRP da como resultado una mayor ocupación de MelR sobre la región -35 (Wade et al. 2001).

Al igual que los activadores, los represores también pueden actuar en conjunto con otras proteínas reguladoras. En este caso podemos encontrar combinaciones entre represores que comparten la región de regulación con activadores o con otros represores.

En el caso de más de un represor, encontramos que en la literatura no se han descrito tan claramente como en el caso de activadores modelos de interacción múltiple. Sin embargo, se darán ejemplos de los casos mejor descritos. El primero de ellos, lo encontramos en el gene *betT*, que codifica al transportador de colina. Este gene es regulado por las proteínas BetI y ArcA (Lamark et al. 1996). La región de regulación contiene dos promotores divergentes responsables de la expresión de los genes *betT* y *betIBA*. La proteína represora BetI regula la transcripción de ambos promotores en respuesta al inductor colina. Así mismo, la región reguladora contiene sitios represores para la proteína ArcA, que en condiciones anaeróbicas reprime a ambos promotores (Lamark et al. 1996). En este caso se encontró que las proteínas reprimen de manera independiente a los promotores debido a que responden a condiciones fisiológicas distintas.

La región reguladora del gene *aroL* contiene sitios de pegado al DNA



Rhodius V.A. and Busby S. 1999

Figura 1.5: Organización de los sistemas de activación. a) los activadores hacen contactos independientes con la RNAP sin requerir interacción directa entre ambos activadores. b) un primer activador ayuda a que un segundo activador alcance una posición en la que es capaz de activar la transcripción. c) uno de los activadores se pega a la región -35 mientras que el segundo activador se une río arriba con respecto al -35. d) dos activadores co-dependientes pueden actuar cooperativamente.

para el regulador TyrR (represor de tirosina) y una para la proteína TrpR (proteína represora de triptófano). Análisis mutacionales mostraron que la proteína TrpR sólo reprime en presencia de triptófano y de la proteína TyrR. A diferencia de TrpR, TyrR en presencia de tirosina reprime al gene *aroL* aún en ausencia de TrpR (Lawley and Pittard 1994). TyrR responde también a triptófano y fenilalanina pero la represión aumenta dos veces menos con respecto a la generada en presencia de tirosina o una combinación de tirosina y triptófano (Lawley and Pittard 1994). Aquí encontramos, que existe represión independiente entre TyrR y TrpR, pero la presencia de TyrR cuando TrpR ocupa su operador incrementa levemente la represión del promotor.

Por último, los reguladores ArcA y FNR responden a diferencias en las concentraciones de oxígeno o a diferentes aceptores finales de electrones en los procesos respiratorios de *E. coli*, asegurando con ello, la adaptación de la bacteria al estado energéticamente más favorable (Spiro and Guest 1991). Ambas proteínas pueden actuar como represoras o activadoras de la transcripción dependiendo del estado energético de la célula (Spiro and Guest 1991). Se ha observado que ArcA y FNR regulan a los operones *sdhCDAB*, y *cyoAB*. Cuando regulan al operón *sdhCDAB*, ArcA puede ocupar 4 sitios en el DNA. Los sitios 1 y 2 para ArcA, activan aeróbicamente la expresión (Shen and Gunsalus 1997). El sitio 3, es necesario y suficiente para reprimir la expresión anaeróbica del operón. Por su parte, el sitio 4, no parece tener un papel fundamental en la expresión del operón (Shen and Gunsalus 1997). Finalmente el control de la expresión relativa al represor FNR, ocurre de manera independiente a la represión generada por el regulador ArcA (Shen and Gunsalus 1997).

Si observamos detenidamente a cada uno de los sistemas regulados por represores, podemos darnos cuenta que en todos los mecanismos presentados, excepto en el caso del gene *aroL* con respecto al regulador TrpR, los represores actúan de manera independiente; esta observación fue importante para establecer una de las reglas de interacciones múltiples de las cuales hablaremos en el capítulo de resultados.

La región reguladora, puede ser ocupada por 2 o más proteínas en donde algunas son activadoras y las restantes que ocupan la misma región son represoras. De este tipo de regiones podemos encontrar una gran cantidad de ejemplos. Sólo daremos cuenta de tres de ellos que nos parecen interesantes representantes de los mecanismos de regulación dual.

Uno de los ejemplos más estudiados que representan a los sistemas duales lo tenemos en el promotor *lac*, que regula al operón encargado de la degradación de lactosa (Jacob and Monod 1961). La región reguladora contiene un sitio de pegado para el activador CRP y tres sitios operadores

O_1 , O_2 y O_3 , que son reconocidos por la proteína represora LacI (Lewis et al. 1996). El represor LacI a diferencia de la proteína CRP (que se pega al DNA en presencia de AMPc), se encuentra activo en ausencia de su molécula efectora alolactosa (Figura 6). Se sugiere, que en conformación adecuada ambas proteínas pueden ocupar sus sitios en el DNA, aunque, al parecer, LacI reprime aún en presencia de CRP (Vossen et al. 1996).

La región de regulación del gene *galM* representada en la figura 6, presenta sitios de pegado al DNA para los reguladores negativos GalR y HU, adicionalmente se ha descrito un sitio de pegado para el activador CRP. Se ha observado, que en presencia de uno o de ambos represores, CRP es incapaz de ocupar su sitio de pegado en el DNA, lo que genera activación sólo en ausencia de los represores (Aki and Adhya 1997).

Finalmente, el represor CytR (figura 6) regula a nueve unidades de transcripción en *E. coli* (Kristensen et al. 1996; Rasmussen et al. 1996; Valentin-Hansen et al. 1996). Aunque hay diferencias entre estas unidades, todas comparten tres características: i) todas están reguladas por CRP, ii) entre los dos sitios de CRP se forman un sitio de unión para CytR, iv) el represor de CytR solo reprime en presencia de CRP cuando aparece su inductor citidina. Estas características sugieren que el regulador CytR, funciona como un anti- activador impidiendo que CRP contacte a la subunidad α de la RNAP (Valentin-Hansen et al. 1996).

Dentro de los tres mecanismos presentados se puede notar que, de manera general, la presencia del represor impide que el activador realice su acción sobre los genes regulados. Está observación como se muestra en la parte de resultados nos ayudó a obtener la regla para mecanismos de regulación dual.

1.6. Circuitos genéticos.

Los circuitos genéticos orquestan la expresión de grupos de genes para producir un patrón apropiado de respuesta celular. El estudio de este fenómeno ha traído como consecuencia el descubrimiento molecular detallado de muchos circuitos genéticos. Sin embargo, estamos todavía lejos de entender, como estos circuitos genéticos se integran a la fisiología celular, ya que nuestro conocimiento es aún fragmentado y descriptivo.

El circuito genético de un organismo conecta al genoma con sus patrones de expresión fenotípica. El genotipo está determinado por la información codificada en la secuencia de DNA, mientras que el fenotipo está determinado por el contexto expresión-dependiente del genoma. El circuito interpreta el contexto y organiza el patrón de expresión.

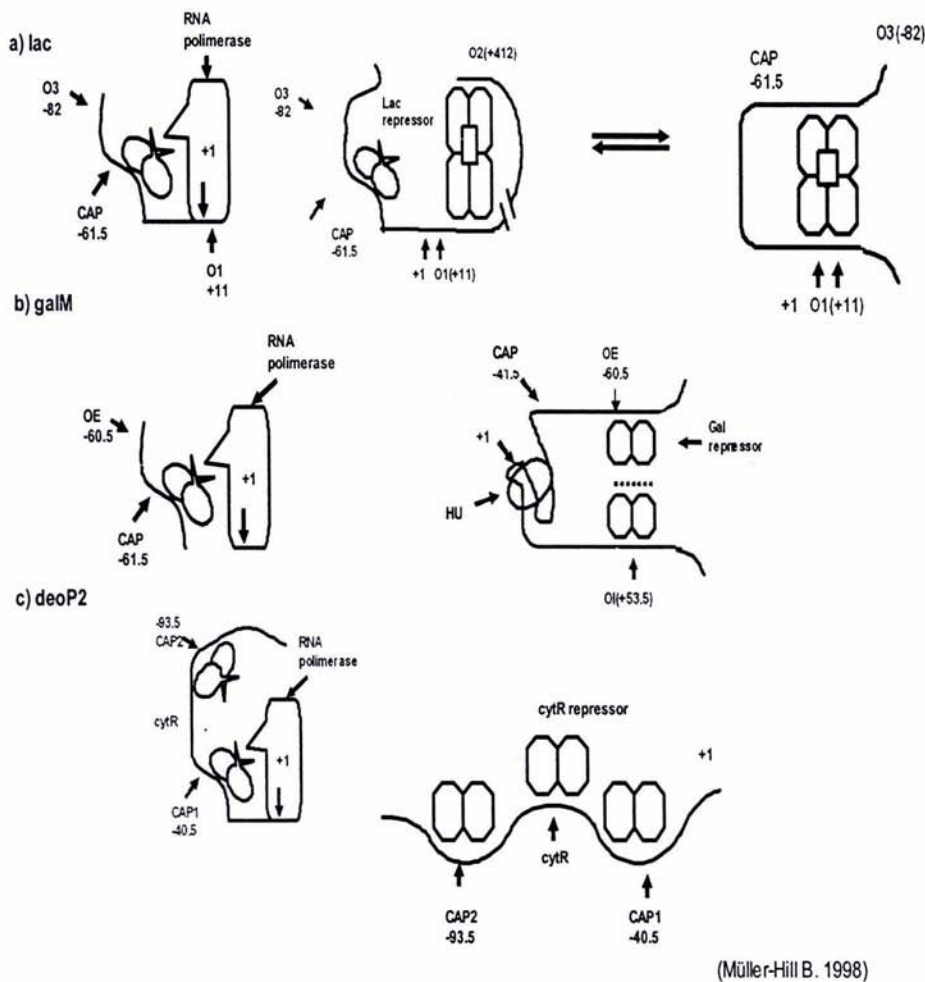


Figura 1.6: Ejemplos de los sistemas duales mejor estudiados en la literatura (*lac*, *galM* y *deoP2*). El panel a representa la organización del sistema en ausencia del represor. El panel b representa la organización de los sistemas en presencia del represor.

El estudio de los circuitos genéticos se ha abordado principalmente mediante modelos que abarcan diferentes niveles de detalle. Un acercamiento utiliza el modelo de representación *lógica o binaria*, en donde cada gene tiene dos estados: *on* u *off*, y la dinámica describe como el cambio en el estado de expresión de un grupo de genes afecta el estado de otros a través del tiempo (Kauffman 1974; Savageau 1974). Este modelo tiene la desventaja de que no permite un nivel de descripción muy detallado del sistema biológico, pero es muy útil para aquellos sistemas estudiados pobremente. Los modelos de *cinética química* o de *cambios cinéticos*, se utiliza cuando se cuenta con un nivel de descripción más detallado (Hasty et al. 2000). En estos modelos, las variables de interés son las concentraciones de proteínas individuales dentro de la célula y la dinámica que describe es la razón de cambio entre la producción y la desaparición de estas proteínas. Estos modelos consisten de un sistema de ecuaciones diferenciales ordinarias que permiten aplicar al modelador técnicas analíticas de dinámica no-lineal. La ecuación que se utiliza es determinística de modo que el ruido biológico inherente a la célula puede ser incorporado como un término que de cuenta del cambio en las concentraciones celulares de los elementos participantes. Los modelos de *cinética estocástica*, admiten el mayor nivel de descripción. Permiten hacer simulaciones del comportamiento de reacciones químicas que involucran a un número pequeño de moléculas, lo cual se ha aplicado a las reacciones de pegado entre el DNA o RNA y proteínas en los procesos de transcripción y traducción (McAdams and Arkin 1997; McAdams and Arkin 1998). El ruido propio derivado de los procesos celulares puede ser incorporado como un término estocástico, cuyas magnitudes son dependientes de la concentración (McAdams and Arkin 1999). En el caso de estos modelos al igual que en el de los modelos cinéticos, se permite dar una visión muy completa del sistema biológico estudiado, pero a diferencia de los modelos cinéticos el costo computacional es muy alto y la interpretación analítica no es inmediata.

El formalismo llamado *power law* combina elementos no lineales (producto de la ley de potencias) que tienen una estructura específica con operadores lineales para formar un grupo de ecuaciones diferenciales que sean capaces de representar convenientemente funciones diferenciales de carácter no lineal. En el caso de las redes genéticas, se ha demostrado que la red de *E. coli* sigue una ley de potencias, en donde los reguladores globales como CRP y FNR tienen un número mayor de sitios de pegado al DNA a diferencia de reguladores específicos como LacI y MalT que presentan pocos sitios de pegado al DNA a lo largo de todo el genoma (Oosawa 2002; Savageau 2000).

Todos los modelos tienen méritos e inconvenientes pero dependiendo de los objetivos del estudio y el nivel de descripción del sistema biológico que se

pretende modelar, la elección de alguno generará resultados satisfactorios.

La generación de datos masivos derivados de la secuenciación de genomas completos, experimentos de expresión global, y la generación de bases de datos bien organizadas que compilen información relacionada con el metabolismo, la regulación y las características de los elementos que participan en la modelación de los dos primeros, hacen posible con ciertas limitaciones una evaluación de los circuitos o redes genéticas. Por ejemplo, en este momento, la tecnología de microarreglos puede medir los niveles de expresión de los genes pero no nos da información relativa al conjunto completo de parámetros que determina el estado de transición de un sistema. En este sentido, aplicar métodos de ingeniería reversa a una red típica de regulación implica no tomar en cuenta todas las dependencias entre los genes y sus reguladores dada la limitada información relativa a los mecanismos de regulación para cada gene u operón, sumada a las interacciones faltantes dentro de la red misma. Adicionalmente, debemos decir que además del ruido biológico los procesos experimentales masivos están plagados de ruido experimental. Esto sin embargo, no desanima al investigador del área aunque sabe que los modelos necesitan aún de corroboración experimental.

1.7. Metodología de los Microarreglos.

Las funciones celulares recaen en los componentes proteicos del organismo. La producción de energía, la biosíntesis de macromoléculas, el mantenimiento de la arquitectura celular y la habilidad de actuar en respuesta a los estímulos intra y extracelulares son todos procesos proteína-dependiente. Cada célula contiene la información necesaria para producir todo el repertorio de proteínas que conforman a un organismo. Esta información se encuentra codificada en el genoma, y es la expresión de los genes la que determina el fenotipo en función. Por ello no es de extrañarse que la transcripción, el primer paso para convertir un gene en proteína, esté altamente regulado a través de una red de control que coordina la actividad celular.

El conocimiento del estado de expresión global de una célula puede dar información de la función de cada gene en diferentes condiciones, de estados metabólicos, de estrés, o de procesos de diferenciación. Esto a su vez puede dar lugar a la comparación entre los diversos estados de expresión de los genes en diferentes condiciones. Es por ello que en estos últimos años se ha dado gran importancia al desarrollo de la tecnología de microarreglos. Principalmente se han desarrollado dos vertientes que nos ayudan a medir el perfil de expresión de un conjunto de genes de manera global. Los llamados

DNA chips o biochips y los arreglos de oligonucleótidos o GeneChips, cuya diferencia básica radica en la forma en que se sintetizan los transcritos. Sin embargo, podemos decir que sin importar cual de las dos tecnologías se utilizara, el principio básico de esta técnica la encontramos en la propiedad de los nucleótidos llamada complementariedad, por medio de la cual un par de secuencias nucleotídicas pueden aparearse específicamente (hibridización).

En los primeros experimentos de expresión global, las copias generadas a partir de secuencias de RNA se inmovilizaban en membranas de nylon y se hibridizaban con RNA mensajero (Lander et al. 2001). Posteriormente, el gran interés en esta nueva tecnología impulsó el desarrollo de un soporte en vidrio en lugar de membranas porosas que facilitaron la miniaturización de los chips, el marcaje con agentes fluorescentes y la inmovilización de un mayor número de genes sobre la superficie. Así mismo, fue posible marcar con dos agentes fluorescentes las muestras de cDNA candidatas a hibridizar con las sondas o blancos de cDNA de cadena sencilla fijas en una posición específica a la placa de vidrio.

Los marcadores fluorescentes se agregan a la muestra de dos modos alternativos: i) el cDNA se marca con nucleótidos fluorescentes durante su síntesis o ii) el cDNA se marca incorporando a la secuencia nucleótidos biotilados que posteriormente se visualizan con streptavidina-fluorescinada. (figura 7 y 8) Posteriormente las placas se lavan, con el objetivo de remover todo el cDNA marcado que no hibridizó con ningún blanco. Acto seguido, el arreglo se ilumina con un láser y se barre con un microscopio confocal.

Finalmente las imágenes son cuantificadas con un scanner y la intensidad de brillo de cada híbrido en el arreglo representando especies de RNA- se almacena computacionalmente para su posterior análisis. Resumiendo, el sistema completo para la generación de un microarreglo, puede dividirse en 3 partes principales: i) preparación de las muestras, ii) generación del arreglo y, iii) análisis del arreglo.

1.8. Análisis del arreglo.

Como se mencionó, una vez que hemos generado el arreglo, es necesario someterlo a un proceso de análisis que nos permita obtener de manera clara que especies de RNA aumentaron, disminuyeron o no presentaron cambio alguno en nuestro experimento. Estos cambios derivados del brillo de cada gene en el arreglo, son traducidos a valores numéricos, que representan la intensidad con la que cada gene se expresó en nuestro experimento. Se ha demostrado que este valor de intensidad correlaciona con el valor absoluto de

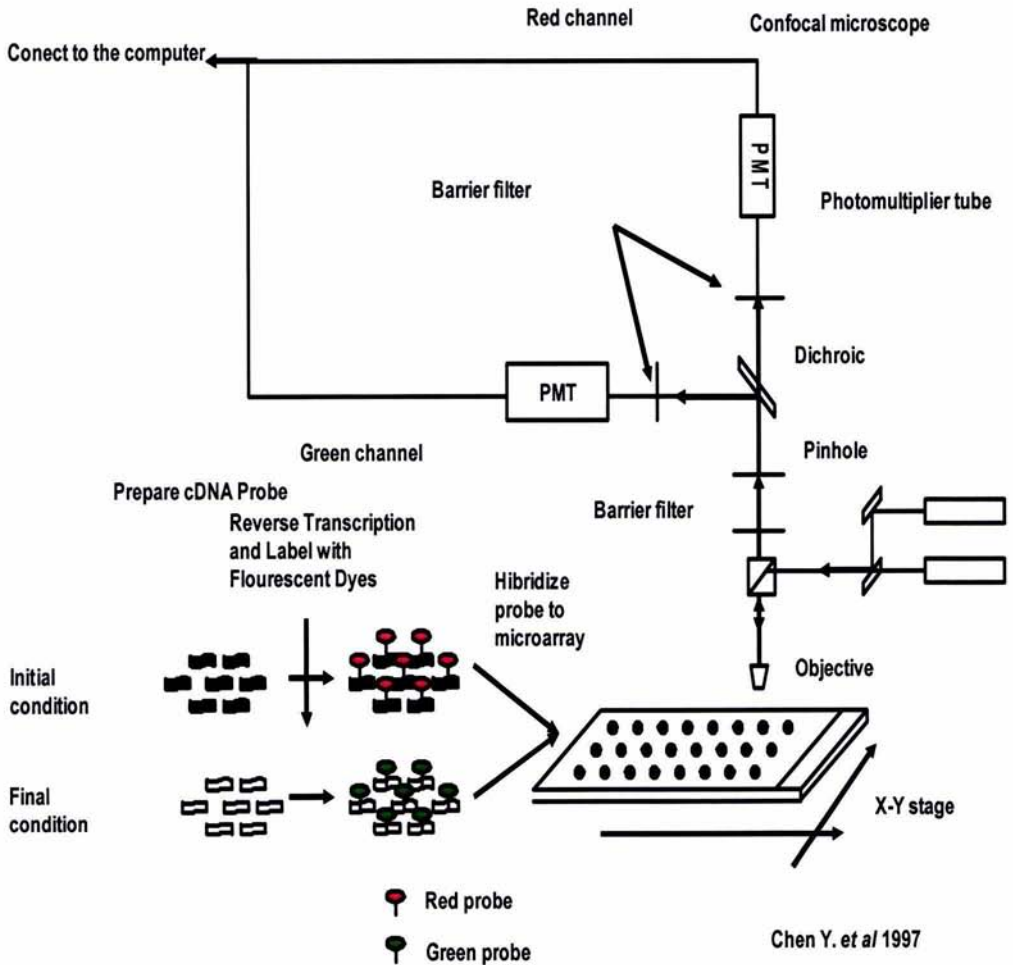


Figura 1.7: Metodología para la elaboración de un microarreglo. El diagrama muestra el proceso completo para la elaboración de un experimento de expresión diferencial global.

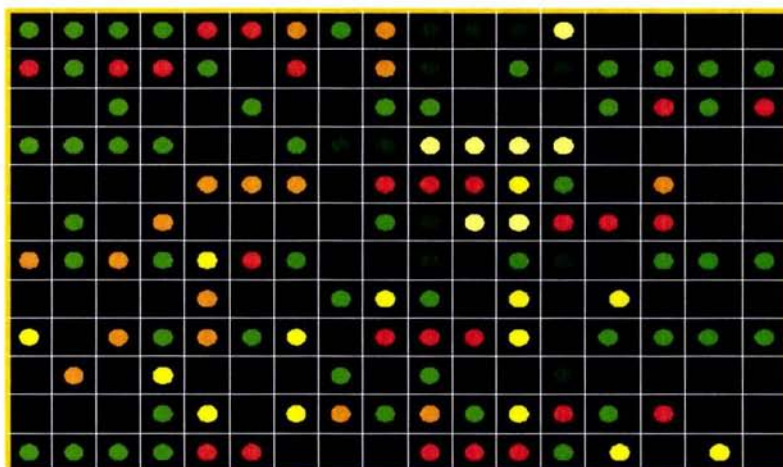


Figura 1.8: Placa revelada de un experimento de expresión diferencial o microarreglo.

RNA de la muestra original (Kohane et al. 2003). No obstante, repeticiones del mismo experimento reportan niveles de RNA no equivalentes entre ellas. Este fenómeno de irreproducibilidad en diferentes publicaciones se ha atribuido al ruido, en donde se asegura que su valor es proporcional al nivel de sensibilidad de nuestro instrumento de medición (Kohane et al. 2003). Para contender con él, la solución práctica se reduce a generar suficientes repeticiones del experimento esperando que la estadística nos ayude a obtener en promedio el resultado aproximado a las intensidades reales.

Una vez que tenemos los valores de intensidad de nuestras repeticiones, debemos transformar los datos del arreglo, de tal forma que puedan compararse unos con otros. A este proceso se le conoce como normalización de los datos. En la literatura se han descrito diversos métodos para normalizar. Sin embargo, existen tres métodos utilizados con mayor frecuencia y que logran resultados satisfactorios. El primero que describiremos se conoce como método de regresión lineal. Este método asume que el error en las mediciones es de naturaleza lineal y sistemática; considera además, que cada muestra a hibridizar con la sonda se comporta de manera similar. Bajo estos supuestos es posible calcular un factor que es utilizado para re-escalar la intensidad de cada gene en el arreglo (Kohane et al. 2003). El resultado de este tipo de normalización intenta corregir el error sistemático, aunque el componente estocástico o azaroso de nuestra medición no se corrigió. El segundo método llamado de distribuciones de probabilidades, trabaja bajo la premisa de que las distribuciones de los niveles de expresión para repeticiones del arreglo son idénticas. Bajo esta premisa, se estima una densidad de probabilidad para la razón entre la intensidad del brillo entre la condición inicial y la condición probada para el k -ésimo elemento en el arreglo. Esta distribución puede ser utilizada en un proceso iterativo que normaliza los cocientes de expresión media a uno y calcula los límites de confianza que pueden utilizarse para identificar la expresión diferencial de cada gene en el arreglo (Chen 1997). El último método presentado, conocido como de genes constantes (housekeeping) asume que ciertas sondas o blancos hibridizan de manera constante a través de todas las condiciones. Por ejemplo, cantidades exactas de RNAs mensajeros particulares se incorporan en la sonda que de antemano se sabe tiene un efecto conocido y determinístico dentro del microarreglo en cualquier condición. En este contexto, esta técnica de normalización cuenta las transformaciones de los datos, que disminuyan o eliminen la divergencia estadística en los valores de intensidad de este conjunto de genes constantes a lo largo de todos los experimentos (Kohane et al. 2003).

1.9. Métodos para el análisis de la expresión.

Después de haber normalizado las señales, el dato respectivo a cada gene se reporta como el logaritmo del cociente de la expresión. Este cociente no es más que el valor normalizado de la intensidad menos el fondo de la condición pregunta entre el valor normalizado de intensidad menos el fondo de la condición inicial. Esta intensidad representa entonces la expresión diferencial de un gene en el arreglo. La mayoría de los valores de intensidad considerados son aquellos con una diferencia de 2 relativa a su base (Quackenbush 2001).

Este corte es arbitrario y no existe una base sólida para establecerlo. A diferencia de esto existen enfoques estadísticos (Chen 1997), que establecen intervalos de confianza con una base matemática a partir de los cuales se encuentran las señales más significativas. Una vez obtenidas las señales significativas, dependiendo del objetivo del experimento todavía es posible aplicar algoritmos computacionales que nos permitan agrupar señales con valores de expresión similares. Básicamente, en cualquier algoritmo de agrupamiento el cálculo de las distancias entre dos objetos es fundamental para establecer la pertenencia de los elementos a un grupo (Quackenbush 2001). En el caso de microarreglos, el principio es el mismo, generando como resultado grupos de genes con expresión cercana. En particular el objetivo en este caso es generar un vector de expresión que representa la distancia entre cada gene. Existen varios métodos para medir estas distancias; pero típicamente podemos dividirlos en dos clases generales: métricos y semi-métricos.

Para ser clasificada como métrica, una distancia d_{ij} medida entre dos vectores i y j deben de obedecer a las siguientes reglas:

- La distancia debe ser definida como positiva o cero.
- La distancia debe ser simétrica $d_{ij} = d_{ji}$, de modo que la distancia entre i y j es la misma que entre j e i .
- Un objeto tiene distancia cero con respecto a sí mismo, $d_{ii} = 0$.
- Cuando consideramos 3 objetos i , j y k , la distancia entre i y k es igual o menor a la suma de la distancia de i y j , y de la distancia de j a k . Con frecuencia esta ecuación es llamada la desigualdad del triángulo.

La distancia más comúnmente utilizada es la Euclidiana que es una generalización del teorema de Pitágoras. En un espacio Euclidiano de tres dimensiones la distancia d_{12} , entre dos puntos (x_1, x_2, x_3) y (y_1, y_2, y_3)

está dada por la siguiente ecuación:

$$d_{12} = \sqrt{\sum_{i=1}^3 (x_i - y_i)^2},$$

en donde (x_1, x_2, x_3) son las coordenadas usuales (x, y, z) . La generalización de esto a nuestro espacio de expresión n-dimensional es como sigue:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

En donde x_i y y_i son los valores de expresión medidos, respectivamente, para los genes X y Y del experimento i y la sumatoria corre sobre los n experimentos en el análisis. En el caso de experimentos con medidas de distancia semi-métrica, se cumple las primeras tres condiciones de las medidas métricas, pero no cumplen la regla del triángulo. Ejemplos de estas aplicaciones pueden encontrarse en diversas publicaciones (Eisen et al. 1998; Ermolaeva et al. 1998). Otro tipo de algoritmos utilizados en la evaluación de microarreglos son los de agrupamiento jerárquico (Heyer et al. 1999; Weinstein et al. 1997; Wen et al. 1998). La diferencia básica entre estos algoritmos, radica en la forma en que las distancia para generar los grupos a partir del conjunto de datos son calculadas.

Capítulo 2

Antecedentes directos y objetivos.

Los patrones generales de regulación mediada por factores transcripcionales, responden a través de mecanismos diversos de respuesta los cuales pueden ser muy complejos. Estos mecanismos de respuesta se disparan ante la exposición de un organismo a un estímulo dado. Estos patrones se han modelado de manera formal mediante la utilización de métodos muy variados que dependen del nivel de descripción del sistema biológico. Lo cual deja como consecuencia, que la modelación sólo pueda ser aplicable a un número reducido de sistemas de regulación para los cuales la información es muy detallada. En esta tesis proponemos reglas generales que describen el proceso de regulación del inicio de la transcripción mediado por proteínas reguladoras. Estas reglas son el producto de un estudio detallado de los sistemas de regulación descritos en la literatura.

De modo que, el objetivo central de esta tesis fue encontrar reglas que describen la acción de las proteínas reguladoras sobre los genes regulados. Particularmente, se desarrollaron reglas de interacción simple que modelan la acción de un regulador sobre su regulón y reglas de interacción compleja que modelan la acción de más de un regulador sobre sus genes regulados, usando como modelo biológico a la bacteria *E. coli*.

La aplicación de las reglas de interacción simple nos llevó a la predicción de los estados tanto activos como inactivos de los reguladores participantes, lo cual permitió aplicarlo a la red de regulación conocida de *E. coli*. Esta red de regulación, se obtuvo de búsquedas en la literatura realizadas en nuestro laboratorio y recopiladas en la base de datos RegulonDB. (<http://www.cifn.unam.mx/RegulonDB>). Los estados activos e inactivos de

los reguladores fueron utilizados para la predicción de los estados de encendido y apagado de sistemas regulados por más de una proteína reguladora. Estos estados de encendido y apagado fueron comparados con los estado de expresión de 4 experimentos de microarreglos en diferentes condiciones de crecimiento, utilizando las reglas desarrolladas para sistemas de interacción compleja. Lo que permitió corroborar el nivel de consistencia entre nuestro modelo y datos experimentales sobre la red de *E. coli*. se probaron sobre microarreglos en 4 distintas condiciones experimentales, obteniendo como resultado final el nivel de consistencia entre nuestras reglas y los experimentos globales.

Capítulo 3

Resultados y discusión

3.1. Manuscrito del artículo.

Cita

Rosa Mara Gutierrez-Ros, David A. Rosenblueth, Jos Antonio Loza, Araceli M. Huerta, Jeremy D. Glasner, Fred R. Blattner, and Julio Collado-Vides
Regulatory Network of *Escherichia coli*: Consistency Between Literature Knowledge and Microarray Profiles *Genome Res.* 2003 **13**: 2435-2443.

Este artículo puede encontrarse en la dirección electrónica:
<http://www.genome.org/content/vol13/issue11/index.shtml>

Regulatory Network of *Escherichia coli*: Consistency Between Literature Knowledge and Microarray Profiles

Rosa María Gutiérrez-Ríos,¹ David A. Rosenblueth,³ José Antonio Loza,¹ Araceli M. Huerta,¹ Jeremy D. Glasner,² Fred R. Blattner,² and Julio Collado-Vides^{1,4}

¹Program of Computational Genomics, Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México (CIFN-UNAM), Morelos 62100, México; ²Genome Center, University of Wisconsin, Madison, Wisconsin 53706, USA;

³Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, 01000 México D. F., México

The transcriptional network of *Escherichia coli* may well be the most complete experimentally characterized network of a single cell. A rule-based approach was built to assess the degree of consistency between whole-genome microarray experiments in different experimental conditions and the accumulated knowledge in the literature compiled in RegulonDB, a data base of transcriptional regulation and operon organization in *E. coli*. We observed a high and statistical significant level of consistency, ranging from 70%–87%. When effector metabolites of regulatory proteins are not considered in the prediction of the active or inactive state of the regulators, consistency falls by up to 40%. Similarly, consistency decreases when rules for multiple regulatory interactions are altered or when “on” and “off” entries were assigned randomly. We modified the initial state of regulators and evaluated the propagation of errors in the network that do not correlate linearly with the connectivity of regulators. We interpret this deviation mainly as a result of the existence of redundant regulatory interactions. Consistency evaluation opens a new space of dialogue between theory and experiment, as the consequences of different assumptions can be evaluated and compared.

[Supplemental material is available online at www.genome.org. The data set and supplemental material are available at http://www.cifn.unam.mx/Computational_Genomics/Consistency/.]

Regulatory gene networks in the cell play an essential role in controlling the expression of specific genes according to environmental changes. The resulting patterns of gene expression vary temporally and spatially, as the outcome of a set of decisions executed by the regulatory network (Oosawa and Savageau 2002).

Extensive molecular studies in *Escherichia coli* have determined details of regulatory mechanisms and have also revealed many global aspects of the gene regulatory network (Neidhardt and Savageau 1996; Oosawa and Savageau 2002). RegulonDB is a database that contains information about transcriptional regulation and operon organization of *E. coli* derived from the careful examination of pertinent literature (Salgado et al. 2001). Although the information is still far from complete, there is currently data for about 83 (of the total possible 314) regulatory proteins (Pérez-Rueda and Collado-Vides 2000) that regulate 600 genes. All of this knowledge together, comprising around 25% of the cellular network, provides us with the largest known regulatory network of a single cell.

The understanding of how structural properties of regulatory networks determine the dynamics of their regulated genes has been the subject of studies for four decades (Kauffman 1974; Savageau 1998; Thieffry et al. 1998; Thomas and D’Ari 1990). The comparison of whole-genome expression profiles with a known network as large as that of *E. coli* provides a great opportunity to test established hypotheses, to assess more recent theoretical

models (Palsson 2001; Pilpel et al. 2001), and in general, to evaluate all aspects on our modeling of the knowledge of gene regulation, such as, for instance, general rules governing the expression by multiple regulators.

Given an initial condition specifying the state of regulatory proteins derived from an experiment, the network of regulatory interactions and conformations of regulators determines theoretical expression states of the regulated genes that can be compared with experimental data. This comparison, which generates a single number, the consistency between experiment and theory, opens a new space of dialogue between theory and experiment. The effects of different assumptions and their corresponding propagation of errors can be tested and compared. Contrary to most recent studies aimed at reconstructing the regulatory wiring from microarray experimental data alone (Eisen et al. 1998; Brown et al. 2000), we used the regulatory interactions described in RegulonDB to establish the network of causal relationships that allowed us to evaluate the congruence between literature and whole-genome expression profiles of *E. coli* in different conditions. The levels of consistency obtained with this approach range from 70%–87%, and fall by up to 40% when metabolites affecting the conformations of regulatory proteins are not considered. The degree of consistency depends both on the quality of the experiment and on the quality and quantity of knowledge about the regulatory elements governing gene expression.

RESULTS

We analyzed expression profiles of *E. coli* under four conditions, minimal medium (the common control condition), heat shock,

⁴Corresponding author.

E-MAIL collado@cifn.unam.mx; FAX 52-777-317-5581.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1387003>.

stationary phase, and anaerobic growth. Three control, independently repeated experiments, showed a correlation coefficient varying between 79% and 87%. The filtering of noise, as explained in the Methods section, left a set of 2157 (49%) genes for all analyses presented here. Similarly, of 170 known regulatory proteins, only 83 (49%) satisfied these conditions and were used in our analyses. We used a relative expression scale to transform expression ratios into on and off states, as described in the Methods section. This discretization could be done because all experimental values are relative to a unique control condition, that of minimal medium. Minimal medium shows a larger fraction of on genes, contrary to anaerobic and stationary phase conditions, consistent with overall results obtained in other laboratories (Richmond et al. 1999; Tao et al. 1999; Oh and Liao 2000).

To get an initial estimate of the expected consistency, we performed a comparison of microarray expression values with well-defined known sets of genes of each stimulus. The results of literature comparison are detailed in Tables 1–4S in the Supplemental material, available online at www.genome.org. Table 1 shows that in all conditions except stationary phase, the expression of genes corresponds to that reported in the literature for 86%–88% of the cases. The lower consistency of 69% in stationary phase is not a surprise, as the experimental settings of deprivation and stresses to induce this condition are rather variable in the literature and do not correspond exactly with those used in the microarray experiment. We considered in stationary phase, the subset of genes induced by sigmaS, which precedes the most significant changes involved in the transcription of most of the genes when the cell enters to stationary phase (Ishihama 2000).

Definitions of Complex, Simple, and Strict Regulons and Homogeneity of Their Expression

Consistency can be evaluated because of the rich structure of the transcriptional regulatory network, in which regulatory proteins work together in several combinations, governing the expression of several genes. We used the term regulon, as initially defined by Maas for the ArgR regulon in *E. coli* (Maas 1964), as the set of genes regulated by only one transcriptional regulator. We call these simple regulons. We define as complex regulons, a group of genes regulated by exactly the same set of several regulatory proteins.

Furthermore, complex and simple regulons are strict regulons if the role (activator or repressor) of each regulatory protein is the same for every gene in the regulon. For instance, the group

of genes *fruB*, *fruA*, *fruK*, and *pykF* define the negative strict simple regulon of FruR. On the other hand, the genes *edp* and *pgk* conform the complex regulon regulated by FruR and CRP. Once we have discretized individual genes, we identify those simple and complex strict regulons that are homogeneously expressed in a given condition (as explained in the Methods section). Now, we can refer to the on or off state for each strict regulon. We performed this process in all cases, excluding the regulator from the regulon set, even if subject to autoregulation, to prevent noise in the homogeneity due to possible conflicts in the expression of the regulatory gene, such as oscillations or other complex behavior. In the four conditions tested, 77% of the regulons, on average, show a homogeneous expression as shown in Table 2. This is rather high, considering the experimental noise inherent in the methodology of microarrays, messenger stability, and the amount of plausible incomplete knowledge of gene regulation in the database, such as unknown additional regulators affecting transcription initiation, and alternative levels of regulation. In addition, we are not certain that the experiments as performed were done under steady-state conditions. This could be an additional source of error that may have a consequence in our estimates of consistency. The subsequent analysis of consistency is limited to regulons with a homogeneous expression profile. We have information for 77 simple regulons in RegulonDB, with an average of 4.71 genes per regulon. Of these, only on average, 18.25 were shown homogeneous in at least one condition. There are 171 strict complex regulons in the database, of which 39.5, on average, were homogeneously expressed and are analyzed. Table 2 also shows that, on average, 57 simple regulons in each condition are regulating just one gene. Of these, only 20, on average, were used in evaluations of consistency. These left us ~78 simple and complex strict regulons to work with.

Prediction of Conformation and State of Regulators

The initial conditions of the network in terms of regulatory proteins being present or absent, and active or inactive, are derived from their discretized expression values in the microarray experiment, as well as from the expression of their corresponding strict simple regulons. Presence or absence of regulators is based on their discretized on or off values. In the absence of the regulatory protein, if it is a repressor, the genes of its simple regulon should be on assuming that a strong promoter transcribes them. In the case of a promoter regulated by an activator, we assumed a weak

Table 1. Comparison of Microarray Expression With Previous Reports in Literature

Condition		Number of genes compared	Fraction of genes consistent with literature
Heat Shock	As reported in literature	19	0.86
	Different from literature	3	0.14
	Total	22	
Stationary Phase	As reported in literature	51	0.69
	Different from literature	23	0.31
	Total	74	
Anaerobic growth	As reported in literature	66	0.88
	Different from literature	9	0.12
	Total	75	
Minimal Medium	As reported in literature	103	0.88
	Different from literature	14	0.12
	Total	117	

Literature search of genes induced or repressed in each condition were compared with expression-discretized values. Consistency ranges from 88% to 69% in the four conditions tested. The complete set of compared genes can be found in Table 1–4S of Supplemental material.

Table 2. Homogeneity of Expression Levels in Strictly Coregulated Groups Coming From Transcriptome Data

Conditions	Two or more genes per complex regulon			Simple regulons			
	on	off	Rejected	Group state Fraction of homogeneous complex regulons	on	off	Total of regulons
Heat shock	21	30	20	0.72	27	31	129
Stationary phase	31	32	11	0.85	28	26	128
Anaerobic growth	30	26	16	0.78	32	26	130
Minimal medium	36	25	18	0.77	25	34	138

Homogeneity of expression within strict regulons was evaluated on the basis of a binomial probability using the Boolean values of expression in the four conditions tested.

promoter, therefore, in the absence of the activator, the gene should be off (Neidhardt and Savageau 1996). We refer to these as the promoter rules when determining consistency. We found 32% of the 83 regulons to be off in all conditions.

When a regulator is present, a more elaborated process, using our basic knowledge of gene regulation, is required. First, we assume that when a gene gives an on value on the microarray, its protein product is present too. But presence of a regulator does not mean it is in a conformation able to exert its positive (if an activator) or its negative (if a repressor) effect on regulated genes. We assume that regulators have two conformations—which is the case, except for very few exceptions—one in which the allosteric metabolite (i.e., cAMP for CRP, allolactose for LacI, arabinose for AraC, etc.) is bound to the protein and one in which it is unbound. In some cases, alternative conformations involve phosphorylation or other processes, which, in the model, are similarly treated. Conformations are defined as active (for present regulators) or inactive (for absent regulators—depending on the presence or absence of its effector—and for absent regulators), and are deduced from the expression of the strict simple regulon for each regulator. Conformations, as defined, can in fact be assigned also for regulators whose effector metabolites have not been identified.

Briefly, we can say that a simple regulon is expressed either when its transcriptional activator is present and active, or when its transcriptional repressor is absent or inactive. We call this the simple rule, as it can be applied only to genes whose promoters are regulated by only one transcription factor, that is to say, to strict simple regulons. Table 3 summarizes the rules used to determine all possible active and inactive states of regulators determining simple regulons to be on or off. The first half of the table, when the regulatory gene is absent, is based on the promoter rules, whereas the second half, when regulatory proteins are present, uses the simple rule. We were able to predict in all of the conditions, on average, the conformation state as active or inactive for 64% of regulators from the initial set of 83. We found that 51% of the regulatory proteins have an inactive state (on average, around 25 are activators and 17 are repressors), whereas 13% are active (around five are activators and six are repressors). For 36% of the regulators (around 30), we were unable to predict whether the proteins were in an active or inactive state, either because the simple regulon did not behave homogeneously, or because there is no known strict simple regulon. With these data at hand, we are now in a position to exploit the structure of the network and evaluate consistency between simple and complex regulons.

Consistency in Expression Between Simple and Complex Regulons

Regulatory proteins govern the expression of groups of genes beyond their specific simple regulons (Neidhardt and Savageau

1996). As described previously, we identified which regulators are on and off, as well as which ones are active or inactive to exert their function. These can be considered the initial conditions of the network, in the sense that they apply only to simple regulons and depend on the experimental values for a given condition. Because these same regulators also work coordinately regulating the expression of a diversity of complex regulons, we can ask whether all of these interconnected complex regulons are expressed consistently by comparing the derived expression values with those from the experiment. An additional required ingredient to answer this question is a set of rules that define the level of expression of genes subject to the action of multiple regulators. Traditionally, logical rules have been used when modeling multiple interactions (Thomas and D'Ari 1990). We evaluated consistency on the basis of two types of rules, logical rules and rules derived from an extensive literature search of the few well-studied complex regulons in *E. coli*. Logical rules stipulate that for multiple repression, one repressor is sufficient to repress, and for multiple activators, one activator is also sufficient to activate. As a result of a literature survey summarized in Table 4, we identified a combination of general rules and specific rules for the case

Table 3. Simple Regulatory Rules

Regulator presence	Regulator function	DNA binding conformation	Effector presence (I)	Discrete expression level
Absence	Positive	P	Absent	→ Off
		P-i	Present	→ Off
		P-i	Absent	→ Off
	Negative	P	Absent	→ On
		P-i	Present	→ On
		P-i	Absent	→ On
Presence	Positive	P	Absent	→ On
		P-i	Present	→ Off
		P-i	Absent	→ Off
	Negative	P	Present	→ On
		P-i	Absent	→ On
		P-i	Present	→ Off

This table describes the rules determining gene expression when a single protein regulates genes. The rules show the regulatory protein complexes (protein and effector-metabolite), when present. We distinguish two types of conformation of regulators notated as P when they bind to their DNA-binding site with no effector, and as P-i regulators when they bind to DNA with the effector present. In the case of an absent protein, we assume that the expression state depends on the promoter strength.

Table 4. An Example of the TU's Used to Infer the Behavior of Multiple Regulated Systems

Protein function	Coregulated group	Regulated promoter(s)	Evidence	Reference	Hypothesis
Activators	CRP-MalT	malK, malE	CRP triggers MalT repositioning to an appropriate activating position	(Richet et al. 1991)	Both activators are required
	CRP-FNR	ansB	Both proteins make independent contacts with RNAPol-activating transcription	(Busby and Ebright 1997)	Both activators are required
	CRP-MeIR	meIAB	MeIR and CRP bind cooperatively at the promoter forming a complex that activates transcription in a codependent manner	(Wade et al. 2001)	Both activators are required
	IHF-NR(I), IHF-Xy1R, IHF-NarL	sigma54 promoters and narG	The first protein bends DNA in such a way that the second protein can interact directly with RNAPol. Both are required for maximal activation. The presence of IHF alone (bending protein) does not activate transcription	(Goosen and van de Putte 1995)	One specific activator is sufficient
	FNR-NarL	napF	Both are required for maximal activation. NarL (proximal site) can activate transcription a few folds compared with the control, but FNR is unable to do it.	(Darwin et al. 1998)	One specific activator is sufficient
	DnaA-FIS	nrd	Both are required for maximal activation. DnaA causes more activation than FIS.	(Jacobson and Fuchs 1998)	One specific activator is sufficient
Dual	CRP-CytR	deo	CytR functions as CRP corepressor, causing CRP to repressed transcription	(Rasmussen et al. 1996)	One repressor is sufficient
	CRP-LacI	lacZYA	When LacI is bound, CRP is unable to reach its regulatory site	(Lewis et al. 1996; Matthews, 1996)	One repressor is sufficient
	CRP-AraC	araBAD	When AraC is bound, CRP is unable to activate transcription.	(Lee et al. 1981, 1992)	One repressor is sufficient
	FNR-FIS-HNS	nirB	FNR anaerobic activation is repressed by the presence of FIS and HNS.	(Wu et al. 1998; Browning et al. 2000)	One repressor is sufficient
	NarL-NarP, -IHF	nirB	NarL-NarP complex displaces IHF repression, permitting RNAPol to initiate transcription.	(Wu et al. 1998; Browning et al. 2000)	The activator displaces the repressor activity
Repressor	CRP-GalR	gal	The GalR repressor could bind alone to its operators' sites, generating repression. When the repressor HU binds, the repression of gal promoter increases. The proteins repress independently, but maximal repression is reached when both are present	(Aki and Adhya 1997)	One repressor is sufficient
	FNR-ArcA(over two different promoters)	cyoAB, sdh	The proteins show independent repression, but maximal repression is reached when both are present.	(Cotter and Gunsalus 1992)	One specific repressor is sufficient
	ArcA-BetI	betT	The proteins repress independently, but maximal repression is reached when both are present	(Lamark et al. 1996)	One specific repressor is sufficient
	TyrR-TrpR	aroLM	TyrR represses in absence of TrpR, but TrpR is unable to repress alone	(Lawley and Pittard 1994)	One specific repressor is sufficient

of CRP. CRP-dependent promoters in the case of positive systems dominate well-studied complex regulons. For positively regulated complex regulons, we define a separate CRP-dependent rule (Lawley and Pittard 1994; Goosen and van de Putte 1995; Darwin et al. 1998; Jacobson and Fuchs 1998; Ishihama 2000; Pilpel et al. 2001), stipulating that both activators have to be active for genes to be expressed. In the case of CRP independent promoters, we propose that the presence of one of the activators in its active conformation is enough for gene expression. Complex regulons subject to negative regulation or dual systems follow the expected logical rule, stipulating that the presence of an active repressor is enough to cause the genes to be turned off. Finally, for systems regulated by two repressors, we observed that the proteins show independent repression on the promoters, but maximal repression is obtained when both repressors are active (analyzed cases can be seen in Table 4). Thus, within our Boolean model, the presence of an active repressor is enough to turn genes off. We did not find enough experiments performed in complex systems involving three or more regulatory proteins in a single promoter with sufficient detail to dissect the role of each regulator separately. We therefore extended the rules proposed here for regulons with more than two regulators, as explained in the Methods section. It is clear that these rules are in some cases excessive generalizations that could, as more knowledge accumulates in these complex mechanisms, be improved in this process of putting things together. Nothing prevents our methodology from having ultimately precise rules for every complex regulon, as long as they are well defined.

The application of these rules, together with the defined active or inactive state of regulator proteins enables us to predict the expression of complex regulons. We assessed the consistency of individual strict complex regulons, comparing our predictions with the observed expression state in the experiments. For example, Arca and FNR are two repressors predicted to be repressing five genes. These five genes appeared off under heat shock, a state that is consistent with both repressors being in an active

state. In this case, the experimental and the predicted expression state coincide, so we define the (Arca, FNR) complex regulon as a case of direct consistency.

All of the combinations of multiple interactions were considered, and different cases and types of rules associated with consistency of complex regulons are listed in the last column of Table 5. We found the highest level of consistency (87%) in heat shock, compared with that of stationary phase, in which we obtained 70% of consistent regulons. These values are similar to our previous estimates, on the basis of a small set of well-studied genes under these different conditions. We obtained a consistency of 76% and 71% for minimal medium and anaerobiosis growth, respectively. The random estimations of consistency on the basis of an average of 1000 simulated values, given the network and the rules (see the Methods section), are 45% for heat shock, 49% for stationary phase, 54% for minimal medium, and 53% in anaerobiosis. The corresponding Z scores of 4.48, 2.49, 2.54, and 2.24 clearly suggest that the observed values are significantly different from random values.

Re-evaluating Consistency: Alternative Model and Error Propagation

The evaluation of consistency involves several assumptions and operational definitions that function as abstractions of our knowledge on gene regulation. Our approach enables us to evaluate how different components of the model, and knowledge involved, affect consistency. It is interesting to compare the results we obtained with a simpler model that does not take into account the different conformations of regulatory proteins. In such a model, if an activator is on, it is activating, and if a repressor is on, it is repressing. Accordingly, an off activator is not activating and an off repressor is not repressing.

As shown in Figure 1, we found that in all conditions tested the difference between the maximum and minimum level of consistency of NC (no conformation) and NCRP (no conforma-

Table 5. Rules for Multiple Regulated Systems

Regulator1	Regulator2	Conformation of regulator1	Conformation of regulator2	Predicted discrete state	Experimental discrete state	Consistency
Any two activators						
activator	activator	activating	activating	on	on	direct
activator	activator	not activating	not activating	off	on	incongruent
activator	activator	not activating	activating	on	on	activator rule
activator	activator	not activating	activating	on	off	incongruent
activator	activator	not activating	not activating	off	off	direct
One activator is CRP						
Crp	activator	not activating	activating	off	off	crp rule
Crp	activator	not activating	activating	off	on	incongruent
Crp	activator	activating	activating	on	on	direct
Crp	activator	not activating	not activating	off	on	incongruent
Repressors						
repressor	repressor	repressing	not repressing	off	on	incongruent
repressor	repressor	repressing	repressing	off	off	direct
repressor	repressor	not repressing	repressing	off	off	repressor rule
Dual groups						
activator	repressor	activating	repressing	off	on	incongruent
activator	repressor	activating	not repressing	on	on	direct
activator	repressor	not activating	not repressing	???	on	basal level
activator	repressor	activating	repressing	off	off	repressor rule
activator	repressor	not activating	repressing	off	off	direct

Example of multiple interaction rules extracted from literature for positive, negative, and dual regulated. The first two columns describe the regulator function. Columns 3 and 4 represent the state of the regulator. Column 5 describes the predicted state. Column 6 gives the microarray discrete state, and column 7 represents the consistency between columns 5 (predicted discrete state) and 6 (experimental discrete state).

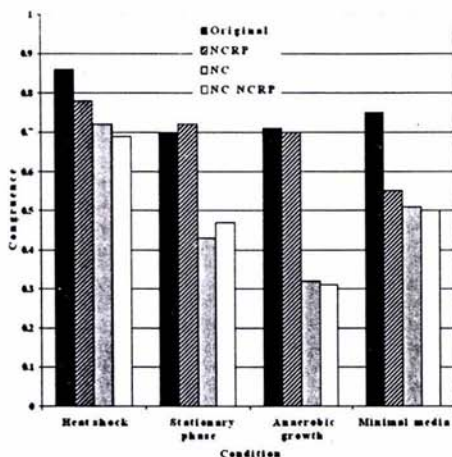


Figure 1 Summary of consistency in each condition. The figure describes the consistency between knowledge extracted from particular experiments written in terms of our multiple rules and microarray data. We show four measures, the first one with the protein state depending on its effector prediction; the second, assuming no CRP rule (NCRP), the third, assuming that all regulatory genes are active (NC) for no conformation, and the last one considering a combination between the second and the third assumption (NC NCRP). In all cases, off regulatory genes were assumed as inactive.

tion and no CRP rule) ranges from around 18% to 40%, with heat shock and anaerobiosis being the most affected conditions. This dramatic decrease illustrates how sensitive these numbers are to different definitions of rules.

We observe that simple negative autoregulated regulons tend to be less consistent in several conditions than simple negative regulons in which there is no autoregulation. This was observed in the autoregulated regulons governed by PdhR, PurR, and FliA and OxyR, contrary to simple regulons like ArcA, DnaA, and OmpR. This observation can be rationalized in terms of the oscillations of expression of homeostatic systems (Thomas and D'Ari 1990). Additionally, the identification of low-expressed regulatory proteins, problems in quantification specific to the experiments analyzed, and our partial knowledge of all regulatory interactions affecting the sets of genes analyzed, also affect consistency estimates. On the other hand, the connectivity of regulatory interactions offers a support to restrict the possible outcomes in an important way. CRP participates in 47 complex regulons, FNR in 20, in addition to their simple regulon, whereas AraC participates in only one complex regulon. Figure 2 shows the connectivity across all regulons, simple and complex, with our current knowledge in *E. coli*. It is this rich structure that was exploited here to evaluate the consistency of expression of global profile experiments in *E. coli*.

The *E. coli* network of regulatory interactions follows a power-law distribution (Oosawa and Savageau 2002) similar to what has been observed in other biological networks (Ravasz et al. 2002), that is to say, few nodes are highly connected, whereas many have a low connectivity. One would expect that highly connected regulons, because they are subject to multiple controls, are more sensitive in their expression values than regions with little dependencies/connectivity. On the other hand, local regulators would tend to define distinguishable expression ranges. In this sense, it is interesting to evaluate the impact on

overall consistency of a wrongly assigned value of initial conditions to a regulatory protein. Figure 3 shows how a wrong-state assignment of a regulator affects consistency as a function of the connectivity of the regulator. In all of the conditions tested, the most connected protein (CRP) affects consistency less (12% on average) than two regulatory proteins in the middle range of connectivity. FNR connected to 20 complex regulons, and ArcA, with 18 connections, diminish consistency on average 16% and 19%, respectively. The less connected protein FIS, affecting eight complex regulons in our set, decreases consistency only 2%. We suggest that the reason for such a phenomenon is that CRP is a global regulator, with regulatory effects that are redundant compared with those already defined by other regulators. Certainly, global regulators tend to collaborate with other regulators quite extensively (Martínez-Antonio and Collado-Vides 2003).

The results presented here suggest that our rule-based approach gives the best levels of congruence, in spite of the noise prevailing in the microarray data and the generalizations made about gene regulation. The results related to each condition are shown in Table S5 in the Supplemental material.

DISCUSSION

The final quantitative result of comparing the consistency between the microarray experiments and the predictions on the basis of the literature is a single number of consistency for each experiment. The range and reproducibility of the high consistency observed shall be strongly dependent on the quality of the experiments. Note that we used a single control as a reference, and all experiments were performed in the same laboratory.

The value of the work presented here is not only the precise degree of consistency, but also an elaborated construction of ideas and knowledge of gene regulation integrated into a rich system whose output is compared with that of the experiment. Furthermore, the virtue of the approach presented here lies in the fact that almost any piece in this construction can be substituted by a different alternative, and can be evaluated by means of the effect on consistency with experimental data. In this sense, this work opens a large window of possibilities for future research. It is also important to emphasize that this setting of ideas and construction is only applicable when the network is known. The large amount of accumulated knowledge on transcription initiation, as well as on operon organization—estimated to represent 25% of the total regulatory network in *E. coli*—is currently rather unique in this respect.

RegulonDB describes knowledge in a discrete and static way, indicating regulatory interactions and their positive and negative effects. Consistency as described in this work was assigned to a single experimental condition. The expression levels of genes in simple regulons were used to assign the active and inactive state of regulators. We made a comparison, using rules of multiple interactions and gene expression levels within complex regulons. In this way, the comparison of these two levels (state assignment and multiple interaction comparison), determine our final measurements.

Given the rules of multiple interactions, one could think that positive regulators have more redundant interactions than negative regulators. However, we did not observe a difference in consistency when comparing repressors and activators; in all of the conditions analyzed, we found, on average, that 51% of the complex regulons regulated only by activators were consistent, and 49% of those regulated only by repressors were consistent also. One would expect that the total connectivity of nonredundant interactions would exhibit a linear relationship with error propagation. If this were the case, the confidence associated with the experimental determination of the expression of genes could depend on their place within the network.

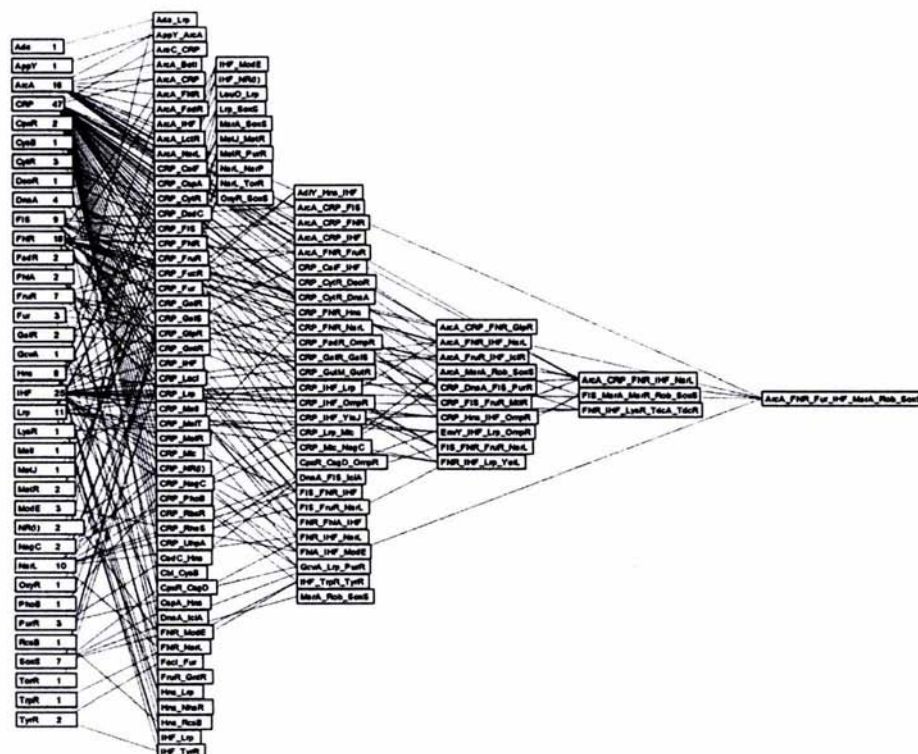


Figure 2 Connectivity across simple regulons and complex regulons. Each box represents a regulon. Simple regulons are alphabetically ordered at left. Complex regulons are ordered in increasing numbers of regulatory proteins. Connections depart only from simple regulons and arrive at complex regulons that share the same protein. In a few cases, the number of connections of a complex regulon is smaller than the number of its proteins. This is because not every regulator has a simple regulon.

Estimation of consistency put together three ingredients, the knowledge of the network and precise interactions, the setting of the initial conditions of the state of regulatory proteins as derived from the experiment, and the rules determining the outcome of multiple interactions. As genome projects, modeling of regulatory networks (Covert et al. 2001), and the associated bioinformatic tools progress, such as predictions of regulatory interactions in upstream regions in a large number of bacterial and eukaryotic genomes (Tao et al. 1999; Dombrecht et al. 2002; Halfon et al. 2002; Ravasz et al. 2002), the construction and approach implemented here could, in principle, be applicable and expandable to many other organisms beyond *E. coli*.

The ability to perform these comparisons opens questions to future research in order to precisely address and improve the adequate level of representation in the modeling of regulatory network interactions, and to integrate our understanding of the regulatory mechanisms as a function of the large set of interconnected regulatory interactions.

METHODS

Growth Conditions

For all experiments, a single colony of *E. coli* strain MG1655 was inoculated into MOPS minimal medium supplemented with

0.2% glucose (Neidhardt et al. 1974) as initial condition. Three control cultures were grown to mid-logarithmic phase in Erlenmeyer flasks at 37°C with constant aeration. The heat-shock experiment was performed by moving a flask with a mid-logarithmic phase culture from a 37°C water bath to a 50°C bath for 5 min. For the anaerobic growth experiment (or perhaps more correctly called microaerobic growth), the culture was grown in a sealed and evacuated flask to mid-logarithmic phase. For the stationary-phase experiment, a culture was grown until cells reached a stable optical density (600 nm) of 1.5.

Microarray Experiments

Total RNA was prepared from cells using QIAGEN RNeasy columns. Each control RNA sample was labeled with Cy3, and the corresponding experimental RNA sample labeled with Cy5 and cohybridized to a microarray as described previously (Richmond et al. 1999). High-density glass microarrays containing spots corresponding to nearly all *E. coli* ORFs were prepared as described previously (Richmond et al. 1999; Tao et al. 1999).

Data Treatment

Microarrays were scanned using a Packard Scanarray microarray scanner. The resulting images files were analyzed by determining the average pixel density (intensity) for each spot in the array using Quantarray detection software. A grid of individual ellipses

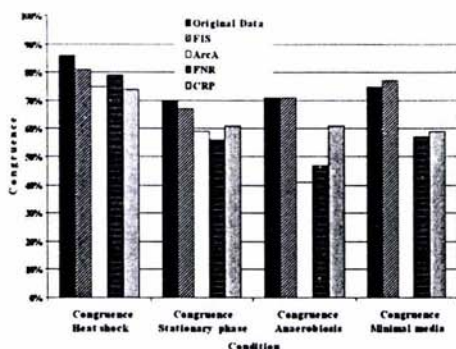


Figure 3 Consistency vs. connectivity. The graph shows the way in which an incorrect predicted state on a large connected regulatory protein influences the level of consistency. We show the cases for four proteins in the four conditions tested. Regulators are ordered from the least to the most connected. Fis coregulating with 15 proteins, ArcA with 18, FNR related 20, and CRP coregulating with 47 proteins.

corresponding to each spot was overlaid on the image to identify each spot to be quantified. Spot-specific and local background intensity values were exported to a Microsoft Access database, where background was subtracted from each intensity value. These processed signals between Cy3 and Cy5 channels were normalized, dividing each intensity value by the total intensity of all the spots on the DNA array.

Because the three conditions tested used the same control, we determined the Pearson correlation coefficient of the logarithmic percentage intensities individually for each gene as a measure of reproducibility on these repeated initial conditions. These varied from 0.79 to 0.87. All analyses performed were restricted to the set of genes whose expression values, defined as (Intensity-background)/background were ≥ 2 in each of the three experimental conditions. A total of 2157 genes satisfied this requirement.

We used the normalized intensities to generate logarithmic expression ratios, as shown in the equation

$$R_{(i)} = \ln \frac{[E_i(n)]}{C_i}$$

in which C_i is the mean normalized expression value of the three repetitions, and $E_i(n)$ is the normalized intensity value for gene i in the stress condition n . We used these values to create a relative expression scale for each gene. We included in our scale the logarithmic expression ratio of the control condition, being zero for all of the genes. Even if we know that this value reports no change in expression, its position in the scale describes the discrete state of the genes in the control condition. The relative expression error $RE(i)$ for each gene i , in each condition n , was computed as

$$RE(i) = \frac{\sigma_i / C_i}{\ln(E_i(n) / C_i)}$$

in which σ is the standard deviation estimated from the normalized expression values for gene i in the three control repetitions.

Discretization in on and off values was performed for each gene by taking the midpoint of the maximum and minimum of the relative values (in which the control is also included with a value of zero). Only those genes whose values and relative errors did not touch the midline were considered as either on if above, or off if below such midline. Note that in this way, we could discretize values for all four conditions, including the control.

Homogeneity in Strict Regulons

Once each gene had an assigned on or off value for each experiment, we grouped the on and off values in regulons strictly coregulated by the same set of proteins having the same function as an activator or a repressor. We accepted or rejected homogeneous strict regulons using a binomial probability on the basis of the total fraction of genes labeled as off. This measurement assumes independence between the results of each experiment, and independence for each gene within a strict regulon. The probability that k genes show an off state in a strict regulon of N members follows a binomial distribution. Now, we define an interval given by

$$\mu - \delta \leq x \leq \mu + \delta$$

in which $\mu = Np$ and $\delta = \sqrt{Np(1-p)}$ are the expected value and standard deviation, respectively, in the binomial distribution, and p is the global frequency of genes in an off state. For a particular strict regulon, frequencies k/N within this interval were rejected, and those outside of the interval were accepted. Values below $\mu - \delta$ are considered off, and values above $\mu + \delta$ are considered on.

Prediction of Conformation and State of Regulators

We used the discrete and homogeneous expression values of strict regulons of a single protein to infer the presence or absence of the allosteric effector of the corresponding regulatory protein; we called this process conformation assignment. In the cases in which a regulator has no simple regulon, but the regulator participates in a complex regulon, their state was deduced after defining that of the other coregulators.

The effector prediction was performed automatically using a program implemented in Prolog, which uses as inputs, (1) the set of on and off values of homogeneous strictly coregulated simple regulons, (2) the known conformations obtained from RegulonDB, and (3) the rules from Table 3 under "Regulator Presence".

Consistency Evaluation

We determined the consistency for each strict complex regulon in each condition. This is done performing a prediction of its discrete expression state given the conformations and states of the regulatory proteins involved, and the rules for multiple interactions. When this prediction matches the observed homogeneous expression of the strict regulon, it is considered as consistent, otherwise, it is inconsistent. Table 5 gives examples of the combinations for two protein-regulated systems. Taking the rules for two regulators and combining their effects, we defined rules for systems involving more than two regulators. In the case of negative and dual complex regulons, the presence of one active repressor is enough to turn genes off, regardless of the number and state of additional regulators affecting the regulon. For complex regulons coregulated only by activators and excluding CRP, the presence of an active positive regulator is sufficient to turn the genes on. For CRP-dependent strict complex regulons, all of the positive regulators have to be active to turn the regulon on; otherwise the genes are off.

We generated 1000 arrays with on and off entries selected randomly. Each of these randomized arrays of complex regulons were compared with the original arrays of complex regulons in such a way that we were able to generate the distribution of matches for each condition tested. With this information, we calculated the expected value of consistent entries and their standard deviation.

ACKNOWLEDGMENTS

R.M.G. had been supported by a Ph.D. fellowship from DGEP-UNAM. This work has been supported by grant 0028 from Conacyt-México, and by grant GM62205-02 from NIH. We thank Edgar Díaz-Peredo, Julio Freyre, Heladia Salgado, César Bonavides, Delfino García, and Víctor del Moral for their computer

support, and Alejandro Garcarrubio, Jaques vanHelden, Socorro Gama-Castro, Agustino Matinez-Antonio, and Gabriel Moreno-Hagelsieb and for fruitful discussions during the performance of this work. We acknowledge the useful comments of an anonymous referee.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Aki, T. and Adhya, S. 1997. Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J.* **16**: 3666-3674.
- Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares Jr., M., and Haussler, D. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci.* **97**: 262-267.
- Browning, D.F., Cole, J.A., and Busby, S.J. 2000. Suppression of FNR-dependent transcription activation at the *Escherichia coli* nir promoter by Fis, IHF and H-NS: Modulation of transcription initiation by a complex nucleosome-protein assembly. *Mol. Microbiol.* **37**: 1258-1269.
- Busby, S. and Ebright, R.H. 1997. Transcription activation at class II CAP-dependent promoters. *Mol. Microbiol.* **23**: 853-859.
- Cotter, P.A. and Gunsalus, R.P. 1992. Contribution of the *fnr* and *arcA* gene products in coordinate regulation of cytochrome *o* and *d* oxidase (*cyoABCDE* and *cydAB*) genes in *Escherichia coli*. *FEMS Microbiol. Lett.* **70**: 31-36.
- Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., and Palsson, B.O. 2001. Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* **26**: 179-186.
- Darwin, A.J., Ziegelhoffer, E.C., Kiley, P.J., and Stewart, V. 1998. *fnr*, *NarP*, and *NarL* regulation of *Escherichia coli* K-12 *napF* (periplasmic nitrate reductase) operon transcription in vitro. *J. Bacteriol.* **180**: 4192-4198.
- Dombrecht, B., Marchal, K., Vanderleyden, J., and Michiels, J. 2002. Prediction and overview of the RpoN-regulon in closely related species of the *Rhizobiales*. *Genome Biol.* **3**: RESEARCH0076.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**: 14863-14868.
- Goosen, N. and van de Putte, P. 1995. The regulation of transcription initiation by integration host factor. *Mol. Microbiol.* **16**: 1-7.
- Halfon, M.S., Grad, Y., Church, G.M., and Michelson, A.M. 2002. Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res.* **12**: 1019-1028.
- Ishihama, A. 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54**: 499-518.
- Jacobson, B.A. and Fuchs, J.A. 1998. Multiple *cis*-acting sites positively regulate *Escherichia coli* *nrd* expression. *Mol. Microbiol.* **28**: 1315-1322.
- Kauffman, S. 1974. The large scale structure and dynamics of gene control circuits: An ensemble approach. *J. Theor. Biol.* **44**: 167-190.
- Lamark, T., Rokenes, T.P., McDougall, J., and Strom, A.R. 1996. The complex bet promoters of *Escherichia coli*: Regulation by oxygen (*ArcA*), choline (*BetI*), and osmotic stress. *J. Bacteriol.* **178**: 1655-1662.
- Lawley, B. and Pittard, A.J. 1994. Regulation of *aroL* expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. *J. Bacteriol.* **176**: 6921-6930.
- Lee, D.H., Huo, L., and Schleif, R. 1992. Repression of the *araBAD* promoter from *araO1*. *J. Mol. Biol.* **224**: 335-341.
- Lee, N.L., Gielow, W.O., and Wallace, R.G. 1981. Mechanism of *arcA* autoregulation and the domains of two overlapping promoters, *Pc* and *PBAD*, in the L-arabinose regulatory region of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **78**: 752-756.
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**: 1247-1254.
- Maas, W.K. and Clark, A.J. 1964. Studies on the mechanism of repression of arginine biosynthesis in *E. coli*. II. Dominance of repressibility in diploids. *J. Mol. Biol.* **8**: 365-370.
- Martinez-Antonio, A. and Collado-Vides, J. 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6**: 1-8.
- Mathews, K.S. 1996. The whole lactose repressor. *Science* **271**: 1245-1246.
- Neidhardt, F.C., Bloch, P.L., and Smith, D.F. 1974. Culture medium for enterobacteria. *J. Bacteriol.* **119**: 736-747.
- Neidhardt, F.S. and Savageau, M.A. 1996. Regulation beyond the operon. In *Escherichia coli and Salmonella: Cellular and molecular biology*, 2nd ed. (eds. F. Neidhardt et al.), vol. 2, pp. 1310-1324. ASM Press, Washington, DC.
- Oh, M.K. and Liao, J.C. 2000. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol. Prog.* **16**: 278-286.
- Oosawa, C. and Savageau, M.A. 2002. Effects of alternative connectivity on behavior of randomly constructed Boolean networks. *Physica D: Nonlinear Phenomena* **170**: 143-161.
- Palsson, S. 2001. The effects of deleterious mutations in cyclically parthenogenetic organisms. *J. Theor. Biol.* **208**: 201-214.
- Perez-Rueda, E. and Collado-Vides, J. 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**: 1838-1847.
- Pilpel, Y., Sudarsanam, P., and Church, G.M. 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* **29**: 153-159.
- Rasmussen, P.B., Holst, B., and Valentin-Hansen, P. 1996. Dual-function regulators: The cAMP receptor protein and the CytR regulator can act either to repress or to activate transcription depending on the context. *Proc. Natl. Acad. Sci.* **93**: 10151-10155.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**: 1551-1555.
- Richet, E., Vidal-Ingigliardi, D., and Raibaud, O. 1991. A new mechanism for coactivation of transcription initiation: Repositioning of an activator triggered by the binding of a second activator. *Cell* **66**: 1185-1195.
- Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., and Blattner, F.R. 1999. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**: 3821-3835.
- Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Diaz-Peredo, E., Sanchez-Solano, F., Perez-Rueda, E., Bonavides-Martinez, C., and Collado-Vides, J. 2001. RegulonDB (version 3.2): Transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.* **29**: 72-74.
- Savageau, M.A. 1998. Rules for the evolution of gene circuits. *Pac. Symp. Biocomput.* **54**-65.
- Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. 1999. Functional genomics: Expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.* **181**: 6425-6440.
- Thieffry, D., Huerta, A.M., Perez-Rueda, E., and Collado-Vides, J. 1998. From specific gene regulation to genomic networks: A global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20**: 433-440.
- Thomas, R. and D'Ari, R. 1990. Biological Feedback. CRC Press, Boston, MA.
- Wade, J.T., Belyaeva, T.A., Hyde, E.I., and Busby, S.J. 2001. A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter. *EMBO J.* **20**: 7160-7167.
- Wu, H., Tyson, K.L., Cole, J.A., and Busby, S.J. 1998. Regulation of transcription initiation at the *Escherichia coli* *nir* operon promoter: A new mechanism to account for co-dependence on two transcription factors. *Mol. Microbiol.* **27**: 493-505.

Received March 28, 2003; accepted in revised form August 18, 2003.

3.2. Consideraciones para el análisis de microarreglos.

Ya hemos mencionado en la introducción algunos de los problemas intrínsecos a la metodología de microarreglos. Uno de estos es la irreproducibilidad que como se mencionó puede ser debida al ruido. En nuestra evaluación tanto el ruido como la irreproducibilidad fueron factores considerados en el análisis. Por ejemplo para el primer caso asumimos que una mediada del doble de intensidad de cada gene con respecto al fondo proporciona un corte razonable. Aunque sabíamos que genes con niveles bajos de expresión -tal y como sucede en el caso de algunos reguladores transcripcionales- quedarían fuera del conjunto de genes analizados, el resultado de este filtrado trajo como consecuencia la pérdida de 35 reguladores transcripcionales de los 118 contenidos en RegulonDB.

Otros autores utilizan para el filtrado métodos estadísticos basados en la distribución de las intensidades de los genes estableciendo límites de confianza para hacer una selección. Al igual que en nuestro análisis los métodos estadísticos no garantizan un filtrado perfecto y tampoco son capaces de proporcionarnos al conjunto de genes biológicamente significativos.

Con respecto a la reproducibilidad, ya hemos mencionado en la sección introductoria los métodos más usuales para contender con este problema. De manera particular en nuestro análisis como los experimentos siempre tuvieron las mismas condiciones iniciales la reproducibilidad se exigió como condición entre los controles de cada experimento. El beneficio de esta exigencia se reflejó en la satisfactoria discretización de los datos. Durante este proceso, notamos además que los datos que provienen del mismo laboratorio son los que proporcionan los mejores índices de correlación entre las repeticiones. La observación anterior, sugiere que los resultados obtenidos a través de la metodología de microarreglos pueden ser utilizados hasta el momento sólo como una guía que necesita de corroboración experimental más fina.

3.3. Contribución de los microarreglos al análisis de la regulación genética.

A pesar de lo discutido en la sección 3.2, los métodos de expresión global han generado mapas de expresión que revelan que genes están inducidos y reprimidos en un tiempo y condición determinada. Como debe esperarse, estos mapas están aún lejos de ayudarnos a completar la anotación de los genomas ya secuenciados de manera que nos sea posible obtener una rep-

representación completa de la red de regulación de un organismo. A pesar de ello, muchos laboratorios han incorporado a los microarreglos como una guía para la caracterización de nuevos genes, operones y regulones (McGuire, A. M. And Church, G:M. 2000, McAdams, H.H., and Shapiro,L. 1999)).

La caracterización de genes incluye secuencias tanto de genes estructurales como reguladores. En particular la caracterización de genes reguladores nos ayudan a encontrar nuevos regulones. Esto ha sido realizado para grupos de genes en diferentes condiciones experimentales en microorganismos como *Saccharomyces cerevisiae*, *E. coli* (McGuire, A. M. And Church, G:M. 2000) y *Bacillus subtilis* (McAdams, H.H., and Shapiro,L. 1999) así como en organismos mas complejos. La forma general de obtener estos regulones o grupos de genes corregulados se hace aplicando a los grupos de experimentos en distintas condiciones experimentales o la misma condición en distintos tiempos, métodos de agrupamiento como los descritos en la sección 1.9. El resultado de estos agrupamientos son probables conjuntos de genes candidatos a ser propuestos como regulones. Estos posibles regulones todavía pueden ser sometidos a una búsqueda computacional de motivos de pegado al DNA para reguladores transcripcionales, lo cual proporciona algo más de certeza en la predicción del regulón. No obstante, todavía se requiere de corroboración experimental mas fina que confirme estas predicciones.

Nuestra propuesta para el análisis de microarreglos .a través de la regulación contempla la evaluación de grupos de genes con expresión homogénea. En este caso la formación de estos grupos se realizó utilizando la información contenida en la base de datos RegulonDB tal y como se describe en la Sección de Métodos en el artículo de la inciso 3.1. Los resultados obtenido revelaron que dentro de un regulón complejo controlado por exactamente el mismo conjunto de reguladores con exactamente la misma función, los patrones de expresión en algunos caso no fueron homogéneos. Esto puede deberse principalmente a dos fuentes de error. La primera se relaciona con los errores experimentales o el ruido inherente a la propia metodología de microarreglos, y la segunda relacionada con la carencia de información en la base de datos relacionada con reguladores no descritos para algunas unidades de transcripción, promotores internos aún no colectados o caracterizados. así como terminadores de la transcripción y/o mecanismos de regulación tales como antiterminación, curvatura del DNA, modificaciones postranscripcionales, degradación de los mensajeros y la propia topología del DNA en cada condición estudiada.

3.4. Aspectos adicionales relacionados con la predicción de efectores.

Como se muestra en el manuscrito del artículo las reglas de interacción simple aunadas a los microarreglos nos permiten inferir cuál es estado de un gene cuando sabemos si los otros elementos de la regla se encuentran presentes o ausentes.

El artículo muestra que las reglas de interacción simple pueden ser reorganizadas de tal forma que cualquiera de sus elementos puede ser predicho en función de los elementos conocidos. Así que tomando esta información, se generó un programa en el lenguaje de programación lógica Prolog, que utiliza los siguientes parámetros:

1. Una tabla de genes y sus estados on u off derivada de los regulones complejos, en la condición probada.
2. Un diccionario que describe las interacciones entre los genes y sus reguladores transcripcionales.
3. Un conjunto de reglas que hacen uso de una fracción de la información contenida en la tabla de interacciones generando con ello una deducción. (Esta deducción consiste en predecir la interacción reguladora. Es decir, que conformación tenía el regulador en el tiempo que se detuvo la evolución del cultivo y que efectores dieron lugar a las conformaciones en los reguladores).

La salida del programa, es la lista de efectores y conformaciones de pegado al DNA de las proteínas reguladoras que dieron lugar al estado de los genes en la condición experimental analizada.

Como resultado de esta metodología predijimos la ausencia o presencia de 36 efectores transcripcionales para un total de 83 reguladores participantes en las cuatro diferentes condiciones como se muestra en la figura 9. La tabla 1 muestra la lista de efectores obtenida para cada regulador. En algunos casos como puede observarse en la tabla, no fue posible hacer una determinación de la ausencia o presencia del metabolito. Esto como consecuencia principalmente de dos factores:

1. El regulador presenta más de una conformación de pegado al DNA.
2. El regulador no presenta conformación alguna reportada en la literatura.

Cuadro 3.1: Predicción de efectores.

Regulator	Active State (p or pi)	Effector Molecule	heat shock	stationary phase	anaerobiosis	exponential growth (MM)
Ada	Ada-methylated agents or Ada	methylating	-	-	-	-
AraC	AraC-arabinose or AraC	arabinose	-	present	absent	-
ArcA	ArcA-phosphorilated	phosphated	present	-	absent	-
ArgR	ArgR-larginine	l-arginine	-	absent	-	present
BetI	BetI	choline.	-	absent	-	present
BirA	BirA-biotin-5-AMP	bio-5-AMP.	-	-	present	absent
CpxR	CpxR-phospholated	phosphate.	-	-	absent	-
CRP	CRP-cAMP	cAMP	absent	-	-	absent
CspA	CspA	-	-	-	present	present
CysB	CysB-thiosulphate or CysB-n-acetyl-serine	thiosulphate or n-acetylserine	present	-	absent	present
CytR	CytR-cytidine or CytR	cytidine	present	-	-	present
DeoR	DeoR-deoxyribose -5-phosphate	deoxyribose -5-phosphate	-	-	-	present present
DnaA	DnaA-ATP	ATP	absent	-	absent	-
ExuR	-	unknown	-	-	-	-
FadR	FadR	acylCoA.	absent	present	-	absent
Fecl	-	unknown	-	-	-	-
FhlA	FhlA-formate	formate.	present	-	-	absent
FIS	-	unknown	-	-	-	-
FNR	FNR-FeII	oxygen	-	absent	absent	present
FruR	FruR or fructose -1-phosphate	fructose16diph - -	- - -	present - -	- - -	- - -
FucR	-	unknown	-	-	-	-
Fur	Fur-iron or magnesium or Fur-magnesium	iron - -	- - -	absent - -	- absent -	- - -
GalR	GalR	galactose.	-	-	-	-
GalS	GalS	galactose.	-	present	-	-
GlpR	GlpR	glycerol-P.	absent	-	-	absent
GntR	GntR-gluconate	gluconate.	-	-	-	absent
Hns	-	unknown	-	-	-	-
IclA	-	unknown	-	-	-	-
IclR	IclR	phosphoenol- pyruvate.	- -	- -	- -	- -

Cuadro 3.2: continuación predicción de efectores

IHF	-	unknown	-	-	-	-
IlyY	IlyY or acetohydroxybutirate.	alfa,	absent	-	-	-
	IlyY-alfa	-	-	-	-	-
	-acetohydroxybutirate	-	-	-	-	-
KdpE	KdpE-phosphorilated	phosphate.	-	-	-	-
LctR	-	unknown	-	absent	-	-
LexA	-	ultraviolet	present	-	-	-
MalI	-	unknown	-	-	-	-
MalT	MalT-maltotriose-ATP	maltotriose-ATP.	present	present	-	-
MarA	-	unknown	-	-	-	-
MeiR	MeiR	melibiose	-	present	-	-
MetJ	MetJ-adenomethionine	adoMet	present	present	absent	-
MetR	MetR-homocysteine or MetR-methionine	homocysteine or methionine	-	present	-	-
Mic	-	unknown	-	-	-	-
MitR	-	unknown	-	-	-	-
ModE	ModE-molibdate	molibdate.	present	absent	absent	-
NadR	-	unknown	-	-	-	-
NagC	-	unknown	-	-	-	-
NarL	NarL-phosphorylated	phosphate.	present	-	-	-
NarP	NarP-phosphorylated	phosphate.	-	-	present	-
NR(I)	NR(I)-phosphorylated	phosphate.	-	-	absent	-
OmpR	OmpR-phosphorylated	phosphate.	present	-	-	-
OxyR	OxyR-oxydize or OxyR-reduced	oxygen	-	absent	-	absent
PdhR	PdhR	piruvate	-	-	absent	present
PhoB	PhoB-phosphorylated	phosphate	present	-	-	-
PurR	PurR-hipoxanthine	hipoxanthine	present	-	present	absent
RcsB	-	unknown	-	-	-	-
Rob	-	unknown	-	-	-	-
SoxR	SoxR-feSoxy or SoxR-feSred	oxydative substances	absent	present	-	-
SoxS	-	oxydative substances	-	-	-	-
SdiA	-	unknown	-	-	-	-
TyrR	TyrR or TyrR-tyrosine or TyrR-phenyalanine or TyrR-tryptophan	tyrosine or phenylalanine or tryptophan	absent	-	-	present

Un ejemplo de esto, es el regulador Ada que presenta dos conformaciones de pegado al DNA (con y sin efector), en el primer caso. Para el segundo caso la tabla 1 muestra 19 reguladores con conformación de pegado al DNA desconocida. Para algunos de ellos como en el caso de HU, IHF y FIS, es posible que los reguladores respondan a señales diferentes a efectores como los casos presentados también en la cuadro 3.1 al 2.

3.5. Predicción de estados de los reguladores transcripcionales.

La metodología anterior nos permitió predecir la conformación de las proteínas reguladoras como activas o inactivas. El manuscrito de la sección 3.1, muestra que para aquellos reguladores sin conformación conocida se tomó la expresión de su regulón como base para predecir el estado del regulador. La cuadro 3.1 muestra la lista de reguladores con sus conformaciones predichas en cada una de las condiciones analizadas. Al igual que en el caso de las predicciones de los metabolitos como puede verse en la cuadro 3.1, no nos fue posible determinar los estado de algunos reguladores. Los casos no predichos fueron los siguientes:

1. Aquellos con más de una conformación descrita.
2. Aquellos reguladores con conformación desconocida o conocida y cuyo grupo de regulones no fue homogénea .

Como se muestra en la figura 10, para todas las condiciones probadas, de un total de 83 reguladores, en promedio, sólo en 13% se encontraron activos, el 52% se encontraron inactivos y un 36% no consiguieron predicción alguna. No es de sorprendernos el bajo porcentaje de reguladores activos. Ya que es muy probable, que sólo aquellos reguladores participantes en la adaptación de la bacteria a la perturbación sometida, estén respondiendo ante el estímulo. No obstante, el uso de un mayor número de condiciones y repeticiones de cada experimento, permitirá verificar estos resultados.

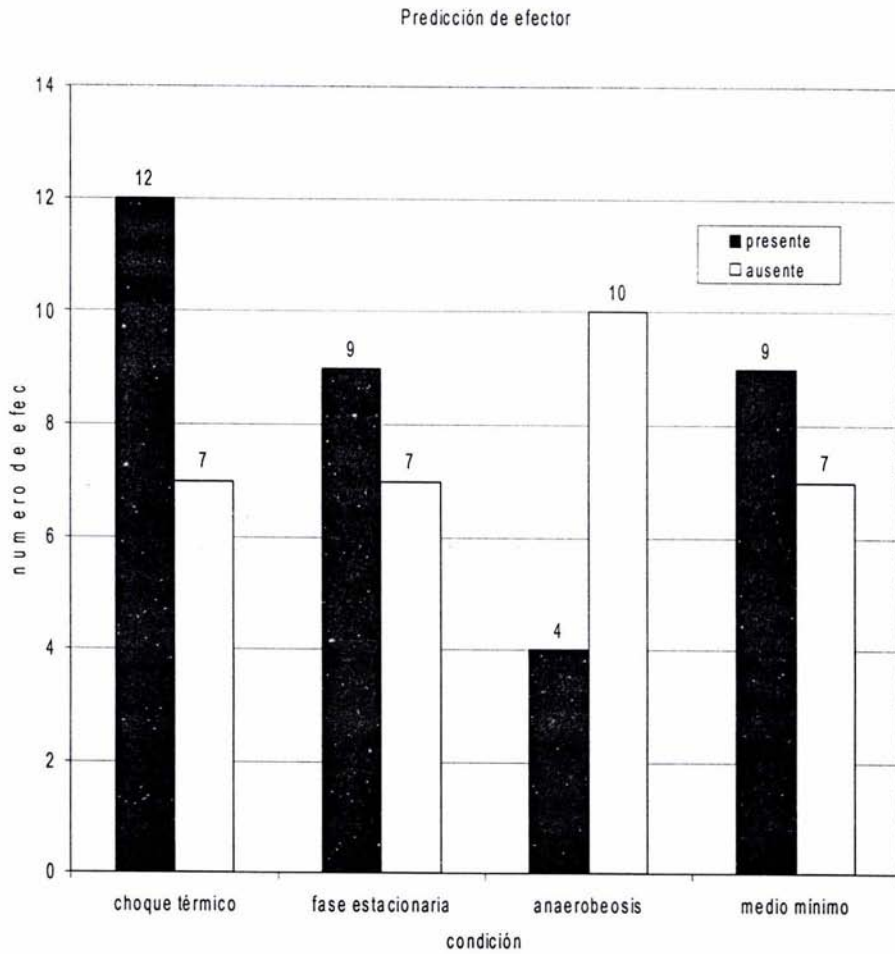


Figura 3.1: Número de efectores por condición.

Cuadro 3.3: Estados activos e inactivos de los factores de transcripción.

Proteins regulating	heat shock	stationary phase	anaerobic growth	minimal media
Ada	not activating	ambiguous	not activating	-
AlpA	not detected	-	not activating	-
AppY	not detected	activating	ambiguous	not activating
AraC	ambiguous	not repressing	not repressing	not repressing
ArcA	repressing	not repressing	not repressing	not repressing
ArgR	not repressing	not repressing	not repressing	repressing
BetI	-	not repressing	not repressing	not repressing
BirA	-	-	repressing	not repressing
CpxR	-	-	not activating	-
CRP	not repressing	not activating	activating	not activating
CspA	not activating	not repressing	repressing	not activating
CysB	activating	not activating	not activating	activating
CytR	repressing	ambiguous	not repressing	repressing
DnaA	not repressing	not repressing	not repressing	not repressing
DeoR	-	not repressing	not repressing	repressing
DsdC	not detected	not detected	not detected	not detected
EnvY	not detected	not detected	not detected	not detected
ExuR	not activating	not repressing	not repressing	-
FadR	repressing	ambiguous	not repressing	repressing
FecI	not activating	activating	not activating	activating
FhIA	activating	activating	not activating	not activating
FIS	activating	not activating	not activating	not activating
FNR	not repressing	not repressing	not activating	not repressing
FruR	not activating	not activating	not activating	not activating
FruR*	not repressing	not repressing	not repressing	not repressing
FucR	not activating	not activating	not activating	not activating
Fur	not repressing	not repressing	not repressing	not repressing
GalR	not repressing	not repressing	not repressing	ambiguous
GalS	not repressing	not repressing	not repressing	not repressing
GcvA	not detected	ambiguous	ambiguous	ambiguous

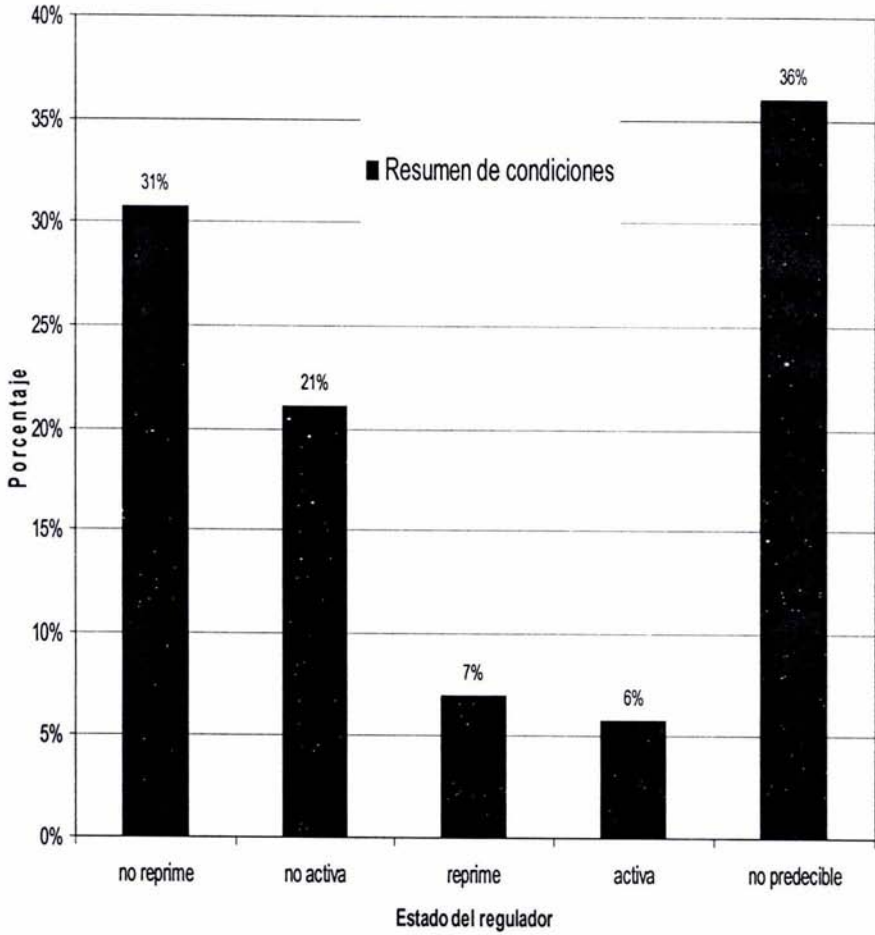


Figura 3.2: Porcentaje de predicción del estado activo o inactivo de los reguladores transcripcionales.

3.6. Otros niveles de regulación de la expresión ¿Pueden ser incorporados en nuestro modelo?

En el capítulo introductorio hemos citado distintos ejemplos de los mecanismos de regulación mediados por proteínas reguladoras. Estos ejemplos junto con 32 otros no citados en el texto fueron, utilizados para modelar las reglas presentadas en el publicación principal de esta investigación. Sin embargo, sabemos que la expresión de los genes esta regulada en muchos niveles. El primero de ellos es el control local ejercido sobre cada operón, de igual forma tenemos un control regional sobre múltiples operones dentro de un regulón y finalmente el control ejercido sobre los regulones a través de un estímulo o modulón (Neidhardt, F.C. and Savageau, M.A. 1996). En este trabajo se han incorporado todos estos niveles de organización de la expresión de los genes a través de las condiciones dentro de los grupos de genes estrictamente coregulados ya sea en su modalidad de regulón simple y regulón complejo. De manera novedosa hicimos uso de la información relacionada con la conformación de pegado al DNA de los reguladores transcripcionales y su función activadora o represora sobre cada gene. Sin embargo, nosotros sabemos que la regulación mediada por factores de transcripción dentro de todos estos niveles depende además de factores como la estructura y composición del DNA. En el caso de la estructura se ha demostrado que el superenrollamiento del DNA afecta a la arquitectura del promotor, modificando la longitud del espaciador regulando con ello el pegado de la RNAP (Hatfield G.W. and Benham C.J, 2002). Del mismo modo, se ha observado que reguladores como HU, IHF, y HNS. mantienen un superenrollamiento negativo modificando con ello la expresión de los genes a los que controlan. Estos parámetros aunque no se han considerado como parte de la regla de manera directa, son tomados en cuenta de manera indirecta en el modelo prediciendo la actividad de HU, IHF, y HNS sobre sus regulones. No obstante, estos reguladores carecen de conformación conocida de pegado al DNA por lo que para la predicción de su actividad de manera más certera es deseable probar condiciones de crecimiento en las que se tenga evidencia previa de modificaciones en el pegado al DNA de estas proteínas.

Este tipo de trabajo hasta el momento no se había reportado, sin embargo, durante el proceso de escritura de esta tesis, fue publicado el trabajo titulado *On schemes of combinatorial transcription logic* (Buchler et al. 2003), que explora los potenciales y limitaciones de un modelo que integra a través de funciones lógicas, las respuestas de los genes como resultado de las combinaciones generadas por los sitios de unión al DNA localizados en *cis*

para las proteínas reguladoras en bacterias. Los autores incorporan en este trabajo las diferentes combinaciones entre elementos reguladores con distintas fuerzas de pegado y de manera importante modelan la interacción de factores de transcripción con sitios de pegado al DNA localizados a distancias lejanas relativas al inicio de la transcripción. Esta última consideración les permite dar un primer acercamiento hacia la modelación de la regulación en organismos eucariotes utilizando este modelo. Por otro lado, los autores incluyen en su trabajo una simulación basada en el método de Shea y Ackers para el fago lambda (Shea and Ackers 1985), integrando a esta simulación los arreglos combinatorios obtenidos en su estudio. Sin embargo, el modelo propuesto para lambda requiere de información difícil de obtener para todos los sistemas de regulación ya caracterizados tanto en bacterias como eucariotes. De modo que como los autores concluyen, no es fácil aplicar este modelo y simular la red genética de un organismo. Así, el encontrar un conjunto de propiedades que puedan probarse sobre la red de regulación de un organismo es una tarea importante y además complicada, que requiere de la exploración de los mecanismos particulares para cada regulón.

3.7. Expansión del modelo a eucariotes unicelulares *Saccharomyces cerevisiae*.

La levadura es un organismos más complejo que cualquier procarionte. Los factores de transcripción que regulan la expresión de los genes se clasifican también como represores y activadores. Sin embargo, no todos los factores de transcripción se unen al DNA directamente, por el contrario, estos factores se unen a otros reguladores que actúan directamente sobre el DNA y con ellos modulan la expresión de los genes. Por otro lado, metabolitos inductores o co-represores tiene la función de dar la conformación a los factores de transcripción para formar los complejos multiproteicos capaces de inducir o reprimir la transcripción.

Un ejemplo de esto es la proteína Gal4 que activa la expresión de los genes *gal* a partir del promotor *gal1*. Se ha visto que en ausencia de galactosa y glucosa, Gal4 se une a la región de activación, pero no activa la transcripción de los genes debido a que se encuentra unida a la proteína Gal80. La proteína Gal3 cuando siente la presencia de galactosa modifica la conformación de Gal80 y permite la activación de los genes vía el factor Gal4. El promotor *gal1* de los genes *gal*, también se encuentra sujeto a represión a través del factor transcripcional Mig1 que se encuentra fosforilado en el citoplasma de la bacteria. La presencia de glucosa inhibe a la kinasa que mantiene fosforilada

a Mig1. Cuando esto sucede Mig1 puede entonces entrar al núcleo, unirse al DNA y reprimir la expresión de los genes *gal*. (Ptashne M. And Gann A. 2002) Como hemos visto, existen consideraciones que deben de hacerse para lograr la expansión de estos modelos a otros genomas: la primera de ellas se relaciona con el nivel de descripción del organismo en estudio; la segunda es un problema de colecta de la información.

De modo que en el primer caso, antes de aplicar directamente nuestras reglas debemos de considerar la organización de las unidades de transcripción en el genoma y determinar si es necesario incorporar reglas *ad hoc* como la relativa a los promotores CRP dependientes en el modelo descrito para *E. coli* (Gutiérrez et al 2003). Para ello será necesario hacer búsquedas en la literatura que nos permitan encontrar al igual que en *E. coli* los patrones de regulación que siguen los factores de regulación en levadura. Para el segundo problema se revisaran todas la bases de datos de levadura para encontrar la información disponible acerca de la función como activadora o represora de los factores de transcripción, conformaciones de pegado al DNA de los mismo así como la identificación de los grupos de genes regulados. La base de datos TRANSFAC (Matys V, et al 2003), consideramos puede utilizarse para estos fines.

Capítulo 4

Conclusiones y perspectivas.

Hemos presentado un método que nos permite evaluar la consistencia entre experimentos de expresión global y la información contenida en la literatura recabada en la base de datos RegulonDB. Los resultados obtenidos reflejan que un modelo que considera la participación de la mayor cantidad de elementos conocidos como es el caso de la conformación de pegado al DNA de los reguladores transcripcionales y las funciones de los factores sobre las unidades de transcripción, dan resultados altamente satisfactorios. Sin embargo, no hemos considerado otros elementos participantes en el proceso de transcripción como lo son por ejemplo, la fortaleza de los sitios de pegado al DNA para cada regulador, antiterminadores e interacciones entre la RNAP y los reguladores, que de ser incorporados, muy probablemente modificarían los niveles de consistencia obtenidos hasta el momento. Así mismo, el análisis de un mayor número de condiciones permitirá no sólo enfatizar los resultados hasta ahora obtenidos sino también obtener mejores predicciones de los estados discretos para los genes. De igual manera, los resultados muestran que el uso de toda la información disponible, organizada adecuadamente, permite encontrar un nivel de representación razonable para la modelación de redes genéticas. No obstante, es importante la generación de datos experimentales que ofrezcan al modelador la posibilidad de expandir los modelos cualitativos a formas cuantitativas, que describan con mayor precisión los eventos de regulación llevados a cabo en la célula.

Ya hemos mencionado en el manuscrito del artículo, que RegulonDB da cuenta sólo del 25 % de la red reguladora de *E. coli*, lo cual significa que conforme esta información aumenta, es posible lograr mejores predicciones sobre la red a nivel cualitativo. Como muestra el artículo en la sección 3.1, encontramos que existen en la red interacciones redundantes sobre reguladores

globales como se ejemplifica para el regulador CRP.

Evaluando de la misma manera, encontramos que el fenómeno de consistencia es sensible a alteraciones de las reglas tal y como mostramos al quitar a los metabolitos o modificar la regla para regulones CRP-dependientes. De modo, que nuestro modelo permite experimentar con diferentes tipos de reglas.

En nuestra búsqueda en la literatura observamos que en algunas UT s el sitio más cercano al promotor domina sobre sitios localizados a distancias más lejanas. Uno pudiera probar una regla de esta naturaleza y evaluar la consistencia en la red, volviendo probablemente a encontrar como resultado que la inserción o modificación en los elementos de la regla dan lugar a diferentes niveles de consistencia. Derivado de lo anterior, podemos decir que una expansión natural de este trabajo es incorporar las predicciones hechas en el genoma de *E. coli* sobre sitios de pegado para proteínas reguladoras (Benítez-Bellón et al. 2002). Las predicciones nos darán por supuesto nuevas construcciones de regulones complejos. Sin embargo, aquí el reto será encontrar sobre cada unidad de transcripción la función específica de los reguladores transcripcionales.

Finalmente, bajo estas condiciones es posible pensar en la elaboración de un modelo de simulación de la expresión. En donde los datos de entrada sean las expresiones discretizadas de los genes en la condición inicial derivadas del control, y la perturbación esté dada por las señales predichas vía esta metodología. Esto genera un modelo de expresión de tipo sincrónico que podría ser modificado introduciendo variables metabólicas tal y como lo ha hecho Bernard Palsson (Covert et al. 2001). En dicho modelo los efectores pueden irse consumiendo o generando como productos del metabolismo cuando sea el caso. Así mismo podemos considerar el uso de las vidas medias de las macromoléculas participantes en el proceso de regulación como elementos modificadores de las concentraciones de dichos elementos.

Así mismo, hemos mencionado que el modelo puede ser extendido a eucariotes sencillos como levadura. Esto como se describió necesitaría de la extracción y colecta de los elementos de regulación.

Apéndice A

Ligas útiles en Internet

BASES DE DATOS DE cDNA UniGene | TIGR Gene Indices | STACK | DoTS

SOFTWARE PARA PROCESAMIENTO DE IMAGENES Axon | BioDiscovery | Imaging Research | NHGRI Microarray Project | TIGR | software tools | Eisen lab

HERRAMIENTAS DE ANÁLISIS BioDiscovery —European Bioinformatics Institute (EBI) Expression Profiler | Eisen lab | Silicon Genetics | Spotfire | X Cluster | TIGR software tools | J express

DATOS DE EXPRESIÓN PÚBLICOS :
<http://genome-www5.stanford.edu/MicroArray/SMD/databases.html>.

OTROS SOFTWARES DIPONIBLES EBI |National Center of Genome Resources | Rockefeller cole Normale Suprieure | Standford

Apéndice B

Tablas.

Cuadro B.1: Genes inducidos y reprimidos en choque térmico.

Bnumber	Gene name	Discrete data	Comparison	Reference
b4171	miaA	on	as literature	9
b4173	hflX	on	as literature	9
b3931	hslU	on	as literature	8
b3347	fkpA	on	as literature	7
b3179	ftsJ	on	as literature	6
b1829	htpX	on	as literature	5
b2571	rseB	off	different from literature	5
b2572	rseA	off	different from literature	5
b1274	topA	on	as literature	4
b0882	clpA	on	as literature	3
b0161	htrA	on	as literature	2
b0439	lon	on	as literature	1
b2573	rpoE	off	different from literature	1
b2614	grpE	on	as literature	1
b3461	rpoH	on	as literature	1
b0014	dnaK	on	as literature	1
b0015	dnaJ	on	as literature	1
b0437	clpP	on	as literature	1
b0438	clpX	on	as literature	1
b3178	hflB	on	as literature	1
b3932	hslV	on	as literature	1,8
b4172	hflq	on	as literature	5,9

Referencias cuadro B1.

1. Gross, C. A. (1996) in it *Escherichia coli* and *Salmonella*: cellular and molecular biology, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 1382.
2. Kim, K. I., Park, S. C., Kang, S. H., Cheong, G. W. and Chung, C. H. (1999) it/ *J Mol Biol* bf/ 294, 1363-74.
3. Thomas, J. G. and Baneyx, F. (1998) *J Bacteriol* **180**, 5165-72.
4. Qi, H., Menzel, R. and Tse-Dinh, Y. C. (1997) *J Mol Biol* **267**, 481-9.
5. Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Diaz-Peredo, E., Sanchez-Solano, F., Perez-Rueda, E., Bonavides-Martinez, C. and Collado-Vides, J. (2001) *Nucleic Acids Res* **29**, 72-4.
6. Tan, J., Jakob, U. and Bardwell, J. C. (2002) *J Bacteriol* **184**, 2692-8.
7. Dartigalongue, C., Missiakas, D. and Raina, S. (2001) *J Biol Chem* **276**, 20866-75.
8. Seol, J. H., Yoo, S. J., Shin, D. H., Shim, Y. K., Kang, M. S., Goldberg, A. L. and Chung, C. H. (1997) *Eur J Biochem* **247**, 1143-50.
9. Tsui, H. C., Feng, G. and Winkler, M. E. (1996) *J Bacteriol* **178**, 5719-31

Cuadro B.2: Genes inducidos y reprimidos en crecimiento anaeróbico

Bnumber	Gene name	Discrete data	Comparison	Reference
b0733	cydA	on	as literature	1
b0734	cydB	on	as literature	1
b0894	dmsA	on	as literature	1
b0904	focA	on	as literature	1
b4122	fumB	on	as literature	1
b2241	glpA	on	as literature	1
b2242	glpB	on	as literature	1
b2243	glpC	on	as literature	1
b0973	hyaB	on	as literature	1
b2996	hyaA	on	as literature	1
b2994	hybC	on	as literature	1
b1224	narG	on	as literature	1
b1276	acnA	off	as literature	1
b0432	cyoA	off	as literature	1
b0431	cyoB	off	as literature	1
b0430	cyoC	off	as literature	1
b0429	cyoD	off	as literature	1
b0428	cyoE	off	as literature	1
b1612	fumA	off	as literature	1
b3426	glpD	off	as literature	1
b0720	gltA	off	as literature	1
b1136	icdA	off	as literature	1
b0116	lpdA	off	as literature	1
b3236	mdh	off	as literature	1
b1109	ndh	off	as literature	1
b0723	sdhA	off	as literature	1
b0724	sdhB	off	as literature	1
b0726	sucA	off	as literature	1
b0727	sucB	off	as literature	1
b0728	sucC	off	as literature	1
b0729	sucD	off	as literature	1
b1676	pykF	on	as literature	1
b1474	fdnG	off	different from literature	1
b4154	frdA	on	as literature	1,5
b2288	nuoA	on	as literature	2
b2287	nuoB	on	as literature	2

Cuadro B.3: continuación genes inducidos y reprimidos en crecimiento anaeróbico

Bnumber	Gene name	Discrete data	Comparison	Reference
b2285	nuoE	on	as literature	2
b2284	nuoF	on	as literature	2
b2283	nuoG	on	as literature	2
b2282	nuoH	on	as literature	2
b2281	nuoI	on	as literature	2
b2280	nuoJ	on	as literature	2
b2279	nuoK	on	as literature	2
b2278	nuoL	on	as literature	2
b2277	nuoM	on	as literature	2
b2276	nuoN	on	as literature	2
b4138	dcuA	on	as literature	3
b3528	dctA	off	as literature	3
b4123	dcuB	off	different from literature	3
b2296	ackA	on	as literature	4
b2525	fdx	on	as literature	4
b0684	fdA	on	as literature	4
b0903	pflB	on	as literature	4
b3924	fpr	off	as literature	4
b4015	aceA	off	as literature	5
b4014	aceB	off	as literature	5

Cuadro B.4: continuación genes inducidos y reprimidos en crecimiento anaeróbico

Bnumber	Gene name	Discrete data	Comparison	Reference
b4018	icrR	off	as literature	5
b2193	narP	on	as literature	6
b4401	arcA	on	as literature	7
b1334	fmr	on	as literature	8
b3365	nirB	on	as literature	9
b0113	pdhR	on	as literature	10
b0114	aceE	off	as literature	10
b0115	aceF	off	as literature	10
b1221	narL	on	as literature	11
b1222	narX	on	as literature	11
b1468	narZ	on	as literature	11
b2297	pta	on	as literature	12
b0356	adhC	off	as literature	12
b3605	letD	off	as literature	13
b0995	torR	off	different from literature	13
b4238	nrdD	off	different from literature	14
b4016	aceK	off	as literature	5
b1187	fadR	off	as literature	5
b4153	frdB	off	different from literature	5

Referencias cuadro B2, B3 y B4

1. Patschkowski T., B. D. M., Kiely J.K., (2000) in *Bacterial Stress Responses*, ed. Storz G., H.-A. R. (ASM Press, Washington D.C.), Vol. 1, pp. 61-68.
2. Bongaerts, J., Zoske, S., Weidner, U. and Unden, G. (1995) *Mol Microbiol* **16**, 521-34.
3. Zientz, E., Janausch, I. G., Six, S. and Unden, G. (1999) *J Bacteriol* **181**, 3716-20.
4. Kessler D., a. K. J. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology*, ed. Neidhardt F., C. R., Ingraham J.L., Lin E.C.C., Brooks-Low K., Magasanil B., Reznikoff S., Riley M., Schaechter M. and Umbarger H.E. (ASM press, Washington, D.C.), Vol. 1st, pp. 199.
5. Cronan J.E., a. L. D. (1996) in *Escherichia coli and Salmonella : Cellular and Molecular Biology*, ed. Neidhardt F., C. R., Ingraham J.L., Lin E.C.C., Brooks-Low K., Magasanil B., Reznikoff S., Riley M., Schaechter M. and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1st, pp. 206.
6. Darwin, A. J., Ziegelhoffer, E. C., Kiley, P. J. and Stewart, V. (1998) *J Bacteriol* **180**, 4192-8.
7. Compan, I. and Touati, D. (1994) *Mol Microbiol* **11**, 955-64.
8. Spiro, S. and Guest, J. R. (1987) *J Gen Microbiol* **133**, 3279-88.
9. Tyson, K. L., Bell, A. I., Cole, J. A. and Busby, S. J. (1993) *Mol Microbiol* **7**, 151-7.
10. Quail, M. A., Haydon, D. J. and Guest, J. R. (1994) *Mol Microbiol* **12**, 95-104.
11. Darwin, A. J. and Stewart, V. (1995) *J Bacteriol* **177**, 3865-9.
12. Gupta, S. and Clark, D. P. (1989) *J Bacteriol* **171**, 3650-5.
13. Iuchi, S., Aristarkhov, A., Dong, J. M., Taylor, J. S. and Lin, E. C. (1994) *J Bacteriol* **176**, 1695-701.
14. Garriga, X., Eliasson, R., Torrents, E., Jordan, A., Barbe, J., Gibert, I. and Reichard, P. (1996) *Biochem Biophys Res Commun* **229**, 189-92.

Cuadro B.5: Genes inducidos y reprimidos en fase estacionaria.

Bnumber	Gene name	Discrete data	Comparison	Reference
b1897	crp	off	different from literature	9
b3961	cyaA	off	different from literature	9
b0978	appC	on	as literature	8
b0979	appB	on	as literature	8
b3806	ftsA	on	as literature	8
b3911	dnaK	on	as literature	8
b0912	nuoI	on	as literature	7
b0973	nuoL	on	as literature	7
b1237	nuoK	on	as literature	7
b1305	nuoA	off	different from literature	7
b1306	nuoB	off	different from literature	7
b1308	nuoG	off	different from literature	7
b1712	nuoF	on	as literature	7
b2149	nuoN	on	as literature	7
b2552	nuoJ	on	as literature	7
b3511	nuoE	on	as literature	7
b4036	nuoM	on	as literature	7
b4396	nuoH	off	different from literature	7
b0854	fepA	on	different from literature	6
b2155	cirA	off	as literature	6
b0756	galM	on	as literature	4
b1236	galU	off	different from literature	4
b2151	galS	on	as literature	4
b2837	galR	off	as literature	4
b3500	mglA	on	as literature	4
b3509	mglC	on	as literature	4
b3510	mglB	on	as literature	4
b0584	glcC	on	as literature	3
b0812	glcB	on	as literature	3
b0435	arcA	off	as literature	2
b4069	acnB	on	as literature	2
b0014	galE	on	as literature	1

Cuadro B.6: continuación genes inducidos y reprimidos en fase estacionaria.

Bnumber	Gene name	Discrete data	Comparison	Reference
b0094	glpD	on	as literature	1
b0118	acnA	on	as literature	1
b0758	hdeB	on	as literature	1
b0759	hdeA	on	as literature	1
b0855	gor	off	different from literature	1
b0882	clpA	on	as literature	1
b1004	pspC	off	different from literature	1
b1274	pspA	off	different from literature	1
b1304	hns	off	different from literature	1
b1440	topA	off	different from literature	1
b1661	cbpA	on	as literature	1
b1749	pspE	on	different from literature	1
b2276	cpxR	off	different from literature	1
b2277	cpxA	off	different from literature	1
b2278	cfa	off	different from literature	1
b2281	wrbA	on	as literature	1
b2282	uspA	on	as literature	1
b2285	rob	on	as literature	1
b2287	potG	on	as literature	1
b2288	potF	on	as literature	1
b2741	pcm	off	different from literature	1

Cuadro B.7: continuación genes inducidos y reprimidos en fase estacionaria.

Bnumber	Gene name	Discrete data	Comparison	Reference
b2743	dnaN	off	different from literature	1
b2980	himA	on	as literature	1
b3357	frdB	on	as literature	1
b3429	hmpA	on	as literature	1
b3430	hyaB	on	as literature	1
b3495	pspB	off	different from literature	1
b3701	galT	on	as literature	1
b3908	proP	off	different from literature	1
b4111	himD	off	different from literature	1
b4401	acs	on	as literature	1
b0758	galT	on	as literature	4.5
b0759	galE	on	as literature	4.5
b1000	bolA	on	as literature	8.1
b2148	oxyR	on	as literature	8.1
b2150	otsB	on	as literature	8.1
b2283	sodA	on	as literature	8.1
b2284	rpoS	on	as literature	8.1
b3912	dps	on	as literature	8.1
b4035	xthA	off	different from literature	8.1
b4153	glgC	on	as literature	8.1
b4154	glgA	on	as literature	8.1

Referencias cuadro B5, B6 y B7.

1. Ishihama, A. (2000) *Annu Rev Microbiol* **54**, 499-518.
2. Cunningham, L., Gruer, M. J. and Guest, J. R. (1997) *Microbiology* **143**, 3795-805.
3. Rabinowitch, H. D., Sklan, D., Chace, D. H., Stevens, R. D. and Fridovich, I. (1993) *J Bacteriol* **175**, 5324-8.
4. Death, A. and Ferenci, T. (1994) *J Bacteriol* **176**, 5101-7.
5. Bohringer, J., Fischer, D., Mosler, G. and Hengge-Aronis, R. (1995) *J Bacteriol* **177**, 413-22.
6. Nystrom, T. (1999) *Curr Opin Microbiol* **2**, 214-9.
7. Wackwitz, B., Bongaerts, J., Goodman, S. D. and Uden, G. (1999) *Mol Gen Genet* **262**, 876-83.
8. Weichart, D., Lange, R., Henneberg, N. and Hengge-Aronis, R. (1993) *Mol Microbiol* **10**, 407-20.
9. Hengge-Aronis, R. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1. pp. 1497.

Cuadro B.8: Genes inducidos y reprimidos en medio mínimo en fase de crecimiento exponencial.

Bnumber	Gene name	Discrete data	Comparison	Reference
b0118	acnB	on	as literature	1
b0723	sdhA	on	as literature	1
b0724	sdhB	on	as literature	1
b0726	sucA	on	as literature	1
b0727	sucB	on	as literature	1
b0728	sucC	on	as literature	1
b0729	sucD	on	as literature	1
b1136	icdA	on	as literature	1
b1276	acnA	on	as literature	1
b1612	fumA	off	diferente from literature	1
b4122	fumB	off	as literature	1
b0720	gltA	on	as literature	2
b1136	icdA	on	as literature	2
b2741	rpoS	on	as literature	2
b3236	mdh	on	as literature	2
b3426	glpD	off	diferente from literature	2
b3495	uspA	on	as literature	2
b2829	metE	on	as literature	2
b0124	gcd	on	as literature	3
b1676	pykF	on	as literature	3
b1723	pfkB	on	as literature	3
b1850	eda	on	as literature	3
b1851	edd	on	as literature	3
b2133	did	on	as literature	1.2
b2779	eno	on	as literature	3
b2926	pgk	on	as literature	3
b3916	pfkA	on	as literature	3
b3919	tpiA	on	as literature	3
b4014	aceB	on	as literature	1.2
b4015	aceA	on	as literature	1.2
b4016	aceK	on	as literature	1.2
b4025	pgi	on	as literature	3
b4388	gapA	on	as literature	3
b0078	ilvH	on	as literature	4
b3670	ilvN	off	diferente from literature	4

Cuadro B.9: continuación genes inducidos y reprimidos en medio mínimo en fase de crecimiento exponencial.

Number	Gene name	Discrete data	Comparison	Reference
b3671	ilvB	off	diferente from literature	4
b3769	ilvM	off	diferente from literature	4
b3770	ilvE	on	as literature	4
b3771	ilvD	on	as literature	4
b3772	ilvA	on	as literature	4
b0720	gltA	on	as literature	5
b0928	aspC	on	as literature	5
b3202	rpoN	on	as literature	5
b3213	gltD	on	as literature	5
b3870	glnA	on	as literature	5
b4053	alr	on	as literature	5
b0032	carA	on	as literature	6
b0071	leuD	on	as literature	4.2
b0072	leuC	on	as literature	4.2
b0073	leuB	on	as literature	4.2
b0074	leuA	on	as literature	4.2
b0273	argF	off	diferente from literature	6
b2818	argA	off	diferente from literature	6
b2938	speA	on	as literature	6
b3172	argG	off	diferente from literature	6
b3237	argR	on	as literature	6
b3359	argD	on	as literature	6
b3767	ilvG	on	as literature	4.2
b3774	ilvC	on	as literature	4.2
b3957	argE	off	diferente from literature	6
b3958	argC	off	diferente from literature	6
b3959	argB	off	diferente from literature	6
b3960	argH	off	diferente from literature	6

Cuadro B.10: continuación genes inducidos y reprimidos en medio mínimo en fase de crecimiento exponencial.

Number	Gene name	Discrete data	Comparison	Reference
b4254	argI	off	diferente from literature	6
b0112	aroP	on	as literature	7
b0388	aroL	on	as literature	7
b0390	aroM	on	as literature	7
b0754	aroG	on	as literature	7
b0908	aroA	on	as literature	7
b1260	trpA	on	as literature	7
b1261	trpB	on	as literature	7
b1263	trpD	on	as literature	7
b1264	trpE	on	as literature	7
b1693	aroD	on	as literature	7
b1704	aroH	on	as literature	7
b2329	aroC	on	as literature	7
b3281	aroE	on	as literature	7
b3389	aroB	on	as literature	7
b3390	aroK	on	as literature	7
b1262	trpC	on	as literature	7.2
b2599	pheA	on	as literature	7.2
b2600	tyrA	on	as literature	7.2
b2601	aroF	on	as literature	7.2
b3572	avtA	on	as literature	4.5
b0116	lpdA	on	as literature	8
b2551	glyA	on	as literature	8
b2903	gcvP	on	as literature	8
b4388	serB	on	as literature	8
b1275	cysB	on	as literature	9
b2421	cysM	on	as literature	9

Cuadro B.11: continuación genes inducidos y reprimidos en medio mínimo en fase de crecimiento exponencial.

Bnumber	Gene name	Discrete data	Comparison	Reference
b3607	cysE	on	as literature	9
b3806	cyaA	on	as literature	9
b0003	thrB	on	as literature	10
b0031	dapB	on	as literature	10
b0166	dapD	on	as literature	10
b0907	serC	on	as literature	8.2
b2472	dapE	on	as literature	10
b2478	dapA	on	as literature	10
b2838	lysA	on	as literature	10
b3433	asd	on	as literature	10
b3809	dapF	on	as literature	10
b4024	lysC	on	as literature	10
b2114	metG	on	as literature	11
b2414	cysK	on	as literature	9.2
b3008	metC	on	as literature	11
b4013	metA	off	diferente from literature	11
b2927	epd	on	as literature	12
b3357	crp	on	as literature	12
b4054	tyrB	on	as literature	7,4,5
b2019	hisG	on	as literature	13.14
b2020	hisD	on	as literature	13.14
b2021	hisC	on	as literature	13.14
b2022	hisB	on	as literature	13.14
b2023	hisH	on	as literature	13.14
b2024	hisA	on	as literature	13.14
b2025	hisF	on	as literature	13.14
b2026	hisI	on	as literature	13.14

Referencias cuadro B8, B9, B10 y B11.

1. Cronan J.E., a. L. D. (1996) in *Escherichia coli and Salmonella : Cellular and Molecular Biology*, ed. Neidhardt F., C. R., Ingraham J.L., Lin E.C.C., Brooks-Low K., Magasanil B., Reznikoff S., Riley M., Schaechter M. and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1st, pp. 206.
2. Tao, H., Bausch, C., Richmond, C., Blattner, F. R. and Conway, T. (1999) *J Bacteriol* **181**, 6425-40.
3. Fraenkel, D. G. (1996) in *Escherichia coli and Salmonella : cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 189.
4. Umbarger, H. E. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 442.
5. Reitzer, L. J. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 391.
6. Glansdorff, N. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Wasington D.C.), Vol. 1, pp. 408.
7. Pittard, A. J. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 458.
8. Stauffer, G. V. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L.,

Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 506.

9. Kredich, N. M. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 514.

10. Patte, J. C. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 528.

11. Greene, R. C. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 542.

12. Ahmad, D. and Newman, E. B. (1988) *J Bacteriol* **170**, 3443-7.

13. Winkler, M. E. (1996) in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reziikoff W.S., Riley, M., Schaechter, M., and Umbarger H.E. (ASM Press, Washington D.C.), Vol. 1, pp. 485.

14. Winkler, M. E., Zawodny, R. V. and Hartman, P. E. (1979) *J Bacteriol* **139**, 993-1000.

Cuadro B.12: Reguladores transcripcionales y sus conformaciones de pegado al DNA.

Regulador	Efeotor	Funcion del regulador	Conforamcion de pegado al DNA
Ada	methylated.	-	pi
Ada	methylated.	+	pi
Ada	methylated.	+	p
AnsC	asparagine.	-	p
AraC	arabinose.	-	p
AraC	arabinose.	+	pi
ArcA	phosphate.	-	pi
ArgR	arginine.	-	pi
ArgR	arginine.	-	pi
ArgR	l-arginine.	-	pi
ArsR	sodium-arsenite-4.	-	p
BaeR	phosphate.	-	pi
BetI	choline.	-	p
BirA	bio-5'-AMP.	-	pi
CadC	cadaverine.	-	p
CadC	lysine.	+	pi
CaiF	carnitine.	+	pi
CpxR	phosphate.	+	pi
CrP	cAMP.	-	pi
Crp	cAMP.	+	pi
CynR	cyanate.	-	p
CynR	cyanate.	+	pi
CysB	thiosulphate.	-	pi
CysB	n-acetylserine.	-	pi
CysB	thiosulphate.	+	pi
CysB	n-acetylserine.	+	pi
DnaA	aTP.	-	pi
DsdC	d-serine.	-	pi
DsdC	d-serine.	+	pi
EvgA	phosphate.	+	pi
FadR	acylCoA.	-	p
FadR	acylCoA.	+	p
FarR	acylCoA.	-	p
FarR	acylCoA.	+	p
FhlA	formate.	+	pi

Cuadro B.13: continuación reguladores transcripcionales y sus conformaciones de pegado al DNA.

Regulador	Efector	Funcion del regulador	Conforamcion de pegado al DNA
FimZ	phosphate.		pi
FlhD	flhC.	+	p
FlhD	flhC.	+	pi
FnR	feSred.	+	pi
FnR	feSoxy.	-	p
FruR	fructose.	-	p
FruR	fructose.	+	p
FuR	magnesium.	-	pi
FuR	iron.	-	pi
GalR	galactose.	-	p
GalS	galactose.	-	p
GcvA	glysine.	+	pi
GcvA	purine.	-	p
GcvA	purine.	-	pi
GlcC	glycolate.	+	pi
GlcC	glyoxylate.	+	pi
GlcC	acetate.	+	pi
GlpR	glycerol-P.	-	p
GntR	gluconate.	-	p
GutR	glucitol.	-	p
IclR	phosphoenolpyruvate.	-	p
IlvY	alfa,acetohydroxybutirate.	-	p
IlvY	alfa,acetohydroxybutirate.	-	pi
IlvY	alfa,acetolactate.	-	pi
IlvY	alfa,acetohydroxybutirate.	+	pi
IlvY	alfa,acetolactate.	+	pi
KdpE	phosphate.	+	pi
LacI	allolactose.	-	p
LacI	iPTG.	-	p

Cuadro B.14: continuación reguladores transcripcionales y sus conformaciones de pegado al DNA.

Regulador	Efector	Funcion del regulador	Conforamcion de pegado al DNA
LexA	ultraviolet.	-	p
Lrp	leucine.	-	pi
Lrp	leucine.	-	p
Lrp	alanine.	+	pi
Lrp	lysine.	+	p
LysR	diaminopimelate.	+	pi
MalT	maltotriose-ATP.	+	pi
MalT	malY.	+	p
MarR	salicylate.	-	p
MelR	melibiose.	-	p
MelR	melibiose.	+	pi
MetJ	adoMet.	-	pi
MetR	metionine.	+	pi
MetR	homocysteine.	+	pi
MIC	iCBGlc.	-	p
ModE	molibdate.	-	pi
NagC	glcNAc.	+	p
NagC	glcNAc.	-	p
NarL	phosphate.	-	pi
NarL	phosphate.	+	pi
NhaR	na+.	+	pi
NrI	phosphate.	+	pi
OmpR	phosphate.	-	pi
OmpR	phosphate.	+	pi
OxyR	oxydize.	-	pi
OxyR	oxydize.	+	pi
OxyR	reduce.	-	pi

Cuadro B.15: continuación reguladores transcripcionales y sus conformaciones de pegado al DNA.

Regulador	Efector	Funcion del regulador	Conforamcion de pegado al DNA
PdhR	piruvate.	-	p
PhoB	phosphate.	+	pi
PhoP	phosphate.	+	pi
PurR	hipoxanthine.	-	pi
RbsR	d-ribose.	-	p
RhaR	rhamnose.		pi
RhaS	rhamnose.	+	pi
SoxR	feSoxi.	-	pi
SoxR	feSred.	-	pi
TorR	phosphate.	-	pi
TorR	phosphate.	+	pi
TreR	trealose6P.	-	p
TrpR	l-tryptophan.	-	pi
TyrR	tyrosine.	-	pi
TyrR	phenylalanine.	-	pi
TyrR	thryptophan.	-	pi
TyrR	tyrosine.	+	pi
TyrR	phenylalanine.	+	pi
TyrR	thryptophan.	+	pi
UhpA	phosphate.	+	pi
UxuR	fructuronate.	-	p
UxuR	mannonic-amide.	-	p
XapR	xanthosine.	+	pi
XylR	xylose.	+	pi
ZntR	zn+.	+	pi

Bibliografía

- [1] Aki, T. and Adhya, S. 1997. Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J* **16**, 3666-74.
- [2] Ansari, A.Z., Bradner, J.E. and O'Halloran, T.V. 1995. DNA-bend modulation in a repressor-to-activator switching mechanism. *Nature* **374**, 371-5.
- [3] Bell, A. and Busby, S. 1994. Location and orientation of an activating region in the *Escherichia coli* transcription factor, FNR. *Mol Microbiol* **11**, 383-90.
- [4] Benitez-Belln, E., Moreno-Hagelsieb, G. and Collado-Vides, J. 2002. Evaluation of thresholds for the detection of binding sites for regulatory proteins in *Escherichia coli* K12 DNA. *Genome Biol* **3**.
- [5] Bertrand-Burggraf, E., Hurstel, S., Daune, M. and Schnarr, M. 1987. Promoter properties and negative regulation of the *uvrA* gene by the LexA repressor and its amino-terminal DNA binding domain. *J Mol Biol* **193**, 293-302 .
- [6] Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-74.
- [7] Buchler, N.E., Gerland, U. and Hwa, T. 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci* **100**, 5136-5141.
- [8] Busby, S. and Ebright, R.H. 1997. Transcription activation at class II CAP-dependent promoters. *Mol Microbiol* **23**, 853-9.
- [9] Chen, Y., Dougherty, E.R. and Bittner, M.L. 1997. Ratio-Based Decisions and the Quantitative Analysis of cDNA Microarray Images. *J. Biomed Opt.* **2**, 364-374.

- [10] Choy, H.E., Hanger, R.R., Aki, T., Mahoney, M., Murakami, K., Ishihama, A. and Adhya, S. 1997. Repression and activation of promoter-bound RNA polymerase activity by Gal repressor. *J Mol Biol* **272**, 293-300.
- [11] Collado-Vides, J. 1989. A transformational-grammar approach to the study of the regulation of gene expression. *J Theor Biol* **136**, 403-25.
- [12] Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, II, Selkov, E. and Palsson, B.O. 2001. Metabolic modeling of microbial strains in silico. *Trends Biochem Sci* **26**, 179-86.
- [13] Dombroski, A.J., Walter, W.A., Record, M.T., Jr., Siegele, D.A. and Gross, C.A. 1992. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* **70**, 501-12.
- [14] Ebright, R.H. and Busby, S. 1995. The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**, 197-203.
- [15] Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**, 14863-8.
- [16] Ellinger, T., Behnke, D., Knaus, R., Bujard, H. and Gralla, J.D. 1994. Context-dependent effects of upstream A-tracts. Stimulation or inhibition of *Escherichia coli* promoter function. *J Mol Biol* **239**, 466-75.
- [17] Enz, S., Braun, V. and Crosa, J.H. 1995. Transcription of the region encoding the ferric dicitrate-transport system in *Escherichia coli*: similarity between promoters for *fecA* and for extracytoplasmic function sigma factors. *Gene* **163**, 13-8.
- [18] Erickson, J.W. and Gross, C.A. 1989. Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high- temperature gene expression. *Genes Dev* **3**, 1462-71.
- [19] Erickson, J.W., Vaughn, V., Walter, W.A., Neidhardt, F.C. and Gross, C.A. 1987. Regulation of the promoters and transcripts of *rpoH*. the *Escherichia coli* heat shock regulatory gene. *Genes Dev* **1**, 419-32.

- [20] Ermolaeva, O., M., R., Pruitt, K.D., Schuler, G.D., Bitter, M.L., Chen, Y., Simon, R., Meltezer, P., Trent, J.M. and Boguski, M.S. 1998. Data management and analysis for gene expression arrays. *Nat. Genet.* **20**, 19-23.
- [21] Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J.L. 1997. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* **61**, 393-410.
- [22] Goosen, N. and van de Putte, P. 1995. The regulation of transcription initiation by integration host factor. *Mol Microbiol* **16**, 1-7.
- [23] Gross, C.A. 1996. Function and Regulation of Heat Shock Proteins. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology 2nd edit. (Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Rezinikoff W.S., Riley, M., Schaechter, M., and Umberger H.E., ed.), Vol. 1, pp. 1382. 2 vols. ASM Press, Washington D.C.
- [24] Hasty, J., Pradines, J., Dolnik, M. and Collins, J.J. 2000. Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci U S A* **97**, 2075-80.
- [25] Haifiel, G.W., and Benham, C.J. 2002, DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.* **36**,175-203.
- [26] Hawley, D.K., Johnson, A.D. and McClure, W.R. 1985. Functional and physical characterization of transcription initiation complexes in the bacteriophage lambda OR region. *J Biol Chem* **260**, 8618-26.
- [27] Hawley, D.K. and McClure, W.R. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* **11**, 2237-55.
- [28] Helmann, J.D. 1991. Alternative sigma factors and the regulation of flagellar gene expression. *Mol Microbiol* **5**, 2875-82.
- [29] Hengge-Aronis, R. 1993a. The role of rpoS in early stationary-phase gene in: Regulation in *Escherichia coli* K12 (Kjelleberg, S., ed.). Plenum Press, New York. Hengge-Aronis, R. 1993b. Survival of hunger and stress: the role of rpoS in early stationary phase gene regulation in *E. coli*. *Cell* **72**, 165-8.

- [30] Heyer, L.J., Kruglyak, S. and Yooseph, S. 1999. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* **9**, 1106-15.
- [31] Hidalgo, E. and Demple, B. 1997. Spacing of promoter elements regulates the basal expression of the *soxS* gene and converts SoxR from a transcriptional activator into a repressor. *EMBO J* **16**, 1056-65.
- [32] Hochschild, A. and Dove, S.L. 1998. Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* **92**, 597-600.
- [33] Ishihama, A. 1988. Promoter selectivity of prokaryotic RNA polymerases. *Trends Genet* **4**, 282-6.
- [34] Ishihama, A. 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**, 499-518.
- [35] Jacob, F. and Monod, J. 1961. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J. Mol. Biol.* **3**, 318-356.
- [36] Kauffman, S. 1974. The large scale structure and dynamics of gene control circuits: an ensemble approach. *J Theor Biol* **44**, 167-90.
- [37] Kim, S.K., Makino, K., Amemura, M., Nakata, A. and Shinagawa, H. 1995. Mutational analysis of the role of the first helix of region 4.2 of the sigma 70 subunit of *Escherichia coli* RNA polymerase in transcriptional activation by activator protein PhoB. *Mol Gen Genet* **248**, 1-8.
- [38] Kim, Y.I. and Hu, J.C. 1997. Oriented DNA binding by one-armed lambda repressor heterodimers and contacts between repressor and RNA polymerase at P(RM). *Mol Microbiol* **25**, 311-8.
- [39] Kohane, I.S., Kho, T.A. and A.J., B. 2003. Microarrays for an integrative genomics. 1st edit, 1. 1 vols, MIT Press, London.
- [40] Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62**, 749-95.
- [41] Kristensen, H.H., Valentin-Hansen, P. and L., S.-A. 1996. CytR/cAMP-CRP nucleoprotein formation in *E. coli*: CytR repressor binds its operator as a stable dimer in a ternary complex with cAMP-CRP. *J. Mol. Biol.* **260**, 113-119.

- [42] Lamark, T., Rokenes, T.P., McDougall, J. and Strom, A.R. 1996. The complex bet promoters of *Escherichia coli*: regulation by oxygen (ArcA), choline (BetI), and osmotic stress. *J Bacteriol* **178**, 1655-62.
- [43] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- [44] Lawley, B. and Pittard, A.J. 1994. Regulation of *aroL* expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. *J Bacteriol* **176**, 6921-30.
- [45] Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. and Lu, P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-54.
- [46] Lopez, P.J., Guillerez, J., Sousa, R. and Dreyfus, M. 1998. On the mechanism of inhibition of phage T7 RNA polymerase by *lac* repressor. *J Mol Biol* **276**, 861-75.
- [47] Maas, W.K. and Clark, A.J. 1964. Studies on the mechanisms of repression of arginine biosynthesis in *Escherichia coli* 11. Dominance of repressibility in diploids. *J Mol Biol* **8**, 365-370.
- [48] Matthews, K.S. 1996. The whole lactose repressor. *Science* **271**, 1245-6.
- [49] Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* Jan 1;31(1):374-8.
- [50] McAdams, H.H. and Arkin, A. 1997. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci* **94**, 814-9.
- [51] McAdams, H.H. and Arkin, A. 1998. Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct* **27**, 199-224.
- [52] McAdams, H.H. and Arkin, A. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* **15**, 65-9.

- [53] McAdams, H.H. and Shapiro, L. 1999. Circuit simulation of genetic networks. *Science* **269**, 650-656.
- [54] McClure, W.R., Hawley, D.K., Youderian, P. and Susskind, M.M. 1983. DNA determinants of promoter selectivity in *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* **47**, 477-81.
- [55] McGuire A.M. and Church, G.M. 2000. Predicting regulons and their cis-regulatory motifs by comparative genomics. *Nucleic Acids Res.* **28**, 4523-4530.
- [56] Meccas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J. and Gross, C.A. 1993. The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev* **7**, 2618-28.
- [57] Merrick, M.J. 1993. In a class of its own—the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**, 903-9.
- [58] Miller, A., Wood, D., Ebright, R.H. and Rothman-Denes, L.B. 1997. RNA polymerase beta' subunit: a target of DNA binding-independent activation. *Science* **275**, 1655-7.
- [59] Niu, W., Kim, Y., Tau, G., Heyduk, T. and Ebright, R.H. 1996. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* **87**, 1123-34.
- [60] Neidhardt, F.C. and Savageau, M.A. 1996. Regulation beyond the operon. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. Washington. DC: ASM.
- [61] Oosawa, C. and Savageau, M.A. 2002. Effects of alternative connectivity on behavior of randomly constructed Boolean networks. *Physica D: Nonlinear Phenomena* **170**, 143-161.
- [62] Pressler, U., Staudenmaier, H., Zimmermann, L. and Braun, V. 1988. Genetics of the iron dicitrate transport system of *Escherichia coli*. *J Bacteriol* **170**, 2716-24.
- [63] Quackenbush, J. 2001. Computational analysis of microarray data. *Nat Rev Genet* **2**, 418-27.
- [64] Rasmussen, P.B., Holst, B. and Valentin-Hansen, P. 1996. Dual-function regulators: the cAMP receptor protein and the CytR regulator can act

- either to repress or to activate transcription depending on the context. *Proc Natl Acad Sci U S A* 93, 10151-5.
- [65] Record, M.T., Jr., W.S. Reznikoff, M.L. Craig, K.L. McQuade, and P.J. Schlax. 1996. *Escherichia coli* RNA polymerase (Es70), promoter and the kinetics of the steps of transcription initiation. In *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edit. (Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Brooks Low, K., Magasanik, B., Reznikoff W.S., Riley, M., Schaechter, M., and Umberger H.E., ed.), Vol. 1, pp. 792-820. 2 vols. ASM Press, Washington D.C.
- [66] Rhodius, V.A. and Busby, S.J. 1998. Positive activation of gene expression. *Curr Opin Microbiol* 1, 152-9.
- [67] Richet, E., Vidal-Ingigliardi, D. and Raibaud, O. 1991. A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. *Cell* 66, 1185-95.
- [68] Rojo, F. 1999. Repression of transcription initiation in bacteria. *J Bacteriol* 181, 2987-91.
- [69] Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262, 1407-13.
- [70] Roy, S., Garges, S. and Adhya, S. 1998. Activation and repression of transcription by differential contact: two sides of a coin. *J Biol Chem* 273, 14059-62.
- [71] Savageau, M.A. 1974. Comparison of a classical and autogenous system of regulation in inducible operons. *Nature* 252, 546-549.
- [72] Savageau, M.A. 2000. Design principles for elementary gene circuits: Element, methods and examples. *Chaos* 11, 142-159.
- [73] Scott, S., Busby, S. and Beacham, I. 1995. Transcriptional co-activation at the ansB promoters: involvement of the activating regions of CRP and FNR when bound in tandem. *Mol Microbiol* 18, 521-31.
- [74] Serrano, M., Salas, M. and Hermoso, J.M. 1990. A novel nucleoprotein complex at a replication origin. *Science* 248, 1012-6.

- [75] Shea, M.A. and Ackers, G.K. 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211-230.
- [76] Shen, J. and Gunsalus, R.P. 1997. Role of multiple ArcA recognition sites in anaerobic regulation of succinate dehydrogenase (*sdhCDAB*) gene expression in *Escherichia coli*. *Mol Microbiol* **26**, 223-36.
- [77] Shingler, V. 1996. Signal sensing by sigma 54-dependent regulators: derepression as a -control mechanism. *Mol Microbiol* **19**, 409-16.
- [78] Spiro, S. and Guest, J.R. 1991. Adaptive responses to oxygen limitation in *Escherichia coli*. *Trends Biochem Sci* **16**, 310-4.
- [79] Valentin-Hansen, P., Sogaard-Andersen, L. and H., P. 1996. A flexible partnership: the CytR anti-activator and cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol* **20**, 461-406.
- [80] Vossen, K.M., Stickle, D.F. and Fried, M.G. 1996. The mechanism of CAP-*lac* repressor binding cooperativity at the *E. coli* lactose promoter. *J Mol Biol* **255**, 44-54.
- [81] Wade, J.T., Belyaeva, T.A., Hyde, E.I. and Busby, S.J. 2001. A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter. *EMBO J* **20**, 7160-7.
- [82] Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace, A.J., Jr., Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L., et al. 1997. An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343-9.
- [83] Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L. and Somogyi, R. 1998. Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci U S A* **95**, 334-9.