

03081 B  
24



UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

CENTRO DE INVESTIGACION SOBRE FIJACION  
DE NITROGENO

CARACTERIZACION GENETICA DE LOS  
ELEMENTOS INVOLUCRADOS EN EL CONTROL  
TRANSCRIPCIONAL DE LA NITROGENASA  
EN *Rhizobium etli*

T E S I S

QUE PARA OBTENER EL GRADO DE:  
DOCTORA EN INVESTIGACION  
BIOMEDICA BASICA  
P R E S E N T A  
MARIA BRENDA VALDERRAMA BLANCO

CUERNAVACA, MOR.

1996

TESIS CON  
FALLA DE ORIGEN

TESIS CON  
FALLA DE ORIGEN



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 está 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.

*Agradezco a los Dres. Jaime Mora, Rafael Palacios, Enrique Morett, Alejandra Covarrubias, Esperanza Martínez, Julio Collado y Mario Soberón sus comentarios durante la revisión de esta tesis.*

*Al Dr. Jaime Mora, su compromiso y su apoyo.*

*A los Dres. Rafael Palacios, Miguel Angel Cevallos y Gabriel Guarneros su participación como Asesores del Proyecto de Investigación. Al Dr. Enrique Morett sus consejos.*

*A mis compañeros del Departamento de Ecología Molecular: Yolanda, Sergio, Carmen, Mere, Araceli, Adriana, Sandra, Mario, Sara, Ramón, Svieta, Jesús, Ernesto, Mike, Alberto, Humberto, Victor, Karine, Mac, Nora y Javier, por su solidaria convivencia del diario y su contagiosa alegría en las celebraciones.*

*He pasado grandes momentos con muchas personas y mencionarlas a todas sería imposible, pero quiero que sepan que los recuerdo: Alma, Lupita, Amparo, Delia, Maluye, Jaime, David, Oscar, Miguel Angel, Sergio, Patrick, Ismael, Toon, Sandra, Luis Ernesto, Ivonne. A Jorge y Gisela. A mis amigos del JBT Humberto, Ricardo, Sergio, Agustino, Josué, Victor, Humber, Gabriel, Fili, Katy, Lety, Jordi, Marco, Lorenzo, Paco, Eugenio, René y Chayo. A Laura, Oscar, Laurita, Paola, Marco y Jesús. A mis amigos del CAD Julio, Linda, Clara, Tony, Tere y Clemen. A Armando, Toño y Nacho.*

*A Sonia, Joanna, Osvaldo, Hiram, Alberto, José Antonio y sobre todo a Juan, porque son la mejor demostración de que la distancia no importa cuando hay cariño.*

*A mis papás por su apoyo incondicional. A mis hermanos, porque mientras más tiempo pasa, más los quiero. A mis tíos, primos y sobrina, por ser una gran familia.*

*Especialmente al Centro de Investigación sobre Fijación de Nitrógeno y a la Universidad Nacional Autónoma de México, por la oportunidad de realizar mis anhelos.*

Este trabajo se realizó bajo la Dirección del Dr. Jaime Mora Celis en el Departamento de Ecología Molecular del Centro de Investigación sobre Fijación de Nitrógeno perteneciente a la Universidad Nacional Autónoma de México.

Durante el desarrollo del proyecto conté con el apoyo de la Beca de Doctorado de la Dirección General de Asuntos del Personal Académico de la UNAM, así como con un complemento del Departamento de Ecología Molecular.

INTRODUCCIÓN .....	1
Generalidades de la expresión génica .....	2
Control del inicio de la transcripción en <i>E. coli</i> .....	4
Localización del promotor y reconocimiento por E $\sigma$ <sup>70</sup> .....	4
Formación del complejo cerrado con E $\sigma$ <sup>70</sup> .....	5
Formación del complejo abierto .....	6
Fuerza y eficiencia del promotor .....	7
Control de la eficiencia de un promotor por factores reguladores .....	8
Interacciones entre factores de la transcripción .....	9
Comparación con el modelo eucariótico .....	9
Factores sigma ( $\sigma$ ) .....	10
Familia de $\sigma$ <sup>70</sup> .....	10
Familia de $\sigma$ <sup>54</sup> .....	11
Transcripción de promotores dependientes de $\sigma$ <sup>54</sup> .....	12
Reconocimiento del promotor .....	12
Formación del complejo cerrado con E $\sigma$ <sup>54</sup> .....	12
Formación del complejo abierto .....	14
Características generales de la familia de proteínas activadoras de promotores	
dependientes de $\sigma$ <sup>54</sup> .....	16
Mecanismo de NtrC .....	17
Mecanismo de DctD .....	19
Mecanismo de NifA .....	20
Los genes <i>nif</i> de <i>Rhizobium etli</i> .....	22
ANTECEDENTES DIRECTOS Y OBJETIVO .....	24
RESULTADOS Y DISCUSIÓN .....	27
Manuscrito del artículo "Regulatory proteins and <i>cis</i> -acting elements involved in the transcriptional control of <i>Rhizobium etli</i> reiterated <i>nifH</i> genes" .....	28
Perspectivas en la aplicación de este trabajo .....	45
Consideraciones sobre la evolución del sistema transcripcional de la nitrogenasa .....	48
Un homólogo de <i>hemN</i> en el plásmido simbiótico de <i>Rhizobium etli</i> .....	54
BIBLIOGRAFIA .....	59
TABLAS Y FIGURAS .....	75

# **ΕΠΙΧΡΟΝΙΣΤΗΡΙΟ**

## Generalidades de la expresión génica

Una unidad transcripcional comprende el ADN entre el promotor, donde inicia la transcripción, y un terminador, donde concluye. Una cadena sencilla del ADN en esta región sirve como templado para la síntesis de una cadena complementaria de ARN, creando una región híbrida ARN-ADN muy corta (2-3 pb) que se mueve a lo largo del ADN junto con la burbuja de transcripción. Caracterizada por su capacidad de incorporar ribonucleótidos al ARN bajo la dirección de un templado de ADN, la ARN polimerasa (ARNP) forma parte de un complejo aparato involucrado en la transcripción: supervisa la fidelidad del apareamiento de bases del híbrido ARN-ADN y cataliza la formación de enlaces fosfodiéster en la cadena nascente de ARN (101) (Fig. 1).

Una de las ARNP bacterianas mejor caracterizadas es la de *Escherichia coli*, la cual es responsable de la síntesis de todos los ARN mensajeros, ARN ribosomales y ARN de transferencia en esta bacteria. Existen alrededor de 7000 moléculas de ARNP en una célula de *E. coli* y dependiendo de la condición de crecimiento de la bacteria pueden ser de 2000 a 5000 de ellas sintetizando ARN al mismo tiempo. La holoenzima ARNP puede ser separada en dos componentes principales: la enzima básica, que es un multímero de estructura  $\alpha 2\beta\beta'$  responsable de la elongación de la cadena de ARN, con un peso molecular de aproximadamente 480 kD y el factor sigma ( $\sigma$ ), una subunidad sencilla que se requiere en la etapa de iniciación para el reconocimiento del promotor. La ARNP de *E. coli* mide aproximadamente 90 x 95 x 160 Å y contiene un canal o zurco de alrededor de 25 Å de ancho en su superficie cuya longitud permite contener hasta 16 pb de ADN (101).

El proceso de transcripción puede ser dividido en las siguientes etapas (Fig.2):

A) Formación del complejo cerrado - Esta etapa comienza con la unión de la ARNP a la doble cadena del ADN en una secuencia específica llamada promotor. La enzima básica tiene una afinidad general por el ADN y la adición del factor  $\sigma$  incrementa su afinidad por los promotores. La estructura formada por la holoenzima unida al promotor se conoce como complejo cerrado.

B) Formación del complejo intermedio - Existen evidencias que indican la presencia de un estado intermedio del complejo de transcripción entre el cerrado y el abierto. Se trata aparentemente de pares de bases de la región de promotor que se desfasan entre sí creando una estructura que pudiera considerarse precursora directa de la burbuja de transcripción.

C) Formación del complejo abierto - La holoenzima lleva a cabo la separación de las cadenas para permitir que la cadena templado se encuentre disponible para la complementación con los ribonucleótidos. La burbuja de la transcripción se crea por un desenrollamiento local que comienza en el sitio de unión de la ARNP. La formación de la burbuja implica un cambio conformacional, tanto de la holoenzima como de las cadenas del ADN, conocido como

isomerización a un complejo abierto.

D) Iniciación - Es la síntesis de los primeros ribonucleótidos enlazados como ARN. La ARNP se mantiene sobre el promotor mientras que se unen los primeros ribonucleótidos, aproximadamente 9. La iniciación puede ser detenida por la ocurrencia de eventos abortivos, en los cuales la enzima sintetiza transcritos cortos (de menos de 9 b) que son liberados, promoviendo la síntesis de un nuevo transcrito. La iniciación termina cuando la ARNP extiende la cadena de ARN más de 9 nucleótidos.

E) Desalojo del promotor - Es la fase durante la cual la enzima se mueve a lo largo del ADN y extiende la cadena creciente de ARN. El factor  $\sigma$  se libera al principio de esta etapa, el patrón de protección de la enzima básica se reduce y la polimerasa pierde su contacto arriba de la posición -35 deslizándose sobre la cadena del ADN dentro de la fase codificadora. Una región de ADN localmente desenrollada, conocida como burbuja de transcripción, se mueve junto con la enzima a partir de este momento hasta el final de la unidad de transcripción. Cuando la cadena naciente de ARN alcanza una longitud de 15 a 20 bases el patrón de protección se reduce aún más, hasta cubrir solamente 30 a 40 pb. Los ribonucleótidos se añaden al extremo 3' de la cadena de ARN, formando un híbrido ARN-ADN en la región desenrollada. Detrás de esta región desenrollada, el templado de ADN se aparea con su contraparte original para formar de nuevo la doble hélice.

F, G, H e I) Elongación - La elongación involucra el movimiento de la burbuja de transcripción al descomponer la estructura del ADN, en la cual la cadena codificadora se aparea con la cadena naciente de ARN en su punto de crecimiento. Durante esta etapa, la polimerasa se transporta de manera discontinua sobre el ADN. Esto ha sido demostrado al provocar artificialmente que la enzima haga pausas en distintos puntos, midiendo en cada uno de ellos la longitud del templado cubierto por la enzima, así como el tamaño y posición de la burbuja de transcripción. Los resultados obtenidos con este método indican que la polimerasa conserva su posición y tamaño hasta que añade de 8 a 10 ribonucleótidos y entonces, súbitamente, brinca hacia adelante. El límite anterior de la burbuja de transcripción permanece constante conforme se añaden los ribonucleótidos mientras que se extiende por el otro extremo. En el momento que la polimerasa brinca hacia adelante la burbuja se reduce de 8 a 10 bases. Estos datos indican que la polimerasa se mueve como una unidad a lo largo del ADN en pasos de aproximadamente 10 pb, y que mientras la burbuja de transcripción se abre continuamente, se cierra de manera discontinua sincronizada con el brinco de la enzima (12,89)

La afinidad de la polimerasa básica por el factor  $\sigma$  disminuye una vez que se compromete en la conformación de elongación. Se ha observado que al abandonar el factor el complejo de elongación, éste se asocia con NusA. Al llegar a un terminador, el complejo de elongación se desprende del ADN y NusA abandona a la polimerasa básica, la cual queda disponible para

interaccionar con otra molécula de factor  $\sigma$  y comprometerse en un nuevo ciclo de la transcripción. Al separarse del ADN, la afinidad de la polimerasa básica por NusA disminuye mientras que aumenta por el factor  $\sigma$  (64) (Fig. 3).

La ARNP es capaz de reconocer el punto donde debe detener la formación de enlaces fosfodiéster entre los ribonucleótidos y el complejo de elongación debe descomponerse. Cuando se añade la última base a la cadena de ARN, la burbuja de transcripción se colapsa mientras el híbrido ARN-ADN se disocia. El ADN se reasocia como doble hélice y la enzima y el ARN se liberan. La secuencia de ADN requerida para estas reacciones se conoce como terminador.

Existen dos tipos de terminadores. Unos son los sitios de terminación intrínseca, que contienen una estructura de tallo-asa rica en pares GC seguida de una corrida de residuos de U. Estas estructuras son reconocidas *in vitro* por la enzima básica sola. Los segundos son los sitios de terminación Rho-dependientes, los cuales requieren de este factor tanto *in vivo* como *in vitro*; la condición común es un tracto de 50 a 90 nucleótidos rico en C y pobre en G precediendo al sitio de terminación. El factor Rho es una proteína esencial que actúa como un factor de terminación adicional, el cual reconoce el ARN y actúa en sitios donde la ARNP se ha pausado. La actividad de terminación requiere de la hidrólisis de ATP.

### **Control del inicio de la transcripción en *E. coli***

La etapa de iniciación de la transcripción puede subdividirse en los pasos secuenciales esbozados en el apartado anterior, todos los cuales son sitios potenciales de control *in vivo*. En esta sección mencionaremos con mayor detalle estos pasos, tomando a *E. coli* como sistema modelo.

Localización del promotor y reconocimiento por  $E\sigma^{70}$  - En *E. coli* se han identificado varios tipos de holoenzimas definidas en base al factor  $\sigma$  que posean. Todas ellas utilizan la misma enzima básica y existen evidencias sobre la competencia de diferentes factores  $\sigma$  por el mismo sitio. Esta competencia está dirigida aparentemente por el equilibrio y el resultado dependerá de la concentración relativa de los diferentes factores  $\sigma$ .

Si tomamos células de *E. coli* creciendo en fase exponencial en ausencia de estrés, encontraremos que el factor  $\sigma$  más representado es  $\sigma^{70}$  (el número indica la masa molecular aparente de la proteína expresada en kD). Aunque otros organismos presentan proteínas homólogas con diferente masa molecular se ha convenido que el de *E. coli* sirva como modelo. Alternativamente y en respuesta a señales extracelulares y de desarrollo se producen otros factores  $\sigma$  con características específicas que participan en la expresión coordinada de genes involucrados en la respuesta al estímulo. En la Tabla 1 se muestran algunas características de estos factores. Debido a la gran similitud a nivel de secuencia de  $\sigma^{70}$  con el resto de los factores  $\sigma$  mencionados en la Tabla 1, se les ha agrupado en la misma familia. En esta sección nos enfocaremos

únicamente en las observaciones llevadas a cabo con  $E\sigma^{70}$ , con el conocimiento de que su mecanismo es esencialmente igual al de los otros elementos de la familia y en la próxima sección hablaremos del único otro factor  $\sigma$  no relacionado.

En principio, aquellos promotores cuyas secuencias son similares compiten entre sí por un tipo particular de holoenzima, es decir, la parte de la población que contenga cierto factor  $\sigma$ . Tal grupo de promotores relacionados podrían definirse como parte de la misma "clase de especificidad". La competencia por la holoenzima entre promotores de la misma clase sería el segundo punto de control de la iniciación de la transcripción, después de la competencia por la ocupación de la enzima básica por los factores  $\sigma$ .

Los promotores bacterianos reconocidos por  $E\sigma^{70}$  están identificados por dos secuencias cortas conservadas en las posiciones -35 y -10 con respecto al inicio de la transcripción y separadas entre sí por 15 a 21 pb.  $E\sigma^{70}$  contacta inicialmente con el elemento -35 (TTGACA) y extiende su contacto hacia el elemento -10 (TATAAT), cubriendo un total de 77 pb. Recientemente se ha identificado en los promotores de los genes ribosomales de *E. coli* un tercer punto de contacto de la polimerasa, denominado elemento UP, que consiste en una secuencia rica en AT localizada alrededor de 20 residuos arriba del elemento -35 (22). La fuerza de la interacción de un promotor dado con la polimerasa correlaciona directamente con la similitud de cada uno de estos tres elementos con su consenso, es decir, con aquella porción de secuencia que une a la respectiva subunidad de la polimerasa con mayor afinidad (22).

Formación del complejo cerrado con  $E\sigma^{70}$  - La formación del complejo cerrado puede ser vista como la reacción de unión de las bases del promotor con  $E\sigma^{70}$ . Dicha reacción posee una constante específica de equilibrio que deriva en el grado de ocupación del promotor, es decir, en la proporción de las secuencias del promotor presentes unidas a  $E\sigma^{70}$ . Esta constante de equilibrio representa la estabilidad de la interacción. El complejo cerrado tiene un potencial definido para isomerizarse espontáneamente a un complejo abierto, caracterizado por la burbuja de transcripción. La apertura del complejo puede considerarse como un proceso cinético con una tasa constante de formación. La reacción de apertura es lenta, pero la reacción contraria es aún más lenta. Esta constante de formación implica no solo la isomerización del complejo de iniciación, sino también la separación de la doble cadena del ADN. Se considera que este proceso es irreversible (183).

Existen evidencias diversas sobre la topología del complejo cerrado formado por  $E\sigma^{70}$ : el análisis de las propiedades electro-ópticas del complejo formado en el promotor A1 del fago T7 demuestra que la holoenzima induce un doblaje de  $45^\circ (\pm 5^\circ)$  sobre el ADN (116). Por otro lado, observaciones directas de complejo formado en el promotor  $P_L$  del fago lambda indican que el doblaje formado por la holoenzima es de aproximadamente  $54^\circ$  (147). A pesar de la diferencia en los valores obtenidos, en ambos casos se comprueba que la adición de la holoenzima induce la formación del doblaje favoreciendo un ángulo específico, probablemente estabilizado por la

estructura de  $E\sigma^{70}$ . Cabe mencionar que estas observaciones se llevaron a cabo sin considerar el cambio conformacional correspondiente al complejo intermedio, por lo que no se precisa si se trata de la última estructura lista para isomerizarse o de algún precursor (Ver mas adelante).

Formación del complejo abierto -  $E\sigma^{70}$  es capaz de interferir con los puentes de hidrógeno que mantienen estabilizada la estructura de doble hélice del ADN, inclusive bajo condiciones que favorecen ésta estructura sobre las cadenas separadas y aún en ausencia de hidrólisis de ATP. Contrario a la idea que se tiene comúnmente, la doble hélice del ADN es una estructura dinámica y sufre disociación-reasociación en cuestión de milisegundos (56). En ausencia de fuerzas externas, un par de bases se separa con una probabilidad de  $1 \times 10^{-5}$ . El tiempo de vida de un par de bases asociado es de  $10^{-2}$  seg en equilibrio con  $10^{-5}$  seg del disociado. Este equilibrio provee una condición en la cuál las proteínas que interaccionan con el ADN puedan contactar con bases que de otra manera estarían secuestradas en la doble hélice (41).

Como se mencionó previamente, estudios cinéticos y estructurales de la iniciación de la transcripción en varios promotores permiten identificar un estado intermedio entre el complejo cerrado y el complejo abierto (2,8,140). La interacción con  $E\sigma^{70}$  induce hipersensibilidad en la región promotora de algunos genes a reactivos químicos o enzimáticos que revelan la presencia de bases mal apareadas. La formación de este intermedio se basa en un proceso de  $E\sigma^{70}$  que prepara al ADN para su disociación. La transición del complejo cerrado al intermedio en este modelo involucra la aplicación de una fuerza sobre la doble hélice, generada por un cambio conformacional de  $E\sigma^{70}$ , que favorecería la disociación de los puentes de hidrógeno de los pares de bases más tensos.

La separación más frecuente observada entre los elementos -10 y -35 es de 17 bases, lo cual corresponde a una distancia de 1.6 vueltas de la doble hélice entre ellos. Sin embargo algunos promotores son más activos (se aumenta su tasa de transcripción), cuando se disminuye el enrollamiento local de 0.3 a 0.5 vueltas, lo que provoca que ambos elementos se encuentren en la misma cara de la molécula con una separación de dos vueltas entre ellos (168).

Este resultado, junto con las evidencias de la existencia del complejo intermedio, pueden interpretarse como prueba de la aplicación por la ARNP de una fuerza que genera una conformación precursora directa de la apertura del complejo. Esta torsión local provoca una tensión entre los pares de bases del promotor que facilitaría su separación posterior. Es importante considerar que todos los cambios conformacionales mencionados son espontáneos, ya que que esta propiedad influye directamente en su modulación.

Todas las subunidades de  $E\sigma^{70}$  participan en las interacciones con el ADN, específicamente,  $\sigma^{70}$  contacta a la posición -3 en la cadena no codificadora, mientras que la subunidad  $\beta$  contacta a la posición +3 en la burbuja de transcripción (162). También se ha encontrado una fuerte asociación, aunque menos específica, de la subunidad  $\beta'$  con la cadena no

codificadora (13). Por otro lado, el extremo carboxilo de la subunidad  $\alpha$  comprende un dominio independiente, capaz de dimerizarse y unirse a ADN en ausencia del resto de la polimerasa. El contacto de este dominio con el ADN es en el sitio UP, localizado arriba del promotor (22).

**Fuerza y eficiencia del promotor** - El termino fuerza del promotor engloba aspectos cuantitativos de la formación del complejo cerrado, su isomerización y la iniciación de la síntesis de ARN, incluyendo la regulación de estos pasos por proteínas reguladoras. La fuerza de un promotor tiene componentes intrínsecos (dependientes de la secuencia) y externos (dependientes de otros factores).

Existen dos partes de la definición de la fuerza de un promotor. La primera es un componente de equilibrio, que relaciona el grado de ocupación del promotor (complejo cerrado) que puede obtenerse con una concentración de  $E\sigma^{70}$  dada. El segundo componente es cinético y relaciona la velocidad a la cuál se isomeriza el complejo de cerrado a abierto. La fuerza del promotor puede definirse como el producto de la constante de equilibrio para la formación del complejo cerrado y la constante cinética de formación del complejo abierto (183).

Estos parámetros pueden evaluarse por separado de manera experimental por medio del ensayo de iniciación abortiva de McClure et al. (109), que aprovecha que durante las primeras etapas de la iniciación el transcrito es relativamente inestable y el complejo de transcripción es susceptible de abortar pequeños productos. En este ensayo, los transcritos se sintetizan utilizando solo algunos de los cuatro ribonucleótidos necesarios para la formación del transcrito completo. Midiendo el tiempo necesario para la producción a tasa constante de estos transcritos abortivos, uno puede determinar la constante de equilibrio y la constante de isomerización del complejo. Valores típicos obtenidos para promotores de *E. coli* van de  $10^7$  a  $10^9$   $M^{-1}$  para la constante de equilibrio y de  $10^{-1}$  a  $10^{-3}$   $seg^{-1}$  para la constante de isomerización.

Después de que los primeros residuos del transcrito naciente se han polimerizado, el complejo de iniciación se convierte en un complejo estable de elongación. Esta conversión involucra la liberación del factor  $\sigma$ , disociación del promotor y un cambio conformacional mayor en la polimerasa básica del complejo. El promotor no es accesible a otra  $E\sigma^{70}$  hasta que la conversión al complejo de elongación de la molécula asociada en el ciclo previo se completa. Estos procesos se conocen como la etapa de desalojo del promotor, las cuales se describen con la misma constante cinética. Adicionalmente, dada la inestabilidad del complejo de iniciación, solo una fracción de todos los eventos de iniciación entrarán a la etapa de elongación. La tasa de desalojo de un promotor determina la disponibilidad de éste para ser ocupado por otra  $E\sigma^{70}$  (183).

Basados en las consideraciones anteriores, es obvio suponer que cada promotor de una clase de especificidad sea caracterizado por una eficiencia particular de iniciar la transcripción. No todos los eventos de entrada de la ARNP dentro de la secuencia codificadora de un gene van a ser exitosos, y el tamaño de esta fracción depende directamente de todos los factores mencionados

hasta ahora. Aún mas, la tasa a la cuál se van a producir transcritos de tamaño completo y funcionales va a ser menor que la tasa de entrada a elongación. Esto se debe a que la polimerasa no siempre llega hasta el final del operón por terminación prematura.

Control de la eficiencia de un promotor por factores reguladores - La eficiencia de la transcripción es modulada por la interacción entre factores que actúan en *trans* y sitios que actúan en *cis*. Un factor que actúa en *trans* es el producto de un gene regulador, usualmente una proteína pero también puede ser ARN. Un sitio que actúa en *cis* en el ADN es una secuencia que funciona al ser reconocida *in situ*, no tiene función codificadora y puede regular solamente aquellas secuencias que se encuentran físicamente contiguas a ella.

La capacidad de la ARN polimerasa de iniciar la transcripción en el promotor es disminuida o aumentada por otras proteínas. Aquellos genes que sean activos a menos que se apaguen son represibles, mientras que aquellos genes que sean activos solamente cuando se les enciende son activables.

Una proteína represora previene que la ARNP se una al promotor o que active la transcripción al unirse a una secuencia específica, llamada operador, que usualmente se encuentra localizada alrededor del inicio de la transcripción. Los operadores son secuencias cortas y frecuentemente palindrómicas, mientras que el represor es usualmente un homomultímero cuya simetría refleja la de su secuencia de unión. La capacidad del represor de unirse al operador depende de la presencia de una molécula pequeña, un inductor (usualmente el sustrato de la enzima codificada por el gene) que previene al represor de unirse mientras que un corepresor lo activa. El mismo represor puede controlar múltiples blancos que contengan la secuencia del operador.

En contraste, otros promotores no pueden ser reconocidos por la ARNP, o son reconocidos pobremente, a no ser que un factor regulador positivo (activador) se encuentre presente. A su vez, las proteínas activadoras pueden ser reguladas por moléculas pequeñas (efectores) o por modificaciones postraduccionales (fosforilación, oxidación, etc.). que median la respuesta a las necesidades de la célula al transducir otras señales químicas, incluyendo la concentración de sustratos, cofactores y mensajeros secundarios. Adicionalmente, la síntesis de los mismos activadores puede ser regulada por procesos metabólicos.

Las proteínas activadoras estimulan la eficiencia de diferentes promotores de *E. coli* de varias maneras: ya sea incrementando la afinidad de  $E\sigma^{70}$  por el promotor a nivel de formación de complejo cerrado, incrementando la tasa de isomerización a complejo abierto o aumentando la velocidad de desalojo del promotor por el complejo de iniciación (183). Los represores están sujetos a los mismos tipos de control metabólico y de desarrollo que los activadores. Un represor puede bloquear un promotor completamente y entonces eliminarlo del grupo que compite por  $E\sigma^{70}$ , previniendo (en un sentido de equilibrio) la formación del complejo cerrado. También es posible que el represor interfiera la isomerización del complejo cerrado a abierto o el proceso de

desalojo del promotor (183).

Interacciones entre factores de la transcripción - La activación o la represión de la iniciación de la transcripción puede involucrar más de una proteína reguladora. Los factores de la transcripción pueden interactuar de manera acoplada o cooperativa para formar complejos reguladores, caracterizados por mayor afinidad o estabilidad.

El ensamblaje de los complejos reguladores involucra múltiples interacciones entre proteínas y ADN, como por ejemplo, la unión secuencial de los reguladores y  $E\sigma^{70}$  al ADN, donde la constante de equilibrio de unión de cada factor aumenta al incorporarse el siguiente. Las especificidades de los contactos proteína-proteína entre  $E\sigma^{70}$  y los factores de transcripción son determinadas a través de segmentos cortos en la superficie expuesta de cada componente.

Se han identificado dos regiones de  $E\sigma^{70}$  involucradas en la interacción con varios factores de la transcripción, una en el extremo carboxilo de la subunidad  $\alpha$  y otra en el extremo carboxilo del factor  $\sigma^{70}$ . Entre los factores que interaccionan con el carboxilo de  $\alpha$  se encuentran: Ada, TrpI, OxyR, OmpR, Crp e IHF (81). Entre los que interactúan con el extremo carboxilo del factor  $\sigma^{70}$  se conocen PhoB y Crp (81). La posición relativa de los sitios de unión a ADN de los factores es dependiente del tipo de interacción que tengan con  $E\sigma^{70}$ , por ejemplo, cuando CRP contacta el carboxilo de  $\alpha$  se une en la posición -60 (*uxuA* o *lac*), mientras que cuando contacta con el carboxilo de  $\sigma$  el sitio se sobrepone con el promotor (*gal*) (81).

Comparación con el modelo eucariótico - La ARNP bacteriana y la eucariótica se componen de múltiples subunidades, a diferencia de las mitocondriales o de fagos, que constan de una sola subunidad. Como se mencionó anteriormente, la ARNP bacteriana consta de 5 subunidades, mientras que la ARNP II eucariótica está compuesta de por lo menos por diez subunidades. Las dos subunidades mayores de la ARNP II están claramente relacionadas a nivel de secuencia con las subunidades  $\beta$  y  $\beta'$  de su homóloga bacteriana. También se han identificado otras dos subunidades con cierta similitud con la subunidad  $\alpha$  (200).

ARNP II reconoce promotores conteniendo la secuencia TATA alrededor de la posición -25 y/o un motivo iniciador en la posición +1, aunque también se han identificado promotores carentes de estos elementos. El complejo de iniciación formado por ARNP II puede ser ensamblado *in vitro*, condición bajo la cual ha sido intensamente estudiado. Algunos otros factores secundarios participan en la formación del complejo cerrado. De estos, TBP, TFIIB, TFIIE y TFIIF presentan una similitud poco significativa con  $\sigma^{70}$  (49).

En contraste con  $E\sigma^{70}$ , la cual lleva a cabo la isomerización del complejo cerrado a abierto espontáneamente, ARNP II requiere de la hidrólisis de ATP. Existen evidencias de la fosforilación de un dominio localizado en el extremo carboxilo de la subunidad grande, el cual consta de un heptapéptido repetido 26 veces en levadura y 52 veces en mamíferos. Aparentemente se requiere la fosforilación intensiva de este dominio para el desalojo del promotor. Interesantemente, también se

requiere de la hidrólisis de ATP para la función de proteínas carentes del extremo carboxilo, por lo que este nucleótido pudiera tener al menos dos papeles diferentes en la isomerización del complejo cerrado (94,142,187).

Al igual que  $E\sigma^{70}$ , el complejo cerrado formado por ARNP II pasa a través de una fase de transcripción abortiva en la cuál algunos transcritos pequeños se sintetizan antes de la formación de un complejo estable de elongación (102). Apparently al entrar en la fase de elongación algunos de los factores iniciales se disocian de la holoenzima.

### Factores sigma ( $\sigma$ )

Los factores  $\sigma$  confieren especificidad a la enzima básica de la ARNP bacteriana para el reconocimiento de promotores. Existen dos grandes familias de factores  $\sigma$ , una relacionada a la secuencia del factor  $\sigma^{70}$ , el factor primario en *E. coli*, y la otra relacionada a  $\sigma^{54}$ , un factor alternativo en *E. coli* originalmente caracterizado por su papel en la transcripción de genes regulados por la fuente de nitrógeno. La existencia de factores  $\sigma$  alternativos provee una forma coordinada de regular la expresión génica en respuesta a condiciones cambiantes o a patrones de desarrollo.

Familia de  $\sigma^{70}$  - Todos los factores de la familia de  $\sigma^{70}$  reconocen promotores compuestos por dos elementos localizados aproximadamente en las posiciones -10 y -35 con respecto al inicio de la transcripción y separados por 15 a 20 pb de secuencia no conservada (Tabla 1). La localización y el espaciamento exactos son específicos para cada factor  $\sigma$ . Las holoenzimas formadas con cualquier factor de esta familia son capaces en principio de isomerizar el complejo cerrado a abierto en ausencia de factores externos aunque es frecuente encontrar que se regulen por otros elementos en *trans* (activadores o represores).

Los factores pertenecientes a esta familia pueden distribuirse en dos grupos, aquéllos que consisten en el factor primario, o sea, aquél más abundante bajo condiciones de crecimiento relajadas (20) y en los alternativos, que se encuentran más representados bajo condiciones de estrés (choque de calor, escasez de fuentes de nitrógeno, diferenciación, etc). Hasta el momento se han caracterizado ambos tipos de factores en bacterias gram-positivas y gram-negativas (ver Tabla 1).

La secuencia deducida para las proteínas de los factores primarios presentan gran similitud, desde 99% de identidad entre las secuencias de *E. coli* y *Salmonella thyphimurium*, hasta 51% entre las secuencias de *Chlamydia trachomatis* y de *Staphylococcus aureus*. Una comparación equivalente de las secuencias de factores alternativos demostró una similitud promedio de 20 a 30% de identidad. Cabe recordar que estos valores corresponden a la comparación de proteínas involucradas en funciones diversas de organismos no relacionados (65).

Con base en estas comparaciones se identificaron cuatro regiones conservadas entre todos los factores estudiados (Figura 4). La región 1, localizada en la porción amino terminal, se encuentra presente solamente en los factores primarios, mientras que las regiones 2, 3 y 4, localizadas en la segunda mitad de las proteínas, están presentes en todos. La región 2 se divide en cuatro subregiones (2.1, 2.2, 2.3 y 2.4) y la región cuatro en dos subregiones (4.1 y 4.2). La subregión 4.2 presenta un motivo hélice-vuelta-hélice (HVH) similar al que se ha demostrado participa en la interacción con ADN en otras proteínas y es aparentemente responsable del reconocimiento del elemento -35 del promotor (65). La subregión 2.1 esta probablemente involucrada en la unión con la enzima básica, mientras que la subregión 2.4 participa en el reconocimiento del elemento -10 del promotor (65). Los determinantes de activación de  $\sigma^{70}$ , con respecto a la separación de las cadenas del promotor, se encuentran localizados en las subregiones 2.1 y 2.3 (65).

Las subregiones 2.4 y 4.2 se encuentran altamente conservadas entre los factores primarios pero no entre los otros miembros de la misma familia. Esto es consistente con su función de reconocimiento de promotores, ya que los factores alternativos varían considerablemente en cuanto a la secuencia específica que reconocen.

Familia de  $\sigma^{54}$  - El factor de iniciación  $\sigma^{54}$  fue identificado originalmente por su papel en la transcripción de genes regulados por la fuente de nitrógeno, como por ejemplo el gene estructural para la glutamina sintetasa en *E. coli* (72,80). Desde entonces se ha demostrado su participación directa en la regulación de la transcripción de genes involucrados en la fijación de nitrógeno, biosíntesis de flagelo, biosíntesis de pili, catabolismo de tolueno y xileno, etc. (44). Recientemente se caracterizó una proteína homóloga tanto funcional como estructuralmente en *B. subtilis*, una bacteria gram-positiva (42).

A pesar que se utiliza el nombre  $\sigma^{54}$  (peso molecular en kD de la proteína de *E. coli*) para identificar a los componentes de esta familia, el peso molecular de las distintas proteínas varia de 47 hasta 59 kD. Sin embargo, y aunque existen otros nombres para este factor ( $\sigma^N$ , RpoN, NtrA)  $\sigma^{54}$  es el más conocido. La comparación de la secuencia de aminoácidos de factores de diferentes organismos permitió identificar tres regiones, dos de ellas fuertemente conservadas (Fig. 5). La región I se localiza en el extremo amino terminal de la proteína y las predicciones de su estructura sugieren que es rica en  $\alpha$ -hélices. También comprende una región de cerca de 50 aminoácidos ricos en glutaminas y leucinas. La región II, de entre 60 y 110 residuos de longitud, es altamente ácida pero no está notablemente conservada. En bacterias entéricas, los primeros 33 residuos de esta región se designan como la repetición de trímeros ácidos, ya que cada tercer aminoácido posee una carga neta negativa (157). La región III es una región compleja de aproximadamente 400 residuos, localizada en el extremo carboxilo terminal de la proteína y contiene dos posibles motivos HVH y una secuencia de 10 residuos altamente conservada conocida como la "caja

RpoN" (112,192)

### Transcripción de promotores dependientes de $\sigma^{54}$

Reconocimiento del promotor - Las proteínas de la familia  $\sigma^{54}$ , a diferencia de las de la familia  $\sigma^{70}$ , reconocen promotores con la secuencia consenso 5'-TGGCAC-N5-TTGCA-3', localizados 11 pb arriba del inicio de transcripción. Los dos elementos invariables de estos promotores son el doblete GG en la posición -24 y el par GC en la posición -12, razón por la cuál se les conoce como promotores -24/-12 (170). Tal como se mencionó anteriormente, gran variedad de genes, con muy diferentes funciones, son transcritos a partir de promotores de este tipo (91).

Dos porciones de la secuencia de  $\sigma^{54}$  presentan una estructura probable de HVH pero solo una de ellas (localizada entre los residuos 366-386 en *K. pneumoniae*) es necesaria para el reconocimiento de la posición -13 del promotor (37,113). Deleciones de la región III que incluyen este motivo abolen la interacción con el elemento -12, sin afectar la unión a ADN por medio del extremo carboxilo de la proteína, probablemente mediante la interacción con el elemento -24 (193). Cerca de la estructura HVH se encuentra otra porción de la proteína que contacta directamente al ADN, tal como se demostró por medio de la formación de enlaces covalentes entre ellos por luz ultravioleta ("UV crosslinks"). Esa subregión tiene una estructura probable de  $\alpha$ -hélice (28).

Formación del complejo cerrado con  $E\sigma^{54}$  - Al igual que  $\sigma^{70}$ ,  $\sigma^{54}$  se une a la polimerasa básica en una relación 1:1 (80). Sin embargo,  $\sigma^{54}$  no comparte las regiones identificadas en  $\sigma^{70}$  como involucradas en la interacción con la polimerasa básica (subregión 2.1, Fig. 4). Tampoco comparte el mecanismo de unión al promotor, ya que  $\sigma^{54}$  no requiere asociarse previamente a la polimerasa básica (16), por el contrario,  $\sigma^{54}$  es responsable de todos los contactos detectados entre  $E\sigma^{54}$  y el ADN, localizados en la misma cara de la doble hélice entre las posiciones -31 y -5 y funcionando como un intermediario-adaptador dentro del complejo (16,24). Por lo tanto, el primer paso para la formación de complejos cerrados con  $E\sigma^{54}$  es la unión de  $\sigma^{54}$  al templado y solo posteriormente se adiciona la polimerasa básica, la cuál estabiliza la interacción  $\sigma^{54}$ -ADN (16).

Como ya se mencionó, no se han detectado contactos del promotor con las subunidades  $\alpha$ ,  $\beta$  o  $\beta'$  en  $E\sigma^{54}$ , por lo que las diferencias observadas en los perfiles de interacción del promotor, ya sea con  $\sigma^{54}$  o con la holoenzima, se deben probablemente a cambios conformacionales de  $\sigma^{54}$  inducidos por la polimerasa básica (129). Estos cambios conformacionales dependen de la secuencia del promotor, sugiriendo que existen diferentes arquitecturas para los complejos cerrados (27,123).

Las funciones de unión al promotor y a la polimerasa básica son independientes en  $\sigma^{54}$  y

se pueden separar mediante la delección de los residuos 178 al 183, con lo que se genera una proteína incapaz de unirse a la polimerasa básica sin lesionar su unión a ADN (193). Por otro lado, esta región es un dominio independiente ya que conserva su capacidad de unión a la polimerasa básica si se aísla del resto de  $\sigma^{54}$  (residuos 107 a 303) (29). Ambas evidencias coinciden en indicar que los determinantes de  $\sigma^{54}$  para la unión con la polimerasa básica están localizados en el extremo carboxilo de la región III (Figura 5). La composición de este dominio es ácida-hidrofóbica, y sustituciones puntuales, tanto de los residuos ácidos como de los hidrofóbicos disminuyen notablemente la función de unión a la polimerasa básica (172). Es posible que el uso de un motivo ácido para la unión a la polimerasa pudiera ser importante para la modulación de la actividad de la holoenzima de forma tal que ocasione que el complejo formado con el promotor se mantenga cerrado en ausencia del activador (193). No se ha determinado aún la subunidad de la polimerasa básica con la que interacciona  $\sigma^{54}$ , aunque se sabe que no es con el extremo carboxilo de la subunidad  $\alpha$ , como se ha demostrado para otros activadores (97).

A pesar de establecer un contacto estable con el promotor,  $\sigma^{54}$  no promueve por sí misma la isomerización del complejo cerrado a abierto, ni siquiera en presencia de los activadores (16). Por tanto, la formación de un complejo cerrado activable depende de alguna interacción crítica entre  $\sigma^{54}$  y la polimerasa básica. Se ha observado que la unión de  $E\sigma^{54}$  al promotor trae como consecuencia una distorsión local del ADN dentro del complejo cerrado, en posiciones adyacentes al elemento -12 (16,129), lo que recuerda al complejo intermedio observado en algunos promotores dependientes de  $E\sigma^{70}$  (Fig. 2). Esta distorsión depende estrictamente de la formación del complejo, y no se observa cuando  $\sigma^{54}$  se une solo (16). La distorsión local es evidente como un sitio hipersensible a reactivos químicos que indican un desempalme entre bases más que la separación de las cadenas.

Aparentemente, el ADN requiere ser desenrollado localmente por la holoenzima como una condición crítica para la isomerización del complejo, en consecuencia, el factor  $\sigma$  no solo tiene un papel en el posicionamiento específico de la polimerasa sobre el promotor, sino también en el cambio conformacional del complejo. Con  $E\sigma^{54}$  la distorsión se detecta inicialmente alrededor del elemento -12 y se propaga hacia el sitio de iniciación de la transcripción, donde finalmente se localiza, cuando una proteína activadora interactúa con el complejo cerrado (129), mientras que con  $E\sigma^{70}$  la distorsión es interna a los elementos del promotor (41). La distorsión alrededor del elemento -12 podría tener un segundo papel, ya que en el caso de ser estabilizada dentro del complejo cerrado (como aparentemente sucede) crearía una barrera energética a vencer por el activador para la formación de la burbuja de transcripción (129). En conclusión, la distorsión de las bases del ADN se encuentra nucleada dentro del complejo formado con  $E\sigma^{54}$  y se define en parte por los contactos existentes con el promotor, pero dependiendo de las interacciones de la polimerasa básica con  $\sigma^{54}$ .

Los determinantes de reconocimiento del promotor han sido asignados al amino terminal de  $\sigma^{54}$ . Una proteína conteniendo una delección en fase de 13 residuos de la región I (del 18 al 31) tiene una mayor afinidad reconocimiento del promotor, pero no contacta ni distorsiona el ADN en la región -12, ni tampoco lleva a cabo la isomerización a complejo abierto (27,129,157,193). Esta región se caracteriza por la presencia de repeticiones regulares de leucinas, alternando una serie de seis leucinas separadas por seis residuos entre ellas (llamada hexada) con una de cuatro leucinas separadas por siete residuos (llamada heptada). Otra copia de la heptada se localiza en el extremo carboxilo de la región III, aunque no se le ha identificado ninguna función. Estas secuencias presentan similitud con un motivo caracterizado en reguladores eucarióticos, llamado "zipper" de leucinas, el cuál forma dos hélices anfipáticas que interactúan entre sí para formar un dominio de unión a ADN (93,182).

Las variaciones observadas tanto por la delección como por saturación de sustituciones de estas leucinas, coinciden en indicar su papel en el reconocimiento del elemento -12 pero no en la unión al ADN por medio del contacto con otros residuos (74). Esta región tiene un papel complejo en la función de  $\sigma^{54}$ , ya que la derivada con la delección no sufre el cambio conformacional inducido por la polimerasa básica en la proteína silvestre. En base a esto se ha propuesto a esta región como sensora y transductora de la presencia de la polimerasa básica al promotor, un requisito indispensable para la isomerización del complejo cerrado (27). Mutaciones puntuales a saturación de las leucinas contenidas en esta región liberan a  $\sigma^{54}$  de la necesidad de un activador para la isomerización del complejo (184), enfatizando su papel como parte de un mecanismo que inhibe la función de apertura del complejo.

Formación del complejo abierto - La región de  $\sigma^{54}$  involucrada en la activación del promotor, o sea, del cambio conformacional de la holoenzima que cataliza la separación de cadenas de ADN y da origen de la burbuja de transcripción corresponde al extremo amino terminal de la proteína. Esta región se caracteriza por la presencia de los residuos de leucina que participan en la isomerización del complejo y que se encuentran intercalados entre residuos de glutamina (Figura 5).

Pequeñas delecciones del extremo amino terminal de  $\sigma^{54}$ , carentes de los residuos 2 a 16, 51 a 76 y 18 a 31, dejan de isomerizar el complejo abierto sin que su capacidad de interactuar con ADN se vea modificada (157,193). Mutaciones puntuales a saturación de estas glutaminas presentan el mismo fenotipo (75). La duplicación de una fracción interna de esta región también modifica la capacidad de isomerización (ver más adelante), todo lo cual indica que esta región forma el dominio de activación.

Una porción de la región II compuesta por residuos ácidos alternados, denominada repetición de trímeros ácidos y localizada entre los residuos 48 y 81, es indispensable para la isomerización del complejo a temperatura ambiente (157). Este motivo se encuentra involucrado en

la modulación de la tasa de formación del complejo abierto, probablemente al disminuir la barrera energética impuesta por los puentes de hidrógeno de la doble hélice (157). Una proteína NifA que contiene una duplicación discreta de esta porción es más eficiente que la proteína silvestre en cuanto a la velocidad de formación, pero no en cuanto al rendimiento (192). Interesantemente ambas proteínas mutantes son igual de eficientes que la silvestre a bajas temperaturas.

Los resultados anteriores indican fuertemente que el amino terminal de  $\sigma^{54}$  contiene todos los determinantes de activación del complejo de iniciación, intercalados de una forma por demás compleja, lo que ha dificultado su análisis. Se ha propuesto recientemente un modelo integrado en el cual el activador contacta con las glutaminas de  $\sigma^{54}$  para generar un cambio conformacional que libera a la holoenzima del bloque impuesto por el motivo de leucinas. Existen evidencias de este cambio conformacional, al observarse diferencias en el perfil de protección del ADN por  $E\sigma^{54}$  durante la activación (16,17). En el modelo, éste cambio conformacional expondría los residuos ácidos sobre la doble cadena del ADN, creando un microambiente de repulsión electrostática que solo podría aliviarse al separarse la doble cadena, dando origen a la burbuja de transcripción (192).

Adicionalmente a las glutaminas localizadas en el extremo amino terminal, otras porciones de la proteína también participan en el contacto con el activador. Específicamente, se ha logrado identificar a la Cis307 de  $\sigma^{54}$  como sitio de contacto con la proteína activadora DctD, aunque por el momento no se sabe si este residuo es indispensable para la activación (99).

La característica más interesante del proceso de isomerización del complejo por  $E\sigma^{54}$  es su estricta dependencia por un activador. Algunos activadores de  $E\sigma^{70}$  incrementan la transcripción al estabilizar la interacción de la holoenzima con el promotor y facilitar la isomerización espontánea del complejo. En contraste, los activadores de  $E\sigma^{54}$  requieren para su función de la hidrólisis de ATP simultánea con la isomerización del complejo (44). Como ya vimos, se considera que la energía obtenida de la hidrólisis es transducida a la holoenzima, la que a su vez la utiliza para superar algún tipo de barrera energética, probablemente a través de cambios conformacionales. Wedel y Kustu (185) establecieron un sistema experimental que les permite la caracterización de la naturaleza de esta barrera al estudiar las formas que adoptan los complejos de iniciación al decaer, considerando como opciones iniciales una barrera cinética o una barrera termodinámica.

Dado que la isomerización de un complejo cerrado estable no se lleva a cabo si el activador no puede hidrolizar ATP aunque todos los demás componentes se encuentren presentes, sumado a que una vez formados, los complejos abiertos decaen muy lentamente, hace muy probable la existencia de una barrera cinética entre las dos formas. Adicionalmente, la posición de equilibrio termodinámico entre las dos formas del complejo favorece a los cerrados, es decir, que la energía libre de los complejos cerrados es menor que la de los abiertos, por lo que la isomerización no ocurre espontáneamente. Por lo tanto, la barrera energética posee ambos componentes, uno

cinético y uno termodinámico. También se demostró que el activador abandona el complejo de iniciación una vez se ha isomerizado, por lo que la transmisión de la energía debe estar acoplada entre la hidrólisis de ATP y la isomerización.

Adicionalmente de la transducción de la energía derivada de la hidrólisis de ATP para promover el cambio conformacional de la holoenzima, Wedel y Kustu exploraron la posibilidad de que el activador pudiera canalizar la energía hacia la separación de las cadenas del promotor, funcionando como una helicasa sitio-específica. Los autores utilizaron una serie de promotores pre-abiertos, es decir, con la porción equivalente a la burbuja de transcripción en forma heteroduplex.  $E\sigma^{54}$  no es capaz de iniciar la producción de un transcrito normal en ninguno los heteroduplex probados en ausencia del activador, con lo que se descarta que la separación de cadenas sea el paso limitante y que el activador funcione como helicasa. Para  $E\sigma^{70}$ , se ha demostrado que la enzima comienza la transcripción espontáneamente a partir de un templado pre-abierto en ambas direcciones aunque no contenga secuencias similares a los motivos consenso, con lo que se demuestra que para promotores dependientes de  $\sigma^{70}$  el paso limitante es el establecimiento de la burbuja de transcripción (173). Todo esto indica que  $E\sigma^{54}$  utiliza la energía transducida por el activador para efectuar una labor mecánica, es decir el cambio conformacional del complejo que lleve a la apertura de la burbuja de transcripción (17,144).

Como conclusión, es aparente que el cambio conformacional de  $E\sigma^{54}$  es el paso limitante para la formación del complejo abierto, no la separación de las cadenas (185). También  $E\sigma^{70}$  sufre un cambio conformacional durante la transición del complejo cerrado al complejo intermedio (Fig. 2) y en ambos casos, el cambio precede la separación de las cadenas, pero solo  $E\sigma^{54}$  necesita superar una barrera energética intrínseca que estabiliza el complejo cerrado, por medio de la hidrólisis de ATP por un activador. Esta barrera puede suprimirse artificialmente al sustituir una serie de leucinas en el extremo amino terminal de  $\sigma^{54}$  (184), lo que indica que  $\sigma^{54}$  posee todos los determinantes estructurales necesarios para la formación del complejo cerrado estable y de su posterior isomerización.

#### **Características generales de la familia de proteínas activadoras de promotores dependientes de $\sigma^{54}$**

Como ya se ha mencionado, una de las características más sobresalientes de los promotores activados por  $E\sigma^{54}$  es su dependencia por una proteína activadora, cuyo papel es, aparentemente, el de transducir la energía derivada de la hidrólisis de ATP para la inducción de un cambio conformacional que permita la isomerización del complejo cerrado y el establecimiento del complejo de elongación.

Este mecanismo de reacción presenta características poco frecuentes en procariotes: la

extraordinaria estabilidad de los complejos cerrados formados por  $E\sigma^{54}$ , la dependencia estricta de una proteína activadora, la necesidad de hidrolizar ATP, etc. La conservación de estas características es notable y ha sido muy fácil encontrar similitudes entre los numerosos regulones de este sistema, tanto a nivel mecánico como estructural. La composición y estructura de las proteínas activadoras no es la excepción. A continuación revisaremos las generalidades encontradas entre los miembros de la familia para centrarnos posteriormente en el mecanismo de reacción de las tres más estudiadas.

Desde que se obtuvieron las primeras secuencias deducidas de elementos de esta familia, fué evidente la notable conservación de su organización general (46,152), especialmente la presencia de módulos con funciones similares. Posteriormente se demostró que se trata de dominios funcionales, es decir, que poseen una función y una estructura local específica, y que conserva ambas al ser separados del resto de la proteína. Normalmente los dominios están enlazados entre sí por porciones flexibles y poco conservadas (194). Esencialmente, las proteínas de esta familia contienen tres dominios: el amino terminal, el central y el carboxilo terminal. El dominio amino terminal se encuentra poco conservado e incluso puede no existir (128). Este dominio corresponde al módulo receptor de la señal en aquellas proteínas que forman parte de un sistema sensor-activador de dos componentes (128). El dominio central se encuentra muy conservado y se ha demostrado que está involucrado en la interacción con  $E\sigma^{54}$  y conserva su capacidad de activación aislado del resto de la proteína (76,78). El dominio carboxilo terminal está poco conservado, a excepción de una estructura semejante a una HVH involucrada en la unión específica al ADN, al igual que el amino, conserva su función de unir ADN aislado del resto de la proteína (124).

Interesantemente, las proteínas activadoras TyrR de *E. coli* y NtrC de *R. capsulatus* presentan la identidad suficiente con otros miembros de esta familia como para considerarlas homólogas, sin embargo no interactúan con  $E\sigma^{54}$ , sino con  $E\sigma^{70}$  (55,79,83). Ambas proteínas carecen de una subregión altamente conservada de 15 residuos localizada en el dominio central que podría estar involucrada en la interacción con  $\sigma^{54}$ . Morett y Segovia propusieron que estas proteínas eran originalmente activadores de promotores dependientes de  $\sigma^{54}$  que cambiaron su especificidad al perder esta subregión (128).

Mecanismo de NtrC - La proteína NtrC (nitrogen regulatory protein C) es el regulador central de la transcripción de los genes estructurales para las enzimas involucradas en la asimilación de amonio. Fue identificada originalmente en bacterias entéricas (*K. pneumoniae*, *E. coli* y *S. typhimurium*), donde ha sido intensamente caracterizada, especialmente en la transcripción del gene *glnA*, el cual codifica para la glutamina sintetasa, enzima clave para la asimilación de amonio. NtrC se une a cinco sitios en la región reguladora de este gene en *E. coli*, pero basta con los dos localizados en las posiciones -140 y -108 antes del inicio de la transcripción

para obtener una expresión máxima. Estos dos sitios presentan un comportamiento similar a los "enhancers" eucarióticos (to enhance - aumentar, incrementar): funcionan eficientemente a grandes distancias del promotor, tanto antes como después del inicio de la transcripción y en cualquier orientación (132,149).

Aunque la proteína nativa puede unirse a estos sitios con alta afinidad, solamente fosforilada puede llevar a cabo la función de activación. La fosforilación de NtrC es catalizada por NtrB (una histidina cinasa), y juntos forman parte de un sistema sensor-activador de dos componentes (141). El circuito controla la función de NtrC como activador en respuesta a la disponibilidad de nitrógeno combinado al modular el estado de fosforilación de la proteína, aumentando la fracción fosforilada en condiciones limitantes de nitrógeno (104,131). El tamaño de esta fracción no se controla directamente por la reacción de fosforilación, sino por la contraria de defosforilación, también dependiente de NtrB (84). La fosforilación de NtrC tiene dos consecuencias: (i) un importante incremento en la unión cooperativa de NtrC-P como dímero a los dos sitios en la región reguladora de *glnA*, y (ii) un cambio conformacional que activa a NtrC como ATPasa (3,115,189).

De acuerdo a la organización general de los miembros de la familia de activadores de promotores dependientes de  $\sigma^{54}$ , la mayor parte de las actividades de NtrC han sido asignadas a uno de sus tres dominios (Fig. 6A). El dominio amino está muy conservado y contiene al sitio de fosforilación (87,119,155). El dominio central es homólogo al presente en todos los activadores de la familia y es capaz de activar la transcripción en ausencia del resto de la proteína, lo que implica que contiene los determinantes de contacto con  $E\sigma^{54}$  (76,128). Este dominio también contiene el sitio de unión de ATP y excepto en NtrC de *R. capsulatus*, la secuencia propuesta por Morett y Segovia como probable sitio de interacción con  $\sigma^{54}$  (128,189). Los determinantes para la dimerización y la unión a ADN se localizan en el extremo carboxilo (86,135,145,161). El determinante para la cooperatividad de la unión de NtrC-P al ADN parece estar dividido en dos componentes, uno localizado en el dominio amino y el otro en el central (33).

La función de los sitios de unión para NtrC es aparentemente la de aumentar la concentración local de la proteína en la cercanía del promotor, con el fin de asegurar una interacción exitosa con  $E\sigma^{54}$ . Por medio del uso de concatémicos se demostró que el sitio de unión no requiere estar en *cis*, pero es necesario que se encuentre cercano al promotor (186). La activación de  $E\sigma^{54}$ , a diferencia de la  $E\sigma^{70}$ , se lleva a cabo a distancia. Específicamente, los sitios de unión para NtrC se encuentran distantes de la polimerasa y no es posible un contacto directo entre las dos estructuras. Observaciones directas de este complejo en el microscopio electrónico indican que la interacción ocurre mediante el doblaje del ADN intermedio (169), lo que ya había sido propuesto anteriormente en base a evidencias genéticas (118) y que parece ser una condición universal para este tipo de complejos (44).

Mecanismo de DctD - La proteína DctD es otro de los activadores de promotores dependientes de  $\sigma^{54}$  que se han estudiado a fondo. Esta proteína regula la transcripción de una permeasa de ácidos dicarboxílicos (DctA) en respuesta a la concentración extracelular de estos compuestos, especialmente ácido succínico. Al igual que NtrC, DctD forma parte de un sistema sensor-activador de dos componentes junto con DctB. DctB se autofosforila bajo condiciones de inducción (presencia de ácido succínico) y transfiere ese fosfato al residuo 55 (ácido aspártico) de DctD (58). DctD-P se une a un par de secuencias localizadas alrededor de 75 pares de bases arriba del inicio de la transcripción de *dctA* tanto en *R. meliloti* como en *R. leguminosarum* (95). Ambas secuencias presentan simetría invertida y son ocupadas por DctD-P *in vitro* en respuesta a la concentración del activador.

La fosforilación de DctD es indispensable para su función de activador transcripcional, al mismo tiempo que aumenta su afinidad por ADN (58). Por otro lado, en ausencia de DctB o en condiciones de no inducción (crecimiento en glucosa), DctD se une a ambos sitios, lo cual ha sido interpretado como un mecanismo de reducción de la activación ilegítima de este promotor por NtrC (92,95). Sin embargo un exceso de DctD también impide la activación legítima por DctD-P, probablemente bloqueando su pegado a ADN (92).

Ambos sitios de unión son necesarios para adquirir niveles máximos de inducción, aunque la afinidad de DctD-P por el proximal es de 50 a 100 veces mayor que por el distal (96). La estructura de DctD-P en solución es aparentemente en forma de dímeros, pero es probable que el grado de oligomerización de la forma activa sea mayor (100). Cinéticas de pegado a ADN obtenidas *in vitro* con variantes en los sitios de pegado indican que la unión de DctD-P es cooperativa y depende de un contacto específico entre las proteínas (96).

A diferencia del comportamiento de los elementos del promotor *glnA* de *E. coli*, donde los sitios de unión para el activador (NtrC) conservan su funcionalidad independientemente de la distancia con respecto al promotor (35), los sitios de unión para DctD en el promotor de *dctA* requieren de una posición específica (96). Este comportamiento es comparable al observado en el promotor *glnH* de *E. coli*, donde inserciones entre el sitio de unión y el promotor anulan la activación por NtrC (34). El drástico efecto posicional de los elementos de estos dos ejemplos se debe a la participación de IHF (Integration Host Factor) una proteína dimerica capaz de doblar el ADN en sitios específicos (63); la cual participa en el establecimiento de una arquitectura productiva a través de la imposición de restricciones topológicas. Estas características están presentes en todos los promotores dependientes de  $E\sigma^{54}$  que se han estudiado (44).

La naturaleza modular de DctD es similar a la de los otros elementos de la familia de activadores de promotores  $\sigma^{54}$ : el dominio amino contiene el módulo de recepción de la señal del sensor (DctB) y el sitio de fosforilación (82). El dominio central es capaz de activar transcripción aislado del resto de la proteína y contiene al sitio de unión de ATP (78). Resultados obtenidos de

la formación de enlaces entre DctD de *R. meliloti* y  $\sigma^{54}$  indican que el dominio central está involucrado en la interacción con la holoenzima, específicamente, con  $\sigma^{54}$  y con la subunidad  $\beta$  (99). Estos datos coinciden con la región propuesta por Morett y Segovia como posible sitio de interacción con  $\sigma^{54}$  (128). El dominio carboxilo contiene la estructura de reconocimiento del sitio de unión a ADN (82) (Fig. 6B).

No se requiere al dominio amino para la dimerización de DctD, lo cual indica su independencia del estado de fosforilación de la proteína (66). Por tanto, los determinantes para la multimerización se localizan aparentemente en el dominio central. Adicionalmente, se demostró que una pequeña región del dominio amino (su extremo carboxilo) inhibe al dominio central en su función de activación (66). Esto contrasta con lo obtenido para NtrC, donde derivadas deletadas del dominio amino son incapaces de activar la transcripción (47).

Una proteína DctD carente del dominio amino presenta una actividad de hidrólisis de ATP independiente de la fosforilación de la proteína (100). Como ya se mencionó anteriormente, la hidrólisis de ATP es un componente fundamental del mecanismo de activación de promotores dependientes de  $\sigma^{54}$  y se ha propuesto que la energía obtenida se transduce al complejo cerrado para su isomerización (185). Estudios de iniciación de la transcripción de *dctA* con esta proteína deletada confirman que existe una dependencia estricta de la isomerización del complejo con la hidrólisis de ATP (100). La actividad de ATPasa de esta derivada de DctD se estimula en presencia de ADN no específico, y notablemente más cuando el ADN contiene los dos sitios de unión canónicos (100). La cinética de la estimulación sugiere que el papel de los sitios de unión de un activador no es solamente incrementar la concentración local de éste, sino la de servir de templado para el ensamblaje de complejos multiméricos activos cerca de los promotores.

Mecanismo de NifA - NifA es el activador central de los genes involucrados en la fijación de nitrógeno en eubacterias, genes *nif* (15,53,139). Al igual que NtrC y DctD, NifA es un activador de promotores dependientes de  $\sigma^{54}$  y pertenece a la misma familia (46,114,128). NifA conserva la estructura modular observada en el resto de los miembros de la familia, pero a diferencia de DctD y NtrC, no es parte de un sistema de dos componentes y no requiere ser fosforilada. Probablemente debido a esta falta de función, el dominio amino de NifA es divergente y su longitud varía de 12 residuos en *R. leguminosarum* bv. *trifolii* a 400 en *Pseudomonas syringae*, careciendo del residuo de aspártico receptor del fosfato presente tanto en DctD como en NtrC (128). El dominio central es altamente similar al de las otras proteínas de la familia, compartiendo el sitio de pegado de ATP y el probable sitio de interacción con  $\sigma^{54}$  (128). Al igual que con el de DctD, este dominio aislado de NifA de *R. meliloti* y *K. pneumoniae* conserva sus propiedades de control positivo (6,76) y contiene al sitio de inhibición por NifL en *K. pneumoniae* (6). Cuando NifA proviene de Rhizobaceas contiene un fragmento extra, localizado entre los dominios central y carboxilo, caracterizado por cuatro residuos de cisteína conservados,

aparentemente involucrados en la inactivación de la proteína por oxígeno (Ver más adelante) (90,128). El dominio carboxilo contiene la estructura HVH involucrada en el reconocimiento del sitio de unión a ADN (128) en un contexto poco conservado (Fig. 6C).

La nitrogenasa es sensible a inactivación por oxidación y para restringir la pérdida energética que implicaría la síntesis improductiva de ésta enzima, por no mencionar las demás proteínas Nif y Fix, todos los organismos fijadores de nitrógeno han desarrollado estrategias de contención en respuesta a altas concentraciones de oxígeno. La estrategia seleccionada por Rhizobaceas consiste en limitar la transcripción de los genes *nif* y *fix* a condiciones de microaerobiosis mediante una cascada reguladora que involucra a las proteínas FixLJ, FixK y NifA (53).

En *B. japonicum* (y aparentemente en todas las especies de *Rhizobium*) la actividad de NifA, y por lo tanto la expresión de los genes *nif*, depende directamente de la concentración de oxígeno. Concentraciones atmosféricas de oxígeno producen la inactivación de NifA, que en estas condiciones ya no lleva a cabo la isomerización del complejo cerrado ni se une al ADN (125) y es más sensible a degradación (77,125). NifA sensa directamente la concentración de oxígeno a través de un cofactor metálico coordinado a la proteína mediante las cisteínas localizadas entre los dominios central y carboxilo (90,128) (Fig. 5D).

En *K. pneumoniae* se ha observado un mecanismo diferente, donde NifA es inactivada no solo por altas concentraciones de oxígeno sino que también responde a la presencia de nitrógeno combinado (121). La inhibición depende de NifL, la cual antagoniza la actividad de NifA de forma estequiométrica, sin involucrar la transferencia de fosfatos o cualquier otra forma de modificación postraduccional obvia (98). La reacción de inhibición ha logrado reproducirse *in vitro*, ya sea con sistemas acoplados de transcripción-traducción o con los componentes puros y se han observado dos diferentes comportamientos. En el primer caso, y utilizando NifL renaturalizado y una derivada fusionada NifA-MalE, ambas de *K. pneumoniae*, se demostró que NifL no sensa directamente la concentración de oxígeno o los niveles de compuestos nitrogenados, sino que aparentemente se requiere otro elemento no identificado (98). Por otro lado, utilizando NifL y NifA puros de *A. vinelandii* se demostró que la inhibición por NifL es sensible a la presencia de ATP y ADP, y que el segundo compuesto la estimula (51). El mecanismo propuesto involucra la existencia de una conformación específica de NifA al unir ADP, diferente a la adquirida cuando une ATP, susceptible de interactuar con NifL, quedando en consecuencia impedida de interactuar con  $E\sigma^{54}$  (126) al mismo tiempo que bloquea su actividad de ATPasa, probablemente al ocupar el sitio de unión de ATP (51). Independientemente del mecanismo real de esta reacción, el extremo carboxilo terminal de NifL contiene todos los determinantes necesarios para interactuar con el dominio central de NifA (6,130).

Al igual que el resto de las proteínas de la familia, NifA puede interactuar con  $E\sigma^{54}$  en

*trans*, pero funciona más eficientemente unida a secuencias específicas colocadas arriba del promotor (TGT-N10-ACA) (19,122). La mayoría de los genes *nif* están precedidos por estos sitios, y se ha propuesto que la cantidad de ellos permite modular finamente la tasa de expresión de un promotor (67). Al igual que con NtrC y DctD, la interacción entre NifA y E $\sigma^{54}$  es estabilizada por IHF (44,73), facilitando su encuentro orientado y estimulando la transcripción de 10 a 30 veces *in vitro* (73). El papel de IHF y la arquitectura que impone a la región promotora son aparentemente parte del mecanismo general de activación de promotores dependientes de  $\sigma^{54}$  (91). Esta estimulación ha sido demostrada específicamente para la expresión de varios genes *nifH*, el modelo más utilizado para el estudio de la actividad de NifA (30,156). La estructura del complejo fue observada en este promotor por medio de microscopía electrónica (73), comprobándose el doblaje en el ADN sugerido previamente en base a evidencias genéticas (18).

Se considera que IHF es utilizado por promotores débiles, es decir, sitios de baja afinidad por E $\sigma^{54}$ , para elevar su eficiencia. Esto coincide con la capacidad de IHF para estimular diferentemente la expresión de varios promotores *nif*, donde mientras más lejana al consenso es su secuencia, son más sensibles a estimulación por este factor (156) (Fig. 7).

#### **Los genes *nif* de *Rhizobium etli***

*Rhizobium etli* es la bacteria endosimbiótica más frecuentemente asociada a raíces de frijol aisladas de la región central de México (108). Anteriormente clasificada como *R. phaseoli* ó *R. leguminosarum* bv. *phaseoli*, recientes evidencias moleculares promovieron su separación como una nueva especie (160). Son bacilos gram-negativos aeróbicos capaces de crecer en medio mínimo con ácidos dicarboxílicos como la mejor fuente de carbono. Forma nódulos en raíces de frijol y es caracterizada a nivel molecular por la presencia de reiteraciones del gene *nifH*, así como la separación de los genes *nodA* y *nodBC* (178). La diversidad genética de aislados de esta especie es la mayor conocida para bacterias (143), lo cual sugiere que proviene de un linaje muy antiguo.

Se han caracterizado hasta el momento en la cepa CE3, considerada la cepa tipo de esta especie, dos copias de los genes *nifDK* y tres de *nifH* (146) los cuales codifican para las enzimas nitrogenasa y nitrogenasa reductasa, así como al gene *nifA* (61) cuyo producto ya fue descrito. También han sido clonados los genes *nodS* (181), *nodBC* (178), *nodA* (178) y *nodIJ* (179) cuyos productos están involucrados en la síntesis y modificación de señales morfogénicas durante la interacción planta-bacteria. Adicionalmente, se requiere la integridad de algunos otros genes para un desarrollo óptimo del proceso simbiótico, aunque sus productos no esten directamente involucrados. De estos últimos, cepas de *R. etli* con una mutación en genes cuyos productos participan en la cadena respiratoria producen nódulos vacíos (163), mientras que cepas incapaces de producir algunos de los lipopolisacáridos presentan deficiencias de diversos grados en el

desarrollo del nódulo (31).

Los genes estructurales para la nitrogenasa están organizados en dos operones *nifHDK* (146) ubicados en el plásmido simbiótico en orientación directa y separados por 120 kb (60). Ambas copias se expresan en nódulo (127) y son necesarias para la expresión óptima de la actividad de nitrogenasa (151). La tercera copia de *nifH* se localizó como independiente de cualquier otro gene (146), y se ubica en el plásmido simbiótico entre las otras dos copias, en orientación invertida (60). Las copias de *nifH* asociadas a los operones se conocen como copias a y b, mientras que la copia independiente se conoce como c (60).

Como se mencionó anteriormente, la presencia de reiteraciones de *nifH* es una característica de *R. etli* y ha sido observada en todos los aislados de esta especie (108). *R. leguminosarum* bv. *phaseoli* está relacionado con *R. etli* y comparte alguna de sus características simbióticas, como la presencia de reiteraciones, el rango de infección de leguminosas, la organización y secuencia de algunos genes, etc (7,9-11,45,117). Una revisión reciente de nuevos marcadores filogenéticos de estas dos especies permitió la reconstrucción de una probable historia evolutiva (160).

Siendo el frijol una planta nativa de mesoamérica, y dada la frecuencia y eficiencia de nodulación de esta planta por *R. etli*, se puede considerar como el endosimbionte original. Al exportarse la planta a Europa, algunas semillas llevaban células de *R. etli*, que una vez en contacto con las poblaciones indígenas de campos europeos transfirieron su plásmido simbiótico a células de *R. leguminosarum* no simbióticas más aptas para sobrevivir en esas condiciones, adquiriendo todos los determinantes característicos de *R. etli*. Evidencias a favor de esta historia son la gran similitud, y en algunos casos identidad, entre los plásmidos simbióticos de ambas especies, comparado con la enorme divergencia entre marcadores cromosomales (150). Una segunda evidencia es la alta conservación genética entre aislados de *R. leguminosarum* bv. *phaseoli*, en contraste con el polimorfismo moderado observado en los biovars *viciae* y *trifolii*, lo cual concuerda con un origen reciente del biovar *phaseoli* (48,160,199).

En conclusión, existen pocas evidencias sobre la regulación de los genes *nif* localizados en *R. etli* y aunque se identificó recientemente al gene que codifica para la proteína reguladora NifA (61) apenas se está comenzando a estudiar el efecto de la carencia de esta proteína.

**ANTECEDENTES DIRECTOS Y  
OBJETIVO**

Nuestro interés particular es estudiar el mecanismo regulador que controla la expresión de los genes reiterados para la nitrogenasa en *R. etli*, los cuales presentan una serie de características particulares. Estos genes, tal como otros genes dependientes de  $\sigma^{54}$  que se han estudiado, presentan una secuencia promotora similar a la consenso 100 residuos arriba del inicio de la traducción. La región codificadora de los tres genes es idéntica, y la identidad se mantiene hacia arriba hasta el promotor. Hacia abajo, dos de los tres genes forman operones *nifHDK* y comparten el mismo patrón de enzimas de restricción (146). Estas copias fueron identificadas como a y b. La tercera copia de *nifH* (copia c) comparte el patrón de enzimas de restricción de las otras dos solamente en su porción codificadora, divirgiendo en las zonas aledañas y no hibridiza contra un detector heterólogo que contiene los genes *nifDK* (146), por lo que se le supuso independiente de estos genes.

La región reguladora de las copias a y b es idéntica hasta por lo menos 100 nucleótidos antes del promotor, y contiene una secuencia similar al sitio de unión de NifA (146). La copia c no tiene parecido con las otras dos en esta porción ni contiene secuencias similares al sitio de unión de NifA. Experimentos sobre la expresión de estos genes, basados en fusiones con el gene para la  $\beta$ -galactosidasa de *E. coli* (*lacZ*), demostraron que las tres copias son transcritas en nódulos de frijol (127). Sin embargo, los niveles absolutos de transcripción no pudieron ser evaluados dada la alta frecuencia de amplificaciones en las cepas recombinantes, las cuales generaban una copia extra de la fusión (127).

Inserciones polares en cada una de las tres copias de *nifH* sirvieron para evaluar su papel en la función simbiótica de la cepa con plantas de frijol (151). Inserciones en las copias a o b disminuyen al 50% la actividad de reducción de acetileno y el contenido de nitrógeno de las plantas, lo que se interpreta como evidencia de un fenómeno de dosis génica entre ambas copias (151). Inserciones en la copia c no presentan fenotipo simbiótico (151).

La presencia de las copias reiteradas de *nifH* ha llamado nuestra atención por dos razones principales. Por un lado, la necesidad de tener dos operones *nifHDK* para alcanzar niveles máximos de nitrogenasa en simbiosis, cuando en todos los demás organismos estudiados basta con una copia. Este fenómeno no es común en bacterias, ya que por lo general basta con la regulación de la expresión de un gene para modular el nivel del producto. Como se mencionó en la Introducción, ésta regulación puede llegar a ser muy compleja sin necesidad de involucrar duplicación. En los casos donde se ha observado, la duplicación de genes estructurales normalmente conduce a su regulación diferencial, por ejemplo *nifH* de *Azorhizobium caulinodans*, donde una copia se expresa en vida libre y la otra en condiciones simbióticas (133). Por otro lado, la existencia de la copia c es difícil de justificar. La identidad de este gene con los otros dos es sorprendente dado que aparentemente no posee una función en simbiosis y por lo tanto, es difícil pensar que su secuencia esté sometida a presión selectiva. En principio, las diferencias en su

secuencia reguladora indican que su perfil de expresión pudiera ser diferente, de hecho, la aparente ausencia de secuencias de unión de NifA sugiere que este gene se expresa de manera independiente al regulón simbiótico.

En dos proyectos anteriores profundizamos la caracterización de la región del plásmido simbiótico de *R. etli* que contiene a la copia c. Inicialmente llevamos a cabo un estudio a nivel estructural, basado en la identificación de secuencias similares a los genes *nifHDK* diferentes de las dos mencionadas. Este trabajo proporcionó evidencias de la presencia de una secuencia similar a *nifD*, pero no a *nifK*, asociada a la copia c (176). Posteriormente llevamos a cabo la subclonación de la copia c y sus porciones aledañas y demostramos que la zona de similitud con *nifD* consiste en la porción amino terminal de éste gene. El gene truncado (*nifD\**) está ligado transcripcionalmente al gene *nifHc* y es idéntico en su secuencia nucleotídica al gene *nifD* de la región b en 292 de sus 307 residuos. No tenemos evidencias sobre la función de *nifD\**, pero sabemos que no es exclusivo de la cepa CE3, sino que se encuentra en todos los aislados de *R. etli* que analizamos, siempre asociado a la tercera reiteración de *nifH* (177).

Simultáneamente comenzamos un estudio sobre la expresión de dos de las reiteraciones, insertando en el plásmido simbiótico fusiones transcripcionales de *nifHa* y *nifHc* con el gene *lacZ* y cuantificando su expresión en vida libre y en simbiosis. Como mencionamos anteriormente, *nifHa* está precedido por secuencias similares a un promotor dependiente de  $\sigma^{54}$  y de un sitio de unión de NifA, mientras que *nifHc* solo presenta la secuencia del promotor. Con base en esto podemos predecir que el primer gene, pero no el segundo, va a ser expresado en microaerobiosis, una condición de inducción reconocida para la actividad de NifA en otras especies de *Rhizobium*. Contrario a lo que esperábamos, ambos genes se expresan en cultivos microaeróbicos, siendo más alta la transcripción de *nifHc* que de *nifHa*. Estas expresiones son exclusivas de bajas concentraciones de oxígeno disuelto y tienen un máximo en 1%. También determinamos la actividad de  $\beta$ -galactidasa en nódulos inducidos en raíces de frijol por las cepas con las fusiones. En estas condiciones, ambos genes se expresaron, pero la actividad de la fusión *nifHc-lacZ* fué mayor (177).

Estos resultados aumentaron nuestro interés en dilucidar el mecanismo de regulación de la expresión de los genes *nifH* reiterados, pues su comportamiento transcripcional era contradictorio con nuestras predicciones, hechas con base en los datos disponibles.

El objetivo principal de este trabajo es el análisis del mecanismo de regulación de la expresión de los genes *nifH* reiterados de *R. etli* cepa CE3, y la identificación de los elementos reguladores involucrados. Como estrategias decidimos continuar con el estudio de las fusiones mientras que comenzamos una caracterización de la región reguladora de *nifHc* a nivel de secuencia. La expresión de las fusiones fué estudiada en dos contextos genéticos, en el original de *R. etli* y en un sistema heterólogo en *E. coli*.

## **RESULTADOS Y DISCUSIÓN**

*Manuscrito del artículo "Regulatory proteins and cis-acting elements involved in the transcriptional control of Rhizobium etli reiterated nifH genes"*

## Regulatory Proteins and cis-Acting Elements Involved in the Transcriptional Control of *Rhizobium etli* Reiterated *nifH* Genes

BRENDA VALDERRAMA<sup>1\*</sup>, ARACELI DAVALOS<sup>1</sup>, LOURDES GIRARD<sup>2</sup>, ENRIQUE MORETT<sup>3</sup>  
AND JAIME MORA<sup>1</sup>

<sup>1</sup>Departamento de Ecología Molecular<sup>1</sup> and Departamento de Genética Molecular<sup>2</sup>, Centro de Investigación sobre Fijación de Nitrógeno, AP565-A and Departamento de Reconocimiento Molecular y Bioestructura<sup>3</sup>, Instituto de Biotecnología, AP510-3, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

In *Rhizobium etli* the nitrogenase reductase genes are reiterated. Strain CE3 has three copies: *nifH<sub>a</sub>* and *nifH<sub>b</sub>* form part of *nifHDK* operons with the nitrogenase structural genes, while *nifH<sub>c</sub>* is linked to a truncated *nifD* homologue. Their sequences are identical up to 6 residues upstream from a  $\sigma^{54}$ -dependent promoter. A remarkable difference among them is the absence of canonical NifA binding sites upstream of *nifH<sub>c</sub>*, while there is one located 200 bp upstream of *nifH<sub>a</sub>* and *nifH<sub>b</sub>*. To evaluate the transcriptional regulation of the reiterated *nifH* genes we constructed fusions of *nifH<sub>a</sub>* and *nifH<sub>c</sub>* with the *lacZ* gene of *Escherichia coli*. Both genes were expressed at maximum levels under 1% oxygen in free living cultures, declining as the oxygen concentration was increased. This expression was dependent on the integrity of *nifA*, and *nifH<sub>c</sub>* was expressed at higher levels than *nifH<sub>a</sub>*. The same pattern was observed in root nodule bacteroids. Expression of both genes in *E. coli* required  $\sigma^{54}$  in addition to NifA bound to UAS. In vivo DMS footprinting analyses showed that NifA binds to the canonical site upstream of *nifH<sub>a</sub>* and to a 161 half-site 6 nucleotides further upstream. NifA protected an imperfect binding site upstream of *nifH<sub>c</sub>* at position 85 from the promoter. IHF stimulated each gene differently, *nifH<sub>a</sub>* being more dependent on this protein. The above results correlate the asymmetric arrangement of cis-acting elements with a differential expression of the reiterated *nifH* genes, both in culture and in symbiosis with bean plants.

Eubacterial nitrogen fixation genes (*nif* and *fix*) are generally transcribed from  $\sigma^{54}$ -dependent promoters located between positions -26 and -11 relative to the transcription start site and having the consensus sequence 5'-TGGCAC-N<sub>5</sub>-TTGCA/T-3' (37).  $\sigma^{54}$  (also known as RpoN, s<sup>N</sup> or NtrA) assists in the assembly of a preinitiation closed complex with the RNA polymerase and the promoter DNA (E $\sigma^{54}$ ). Isomerization of this stable closed complex is dependent on the NifA regulatory protein and hydrolysis of a nucleoside triphosphate (37). NifA binds to a conserved sequence motif (5'-TGT-N<sub>10</sub>-ACA-3'), located at a distance ranging from 80 to 150 nucleotides upstream from the transcriptional start site (known as Upstream Activator Sequences or UAS) (36). It has been proposed that the function of the UAS is to increase the local concentration of NifA and to correctly orientate it in the vicinity of a given promoter to facilitate a productive interaction with E $\sigma^{54}$  (8,25). Once NifA is bound to its UAS it approximates the preinitiation closed complex by looping the intervening DNA (8,25). In this regard, expression of several  $\sigma^{54}$ -dependent promoters is enhanced by the Integration Host Factor (IHF), a DNA-bending dimeric heteroprotein that binds specifically between the UAS and the promoter (25,29). Although remote activation is a general mechanism for the expression of  $\sigma^{54}$ -dependent genes, certain strong promoters, whose

sequences strongly resemble the consensus, can be partially activated even in the absence of UAS or by NifA proteins impaired on their DNA binding ability (4,7,26,38), indicating that the DNA binding and the activation of transcription are different functions of the same regulatory protein. Gubler (23) proposed that the presence and number of UAS in a NifA-dependent promoter provides a fine-tuning mechanism for the regulation of gene expression.

Oxygen concentration is the major environmental signal regulating *nif* gene expression. This regulation is enacted by controlling both the expression and the activity of NifA. Autoactivation has been observed in several species (19). In *Rhizobium meliloti* *nifA* expression is additionally controlled by the oxygen sensing, two-component FixLJ system (19), whilst its activity is directly controlled by the oxygen concentration, probably through the oxidation of a metal cofactor linked to a conserved set of cysteine residues (19).

All *Rhizobium etli* (formerly *R. phaseoli*) (53) strains characterized to date have three copies of the nitrogenase reductase (*nifH*) gene (32,46). As in most of the nitrogen fixing eubacteria, in the type strain CE3 *nifH<sub>a</sub>* and *nifH<sub>b</sub>* are transcriptionally coupled with the nitrogenase structural genes forming *nifHDK* operons, while no homologues of the latter genes have been found downstream of *nifH<sub>c</sub>* (47). The three *nifH* genes

are actively expressed during symbiosis with bean plants, but it seems that only the integrity of both *nifHDK* operons is required to achieve full nitrogenase activity (39,50). The sequence of the three genes is identical up to 6 nucleotides upstream from a putative  $\sigma^{54}$ -dependent promoter (5'-ATGGCACGGGTTTCAA-3'). The identity between *nifHa* and *nifHb* extends at least up to 110 nucleotides further upstream and comprises a sequence resembling a NifA binding site (UAS). In contrast, no UAS has been found upstream of *nifHc* (47). Although there are no functional evidences for any of these *cis*-acting regulatory elements, their asymmetric arrangement suggest that the reiterated *nifH* genes are differentially expressed.

In this report we analyzed the expression of the reiterated *nifH* genes of *R. etli*. We show that *nifHa* and *nifHc* are differentially transcribed both in symbiosis and in free living conditions, where they are dependent on NifA,  $\sigma^{54}$  and low oxygen concentrations. Expression and *in vivo* DMS footprinting analyses showed that the DNA binding ability of NifA is essential to activate both genes, although NifA protected different sites upstream of each gene. We also observed a distinct dependence on IHF, with *nifHa* being more dependent than *nifHc*.

## MATERIALS AND METHODS

**Bacterial strains, plasmids and media.** Plasmids and strains used in this work are described in Table 1. *Rhizobium* strains were grown at 30°C in PY complex medium or minimal medium containing succinate and ammonium chloride as sole carbon and nitrogen sources (18). *E. coli* strains were grown at 37°C in LB complex medium or at 30°C in minimal medium containing 10 mM glucose, 10 mM ammonium chloride or glutamine as nitrogen source, 11 mM K<sub>2</sub>HPO<sub>4</sub>, 51 mM KH<sub>2</sub>PO<sub>4</sub>, 4 mM MgSO<sub>4</sub>·7H<sub>2</sub>O. Antibiotics for *E. coli* strains were added in the following final concentrations: 10 µg/ml tetracycline, 100 µg/ml ampicillin, 15 µg/ml chloramphenicol and 30 µg/ml kanamycin.

**Construction of *nifH-lacZ* transcriptional fusions.** EcoRI fragments carrying the *nifHa* and *nifHc* regions from *R. etli* strain CE3 (47) were subcloned in pSUP205 (54) to generate plasmids pEM15 (39) and pDEM23 respectively. An interposon containing the promoterless structural gene for  $\beta$ -galactosidase and a kanamycin-resistance gene (28) was cloned in the *nifH* internal BglII site and the orientation was checked, resulting in plasmids pDEM153 (*nifHa-lacZ*) and pDEM233 (*nifHc-lacZ*). *R. etli* strain CE3 derivatives DEM153 and DEM233 were obtained as one step double recombinants of the respective plasmid (pDEM153 or pDEM233) into the symbiotic plasmid using the suicide vector method (24). Selection was carried out at 15 µg/ml kanamycin and 20 µg/ml nalidixic acid. All strains were regularly analyzed for genomic stability (i.e. absence of rearrangements) by Southern analysis (51).

**Construction of pMJ220tc.** Plasmid pMJ220 carries the *nifA* gene from *K. pneumoniae* cloned in the pEMBL8 expression vector (36). An omega interposon carrying a tetracycline resistance gene (45) was cloned in

the EcoRI site internal to the chloramphenicol resistance gene of the vector. This construction, plasmid pMJ220tc, preserved the activation ability of the pMJ220.

**Ex-planta induction system.** Cells from exponential phase cultures of *R. etli* strains grown in PY medium were collected and washed with minimal medium. Flasks containing minimal medium were inoculated at an initial A<sub>540</sub> of 0.05. Aliquots (20 ml) were injected into sealed 150 ml bottles previously flushed with several volumes of the appropriate oxygen/argon mixture (analytical grade, Linde, México). After incubation for 8 hours, samples for  $\beta$ -galactosidase and protein estimations were withdrawn with syringes and assayed as described below.

**Heterologous expression system in *E. coli*.** *E. coli* strains were transformed by the TSS method (51). Once the presence of the desired plasmid(s) was confirmed by restriction analysis, the strains were grown in LB complex medium with antibiotics until saturation. Minimal medium containing either ammonium chloride or glutamine, as needed, was inoculated at an initial A<sub>600</sub> of 0.05. Samples (0.1 ml) were removed from exponentially growing cultures and assayed for  $\beta$ -galactosidase activity as described below.

**Plant growth conditions.** Surface sterilized *Phaseolus vulgaris* L. cv. Negro Jamapa seeds were germinated under sterile conditions for three days in darkness. The seedlings were transferred to plastic pots containing sterile vermiculite as a support and inoculated with 1 ml of an overnight culture grown in PY. Plant growth and watering were carried out under aseptic conditions in a greenhouse.

**Estimation of  $\beta$ -galactosidase activity.** Nodules from single plants were crushed in cold Z buffer (51), two drops of chloroform were added and the preparation was vortexed and centrifuged at 4°C.  $\beta$ -galactosidase activity was measured as an increase in sample absorbance at 420 nm in the presence of 0.8 mM ONPG. Activities in culture were measured as recommended (51). Aliquots of nodule extracts (0.05 ml) or bacterial cultures (1.5 ml) were precipitated with one volume of 10% trichloroacetic acid and the protein content was estimated by the Lowry method. Specific activities are reported as nmoles of ONP produced per min per mg of protein. Miller units were calculated as described (51).

**DMS footprinting.** *E. coli* cells were inoculated in minimal medium including 2 mM IPTG and grown to an A<sub>600</sub> of 0.6. At this point,  $\beta$ -galactosidase activity was estimated to ensure that induction had occurred. Dimethyl sulphate (DMS) was added (to a 0.1% final concentration) from a fresh 2% solution in the same medium. Cells were incubated for 1 min, diluted in cold saline phosphate (SP), collected by centrifugation and washed once with SP. Plasmid DNA prepared from a 100 ml culture was obtained by the alkaline lysis method (51) and resuspended in 0.1 ml of water. Piperidine was added to a 10% concentration and heated at 90°C for 30 min to cleave the DNA at methylated guanine residues. Piperidine was vacuum evaporated, the plasmid DNA washed and resuspended in 0.1 ml of water. DNA was annealed with 0.5 pmol of 5'-end-labelled oligonucleotide (5'-GTAAAATGCGATTTGACGC-3') and extended with Klenow DNA polymerase (Amersham, UK) for 10 min at 50°C. The reaction was precipitated with ethanol and resuspended in formamide dye mix.

**DNA sequencing and analysis.** Double-stranded templates were sequenced by the dideoxy-method (51) using the Sequenase kit (Amersham, UK). Sequencing reactions mixtures were electrophoresed in 6% polyacrylamide-8 M urea gels in a Bio Rad Sequi-gen sequencing apparatus. Computer analysis of the sequences

was accomplished with the Genetics Computer Group package programs (15) and the Seqscan program, kindly provided by Tracy Nixon, and now available at the URL <http://www.bmb.psu.edu/seqscan>. The *R. etli* DNA sequences mentioned here have been deposited in GenBank and assigned accession numbers L14672, L14671, U38245 and U38244.

## RESULTS

**Free living expression of the *nifH-lacZ* fusions.** Multiple identical nitrogenase genes in *R. etli* precludes the appropriate identification of messenger RNA coming from individual genes. In order to characterize the mechanism involved in their transcriptional regulation we constructed locus-specific fusions with the *E. coli lacZ* gene by gene replacement. This approach has been utilized previously but the high frequency of DNA amplifications in the *R. etli* symbiotic plasmid generated a second copy of the *nifH-lacZ* fusion by homologous recombination, hampering the accurate determination of their expression (39,49). Strains containing either the *nifHa-lacZ* (DEM153) or the *nifHc-lacZ* (DEM233) gene fusions were constructed and their free-living expression estimated under different oxygen concentrations. Strains DEM153 and DEM233 were cultured in liquid minimal medium under 1, 5 and 20% oxygen concentrations (Table 2). For both strains, maximal  $\beta$ -galactosidase activities were always observed when grown under 1% oxygen. Activities declined as the oxygen concentration was increased. Interestingly, at the maximal induction point, strain DEM233 presented 2.3 fold higher  $\beta$ -galactosidase specific activity than strain DEM153. This oxygen-sensitive expression suggested a positive regulation of the reiterated *nifH* genes by NifA as observed for other *nifH* genes (19).

To determine the role of NifA in the microaerobic induction, *nifHa-lacZ* and *nifHc-lacZ* gene fusions were also introduced into the *nifA* mutant derivative CFNX247 to generate strains DEM153-1 and DEM233-1, respectively. Both strains failed to induce significant  $\beta$ -galactosidase activity even when grown under 1% oxygen (Table 2). This result indicates that NifA is required for the free living induction of the *nifH* promoters tested. In support of this, expression of the *R. meliloti nifH* gene, whose activation has been shown to be dependent on NifA (5,17), was also triggered under low oxygen concentrations in *R. etli* in a NifA-dependent manner (Table 2)

**Symbiotic phenotype of the *nifH-lacZ* fusion strains.** After thorough verification of the genomic stability of our fusion strains (not shown), we evaluated the symbiotic transcriptional activity of the *nifHa* and *nifHc* promoters. *In planta*  $\beta$ -galactosidase

activity was measured in extracts of root nodules collected 13 and 23 days after inoculation (Fig. 1). As observed in free living cultures,  $\beta$ -galactosidase activity of nodules induced by strain DEM233 was four fold higher than those induced by strain DEM153. Symbiotic effectiveness was also evaluated and, as previously reported (39,50), disruption of *nifHa* diminished nitrogenase specific activity approximately to 50%, by abating the expression of the transcriptionally linked *nifDKa* genes, while disruption of *nifHc*, which does not form part of a *nifHDK* operon, had no effect (data not shown).

**Heterologous expression analysis of the *nifH-lacZ* fusions.** To further analyze the role of the putative *cis*- and *trans*-acting elements that control the reiterated *nifH* gene expression, we made use of an heterologous system in *E. coli*. We evaluated the role of NifA in the expression of the different *nifH-lacZ* fusions by estimating the  $\beta$ -galactosidase activity of strains carrying a plasmid that harbors the *K. pneumoniae nifA* gene. All the four fusions tested were expressed only in the presence of *nifA* (Table 3, columns I and II). The expression of *nifHa* was at a similar level than the expression of the *K. pneumoniae* and *R. meliloti nifH*. Interestingly, *nifHc* was expressed at a lower level than *nifHa*, contrasting to the two fold ratio obtained in *R. etli*. A plausible explanation to this observation is discussed later.

Sequences highly resembling  $\sigma^{54}$ -dependent promoters (5'-TGGCACGGGTTTTGAA-3') are located upstream of the three reiterated *nifH* genes (47). In order to evaluate the requirement of  $\sigma^{54}$  for the expression of the *nifHa*- and *nifHc-lacZ* gene fusions, we determined their  $\beta$ -galactosidase activities in strain ET8045 (*rpoN* mutant derivative of ET8000) harboring plasmids pDEM153 or pDEM233 respectively. These strains failed to induce  $\beta$ -galactosidase activity even in the presence of the *nifA* gene, showing that  $\sigma^{54}$  is essential for the expression of both *nifHa* and *nifHc* and supporting the functionality of the putative promoter sequences (Table 3, column V).

It has been shown that certain  $\sigma^{54}$ -dependent promoters that highly resemble the consensus sequence, as the *R. meliloti nifH* promoter, are prone to be partially activated by other proteins of the NifA family, such as NtrC (30,33,55). This results from the strong binding of  $E\sigma^{54}$  to the promoter. Nucleotide sequences upstream of the reiterated *nifH* genes differ from the consensus  $\sigma^{54}$ -dependent promoters only at the conserved -12 position, where they have an A instead of a C. We measured the proficiency of these promoters to be cross-activated by NtrC in strain ET8000 by growing it on glutamine (where NtrC is active) or on

ammonium chloride (where NtrC is inactive). In contrast to the *R. meliloti nifH* promoter (that was cross-activated by NtrC in the absence of NifA, Table 3, column IV) the  $\beta$ -galactosidase activities of strains carrying the *nifHa* and *nifHc* promoters were negligible and very similar in both conditions (Table 3, columns I and IV). As previously shown, the same result was obtained for the strain carrying the weak *K. pneumoniae nifH* promoter (7,9). Thus it seems that the *R. etli nifH* promoters are weak  $E\sigma^{54}$  binding sites, despite having a sequence highly resembling the consensus.

**Influence of the DNA binding function of NifA on the activation of the *nifH* genes.** The specificity in the activation observed for both *R. etli nifH* genes suggests that NifA must act in *cis*, interacting with the DNA in the vicinity of the *nifH* promoters. The *nifHa* gene is preceded by a consensus UAS (47), however, and in spite of being highly expressed, no canonical NifA binding sites have been detected up to 200 nucleotides upstream of *nifHc* (47). To assess the dependence of the *R. etli nifH* genes on the DNA binding function of NifA, we made use of a *K. pneumoniae* NifA mutant protein that carries a single amino acid substitution in the recognition helix of the helix-turn-helix motif (Tyr512-Phe) (36). This mutation abolishes the ability of NifA to bind to DNA, but it does not affect its positive control properties (36). As shown in Table 3 (column III), the mutant *K. pneumoniae nifA* gene failed to induce the expression of the *nifHa* and *nifHc* promoters. A similar result was obtained for the weak *K. pneumoniae nifH* promoter, while the *R. meliloti nifH* promoter, which has been shown to be less dependent on UAS, was partially activated by this DNA-binding defective protein.

The expression results shown above strongly suggested the presence of NifA binding sites upstream of *nifHc*. In the search for these predicted sites, we extended the nucleotide sequence up to 500 nucleotides upstream of this gene. As shown in Fig. 2A, two elements resembling UAS were identified at positions 85 (5'-TGG-N<sub>10</sub>-ACA-3') and 167 (5'-TGT-N<sub>10</sub>-ACG-3') upstream from the promoter. The former resulted from the correction of the previously reported sequence (47).

**In vivo footprinting analysis of the *nifH* upstream promoter regions.** The resemblance of the elements located upstream of *nifHa* and *nifHc* to UAS suggested their function as NifA binding sites. In order to identify residues contacted by this protein in the regulatory regions of both genes, we performed *in vivo* methylation protection experiments with dimethyl sulfate (DMS). Figure 3 shows the pattern of cleavage

products obtained from these regulatory regions after exposure of ET8000 cells harboring plasmids pDEM153 or pDEM233 to DMS, both in the presence or absence of plasmid pMJ220. In the presence of NifA, residues 95, 96, 98 and 99 upstream from the *nifHc* promoter were protected from methylation. These residues constitute the promoter-proximal putative NifA binding site depicted in Fig. 2A. No consistent protection was observed at any other position, not even around the putative promoter-distal UAS (data not shown). In the regulatory region of *nifHa*, residues 100, 104, and 113 upstream from the promoter were protected in the presence of NifA (Fig. 2B). The first two positions correspond to the putative canonical site, while the latter is within a NifA binding half-site (TGT), located just 6 nucleotides upstream (Fig. 2B).

**Influence of the Integration Host Factor in the expression of the reiterated *nifH* genes.** It is well known that IHF has a major role in the activation of several  $\sigma^{54}$ -dependent promoters, including those of *K. pneumoniae* and *R. meliloti nifH* (11,25,52). A computer aided search for IHF binding sites upon the regulatory regions of both *R. etli nifH*, based on the algorithm of Goodrich *et al.* (22) identified putative sites between the UAS and the promoter at 15 and 18 nucleotides upstream from the *nifHc* and *nifHa* promoters, respectively (Fig. 2).

In order to evaluate the role of IHF in the expression of the reiterated *nifH* genes, we determined the  $\beta$ -galactosidase activity of strain SE100 (*himD* mutant derivative of ET8000) carrying the different *nifH-lacZ* gene fusions. In spite of the pleiotropic effect of the *himD* mutation, the  $\beta$ -galactosidase activities observed in the absence of *nifA* were equivalent to the values obtained in the isogenic ET8000 wild type strain (Table 3, columns I and VI) When *nifA* was present in the cells, induction of the *nifH-lacZ* fusions expression was observed, however this was lower than in the wild type strain (Table 3, columns II and VIII), being *nifHa* much more dependent on IHF. As previously shown, *K. pneumoniae nifH* was highly dependent on IHF, while *R. meliloti nif* was only partially dependent (52). These results suggest that the NifA-dependent transcription of the reiterated *nifH* promoters is differently stimulated by IHF, despite having the same sequence that highly resembles the consensus. It is noteworthy that this is the only condition in *E. coli* where *nifHc* was expressed at higher levels than *nifHa*, as observed in *R. etli* (Table 2).

It has been shown that the sharp bend induced by IHF on the promoter region may restrict access of non-bound activators to the  $E\sigma^{54}$ -promoter complex (43). The absence of cross-activation by NtrC observed in the

*nifHa* and *nifHc* promoters (Table 3, column IV) could be due either to the lack of a strong E $\sigma$ 54-binding site or to a hindrance produced by IHF. As shown in Table 3 (column VII), neither *nifHa* nor *nifHc* were cross-activated by NtrC in the absence of IHF. In contrast, expression of the strong *R. meliloti nifH* promoter was increased about three fold in the absence of IHF. These observations support our proposition that the *nifHa* and *nifH* promoters are low affinity E $\sigma$ 54 binding sites.

**Nucleotide sequence downstream of *nifHc*.** The *nifHc* gene is actively expressed during symbiosis with bean plants (Fig. 1) and in cultures under low oxygen concentrations (Table 2). Unlike *nifHa* and *nifHb*, *nifHc* seemed not to form part of a *nifHDK* operon (47). Further sequencing of the downstream region of *nifHc* allowed us to identify two open reading frames: ORF70 and ORF251. Interestingly, the deduced amino acid sequence of ORF70 bears high similarity (up to 91%) to the first 70 residues of the nitrogenase NifD protein from several nitrogen fixing bacteria (data not shown), thus this ORF is a truncated *nifD* gene of unknown function. The deduced amino acid sequence of ORF251 is 56 to 58% similar to the oxygen-independent coproporphyrinogen III oxidase protein from *Salmonella thyphimurium* (60), *Rhodobacter sphaeroides* (14) and *E. coli* (57), coded by the *hemN* gene. A sequence resembling a  $\sigma^E$ -dependent promoter is located upstream of ORF251. No expression of ORF251-*lacZ* fusions was detected in culture under any oxygen concentration tested, but it was actively transcribed during nodule development (data not shown). The nucleotide sequence determination downstream of *nifHc* allowed us to show that the identity between the *nifHc* and the *nifHb* regions extends throughout the promoter until 13 nucleotides before the ORF70 translation stop codon.

## DISCUSSION

In this work we characterized the complex *cis*- and *trans*-acting regulatory elements involved in the reiterated *R. etli nifH* gene expression. Strain CE3 contains three *nifH* genes, two of which are functional *nifHDK* operons separated by 120 kb in the symbiotic plasmid, while *nifHc* is located in between (21) and seems to be unlinked to other *nif* genes (32,47). Here we show that *nifHc* is transcriptionally coupled to a truncated *nifD* homologue comprising the first 70 amino acids of the protein. Directly downstream, we identified ORF251 which is similar to several *hemN* genes. Restriction fragment length polymorphism (RFLP) analyses of other distantly related isolates of *R.*

*etli*, previously classified and characterized (32,44), suggests that this arrangement is general in this species (not shown). Although the truncated *nifD* and the *hemN* homologues were readily transcribed in nodules, we do not have any direct evidence for their functionality. As has been shown for other multimeric enzymes as the hepatocyte growth factor (12) and the Fos/Jun transcriptional factor (40), the incorporation of truncated, non-functional monomers instead of the full length polypeptides can regulate their activity. The first 70 residues of the NifD protein do not form an independent domain, do not interact with the NifK protein nor with any of the metal cofactors (27). Consequently it is unlikely that the truncated NifD protein could be involved in the regulation of the nitrogenase activity by this mechanism. Homologues to ORF251 have been characterized in *E. coli*, *S. thyphimurium* and *R. sphaeroides* (14,57,60), these *hemN* genes code for the oxygen-independent coproporphyrinogen III oxidase, an enzyme involved in the alternative biosynthesis of the heme cofactor under anaerobic conditions. In contrast to *hemA* mutations (16), gene disruption in ORF251 had no symbiotic effect (not shown).

The different nucleotide sequences of the *nifHa* and *nifHc* upstream regulatory regions and the asymmetric arrangement of their NifA binding sites suggested a differential transcription of these genes. As it is well known, low oxygen concentration is the most important effector in the induction of *nifH*, so we estimated the expression of the reiterated *R. etli nifH* genes under different oxygen concentrations. For both genes, maximal induction was observed under 1% oxygen concentration, being *nifHc* transcribed at higher levels than *nifHa* (Table 2). By using a *nifA* mutant derivative of strain CE3 we demonstrated that expression of the reiterated *nifH* genes, as reported for *R. leguminosarum* bv phaseoli (34), was dependent on NifA in an oxygen concentration responsive manner. Evidences obtained by two different approaches show that the activation is mediated by NifA bound to UAS elements: (i) induction of both *nifH* genes was strictly dependent on the DNA binding ability of NifA, since a derivative impaired in its specific binding to the UAS was unable to activate transcription (Table 3, column III), and (ii) *in vivo* methylation protection of residues located within putative UAS was dependent on the presence NifA (Figs. 2 and 3). NifA exerts its activation properties by binding to a minimal DNA sequence located around 100 residues upstream from the promoter and having the sequence 5'-TGT-N<sub>10</sub>-ACA-3' (36). Here we show that the NifA-dependent activation of the *nifHc* promoter is mediated by the specific

binding of NifA to a divergent UAS located 85 residues upstream from the promoter (Figs. 2A and 3A). Fully functional divergent NifA-UAS have been also described upstream the *K. pneumoniae nifJ* (10) and the *B. japonicum fixRnifA* promoters (3). The protection pattern observed upstream *nifHc* resembled the one observed at the *K. pneumoniae nifH* UAS (36), where the G of the TGT conserved motif and the internal G residues at positions 5 and 6 were protected, suggesting a similar mode of contacting the DNA (Figs. 2A and 3A). At the *nifHa* regulatory region, in addition to protection of the conserved G of the TGT motif, the G residue at position 6 and a G residue located 8 nucleotides upstream from the UAS were also protected (Figs. 2B and 3B). The latter residue forms part of a TGT UAS half-site. The same pattern was observed by Cannon *et al.* (10) in the upstream region of the *K. pneumoniae nifE* gene, where a canonical UAS and an additional half site exactly 6 nucleotides upstream, were protected by NifA. Protection of this half-site could be due either to the binding of an additional NifA molecule or to the DNA bending induced by NifA, allowing the establishment of an additional contact with the DNA. This half site could help to increase the number of NifA molecules in the vicinity of the *nifHa* promoter, raising the probability of productive interactions with  $E\sigma^{54}$ . In support of this enhancing role, mutations in the UAS half-site diminished the NifA-dependent activation of the *K. pneumoniae nifE* promoter (10). Protection of a TGT UAS half-site in the bottom strand has also been detected in the *K. pneumoniae nifU* regulatory region (11)

The absolute dependence on NifA for expression of the reiterated *nifH* genes was also demonstrated in *E. coli*, by the use of a plasmid harboring the *K. pneumoniae nifA* gene (Table 3). Under this conditions, *nifHa* was transcribed at higher levels than *nifHc*, contrasting to the ratio observed in *R. etli* (Table 2). Although a reproducible NifA-dependent protection pattern was observed in the divergent UAS located upstream *nifHc* under saturation conditions (see Fig. 3A), we cannot rule out a lower affinity of *K. pneumoniae* NifA to this site.

The reiterated *nifH* genes are preceded by identical sequences highly resembling consensus  $\sigma^{54}$ -dependent promoters (47). Here we presented evidences on their functionality by showing the requirement of  $\sigma^{54}$  for their expression (Table 3, column V). These promoter sequences differ from the consensus only at position -12, having an A instead of a C. This position seems to be relevant for the promoter recognition by  $E\sigma^{54}$ , since it is highly conserved and an equivalent

transversion in the *K. pneumoniae nifH* promoter abolishes the promoter capacity to be cross-activated by NtrC (9). Most of the  $\sigma^{54}$ -dependent promoters described so far have the conserved C in position -12, but some exceptions are known among Rhizobia: in the *nifN* gene of *R. meliloti* (1) and in the *nifH* genes of *R. leguminosarum* (48), *R. sp.* ANU240 (2), *R. leguminosarum* bv trifolii (59) and *R. leguminosarum* bv phaseoli (34). It is interesting to note that in the latter three species the *nifH* genes are reiterated as in *R. etli*.

If a given promoter, as *K. pneumoniae nifH*, is a weak  $E\sigma^{54}$  binding site, it is more dependent on IHF and UAS and consequently it is less prone to cross-activation (7,9,37,52). Conversely, promoters that match to the consensus, as in *R. meliloti nifH*, are strongly bound by  $E\sigma^{54}$  and are less dependent on IHF and UAS, but are readily cross-activated (7,26,37,52,56). Although IHF has not been characterized in *Rhizobium*, an homologous activity is likely to be present, since the stimulation of expression by IHF in *E. coli* has been specifically demonstrated for several *nif* genes (11,52). As shown above, expression of the reiterated *nifH* genes promoters is dependent on the upstream binding of NifA and they are not cross-activated by NtrC, despite having a sequence that highly resembles the consensus. Therefore, they probably bind  $E\sigma^{54}$  with low affinity and their expression is likely to be dependent on IHF. Results in *E. coli* showed this dependence and also that the requirement for this protein differs, with the *nifHa* promoter being more dependent than the *nifHc* (Table 3, column VII). This difference could be the result of the relative orientation of all three binding sites which has been shown to be critical for the expression of other  $\sigma^{54}$ -dependent promoters as *K. pneumoniae nifH* (35) and *E. coli glnH* (13). Interestingly, the UAS of *nifHc* is located at 8 helical turns from the promoter, while at the *nifHa* it is located at 8.5 turns. Furthermore, the partial independence of the *nifHc* promoter on IHF could also be the result of an intrinsically bent DNA structure between the UAS and the promoter that brings bound NifA and  $E\sigma^{54}$  in close contact. This structure is either absent in the *nifHa* regulatory region or given the half-a-turn different phasing of the UAS does not allow a productive contact of the bound proteins.

In having reiterated genes an organism gains additional flexibility in the control of gene expression. Two identical genes can be differentially expressed in response to a changing environment or can be simultaneously transcribed, to ensure functional concentrations of their product. In *A. caulinodans* two

*nifH* genes are alternatively expressed under free living nitrogen fixation conditions and in symbiosis with *Sesbania rostrata* (42). In bacteria of the *Azotobacter* genus, three different sets of *nifHDK* gene homologues are expressed in response to the metal availability (58). In *R. etli*, it has been shown that both *nifHDK* operons are required to achieve full nitrogenase activity during symbiosis with *Phaseolus vulgaris* (39,50), while in this study we show that *nifHa* and *nifHc* are differentially expressed. If a promoter is selectively modified to narrow its activation specificity range and avoid cross-activation, then a decrease in its transcription rate is likely to occur. Duplication of the gene would be an advantageous alternative to restore the original product level without losing the stringent control on its expression. Then, the asymmetric arrangement of the *cis*-acting factors described here,  $\sigma$ 54-dependent promoters, IHF binding sites and NifA-UAS, contribute to the differential expression of the reiterated *nifH* genes, by creating a finely regulated architecture that assures optimal levels of the nitrogenase enzyme in *R. etli*.

#### ACKNOWLEDGEMENTS

We thank Drs. Lorenzo Segovia and Michael Dunn their comments on the manuscript and Dr. Tracy Nixon for kindly providing the Seqscan computer program prior to publication. This work was supported by Dirección General de Asuntos del Personal Académico from Universidad Nacional Autónoma de México (IN 204489). BV was recipient of a scholarship from DGAPA.

#### REFERENCES

1. Aguilar, O.M., H. Reiländer, W. Arnold and A. Pühler. 1987. *Rhizobium meliloti nifN (fixF)* gene is part of an operon regulated by a *nifA*-dependent promoter and codes for a polypeptide homologous to the *nifK* gene product. *J. Bacteriol.* 169:5393-5400.
2. Badenoch-Jones, J., T.A. Holton, C.M. Morrison, K.F. Scott and J. Shine. 1989. Structural and functional analysis of nitrogenase genes from the broad-host-range *Rhizobium* strain ANU240. *Gene* 77:141-153.
3. Barrios, H. and E. Morett. Submitted.
4. Berger, D.K., F. Narberhaus and S. Kustu. 1994. The isolated catalytic domain of NIFA, a bacterial enhancer-binding protein, activates transcription *in vitro*: Activation is inhibited by NIFL. *Proc. Natl. Acad. Sci. USA.* 91:103-107.
5. Better, M., G. Ditta and D.R. Hellinski. 1985. Deletion analysis of *Rhizobium meliloti* symbiotic promoters. *EMBO J.* 4:2419-2424.
6. Buck, M. and W. Cannon. 1987. Frameshifts close to the *Klebsiella pneumoniae nifH* promoter prevent multicopy inhibition by hybrid *nifH* plasmids. *Mol. Gen. Genet.* 207:492-498.
7. Buck, M. and W. Cannon. 1989. Mutations in the RNA polymerase recognition sequence of the *Klebsiella pneumoniae nifH* promoter permitting transcriptional activation in the absence of NifA binding to upstream activator sequences. *Nucl. Acids Res.* 17:2597-2612.
8. Buck, M., W. Cannon and J. Woodcock. 1987. Transcriptional activation of the *Klebsiella pneumoniae* nitrogenase promoter may involve DNA loop formation. *Mol. Microbiol.* 1:243-249.
9. Buck, M., H. Khan and R. Dixon. 1985. Site-directed mutagenesis of the *Klebsiella pneumoniae nifL* and *nifH* promoters and *in vivo* analysis of promoter activity. *Nucl. Acids Res.* 13:7621-7638.
10. Cannon, W., W. Charlton and M. Buck. 1991. Organization and function of binding sites for the transcriptional activator NifA in the *Klebsiella pneumoniae nifE* and *nifU* promoters. *J. Mol. Biol.* 220:915-931.
11. Cannon, W.V., R. Kreutzer, H.M. Kent, E. Morett and M. Buck. 1990. Activation of the *Klebsiella pneumoniae nifU* promoter: identification of multiple and overlapping upstream NifA binding sites. *Nucl. Acids Res.* 18:1693-1701.
12. Chan, A.M.L., J.S. Rubin, D.P. Bottaro, D.W. Hirschfeld, M. Chedid and S.A. Aaronson. 1991. Identification of a competitive HGF antagonist encoded by an alternative transcript. *Science* 254:1382-1385.
13. Claverie-Martin, F. and B. Magasanik. 1992. Positive and negative effects of DNA bending on activation of transcription from a distant site. *J. Mol. Biol.* 227:996-1008.
14. Coomber, S.A., R.M. Jones, P.M. Jordan and C.N. Hunter. 1992. A putative anaerobic coproporphyrinogen III oxidase in *Rhodobacter sphaeroides*. I. Molecular cloning, transposon mutagenesis and sequence analysis of the gene. *Mol. Microbiol.* 6:3159-3169.
15. Devereux, J., P. Haeblerli and O. Smithies. 1984. A comprehensive set of analysis programs for the VAX. *Nucl. Acids Res.* 12:387-395.
16. Dickstein, R., D.C. Scheirer, W.H. Fowle and F.M. Ausubel. 1991. Nodules elicited by *Rhizobium meliloti* heine mutants are arrested at an early stage of development. *Mol. Gen. Genet.* 230:423-432.
17. Ditta, G., E. Virts, A. Palomares and K. Choong-Hyun. 1987. The *nifA* gene of *Rhizobium meliloti* is oxygen regulated. *J. Bacteriol.* 169:3217-3223.
18. Encarnación, S., M. Dunn, K. Willms and J. Mora. 1995. Fermentative and aerobic metabolism in *Rhizobium etli*. *J. Bacteriol.* 177:3058-3066.
19. Fischer, H.M. 1994. Genetic regulation of nitrogen fixation in Rhizobia. *Microbiol. Revs.* 58:352-386.
20. Girard, L. Unpublished.
21. Girard, M. de L., M. Flores, S. Brom, D. Romero, R. Palacios and G. Dávila. 1991. Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. phaseoli. *J. Bacteriol.* 173:2411-2419.
22. Goodrich, J.A., M.L. Schwartz and W.R. McClure. 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). *Nucl. Acids Res.* 18:4993-5000.
23. Gubler, M. 1989. Fine-tuning of *nif* and *fix* gene expression by upstream activator sequences in *Bradyrhizobium japonicum*. *Mol. Microbiol.* 3:149-159.
24. Hahn, M. and H. Hennecke. 1984. Localized mutagenesis in *Rhizobium japonicum*. *Mol. Gen. Genet.* 193:46-52.
25. Hoover, T.R., E. Santero, S. Porter and S. Kustu. 1990. The integration host factor stimulates interaction of RNA polymerase with NifA, the

- transcriptional activator for nitrogen fixation operons  
Cell 63:11-22.
26. Huala, E. and F.M. Ausubel. 1989. The central domain of *Rhizobium meliloti* NifA is sufficient to activate transcription from the *R. meliloti* *nifH* promoter. J. Bacteriol. 171:3354-3365.
27. Kim, J., D. Woo and D.C. Rees. 1993. X-ray crystal structure of the nitrogenase molybdenum-iron protein from *Clostridium pasteurianum* at 3.0-Å resolution. Biochemistry 32:7104-7115.
28. Kokotek, W. and W. Lotz. 1989. Construction of a *lacZ*-kanamycin-resistance cassette, useful for site-directed mutagenesis and as a promoter probe. Gene 84:467-471.
29. Kustu, S., E. Santero, J. Keener, D. Popham and D. Weiss. 1989. Expression of  $\sigma^{54}$  (*nifA*)-dependent genes is probably united by a common mechanism. Microbiol. Revs. 53:367-376.
30. Labes, M., V. Rastogi, R. Watson and T.M. Finan. 1993. Symbiotic nitrogen fixation by a *nifA* deletion mutant of *Rhizobium meliloti*: the role of an unusual *nifC* allele. J. Bacteriol. 175:2662-2673.
31. Macneil, T., G.P. Roberts, D. Macneil and B. Tyler. 1982. The products of *glnL* and *glnG* are bifunctional regulatory proteins. Mol. Gen. Genet. 188:325-333.
32. Martínez, E., M.A. Pardo, R. Palacios and M.A. Cevallos. 1985. Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. J. Gen. Microbiol. 131:1779-1786.
33. Merrick, M.J. 1983. Nitrogen control of the *nif* regulon in *Klebsiella pneumoniae*: involvement of the *nifA* gene and analogies between *nifC* and *nifA*. EMBO J. 2:39-44.
34. Michiels, J., I. D'hooghe, C. Verreth, H. Pelemans and J. Vanderleyden. 1994. Characterization of the *Rhizobium leguminosarum* biovar *phaseoli* *nifA* gene, a positive regulator of *nif* gene expression. Arch. Microbiol. 161:404-408.
35. Molina-López, J.A., F. Govantes and E. Santero. 1994. Geometry of the process of transcription activation at the  $\sigma^{54}$ -dependent *nifH* promoter of *Klebsiella pneumoniae*. J. Biol. Chem. 269:25419-25425.
36. Morett, E. and M. Buck. 1988. *nifA*-dependent in vivo protection demonstrates that the upstream activator sequence of *nif* promoters is a protein binding site. Proc. Natl. Acad. Sci. USA. 85:9401-9405.
37. Morett, E. and M. Buck. 1989. In vivo studies on the interaction of RNA polymerase- $\sigma^{54}$  with the *Klebsiella pneumoniae* and *Rhizobium meliloti* *nifH* promoters. J. Mol. Biol. 210:65-77.
38. Morett, E., W. Cannon and M. Buck. 1988. The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognizes the upstream activator sequences of *nif* promoters and can be separated from the positive control function of NifA. Nucl. Acids Res. 16:11469-11488.
39. Morett, E., S. Moreno and G. Espín. 1988. Transcription analysis of the three *nifH* genes of *Rhizobium phaseoli* with gene fusions. Mol. Gen. Genet. 213:499-504.
40. Nakabeppu, Y. and D. Nathans. 1991. A naturally occurring truncated form of FosB that inhibits Fos/Jun transcriptional activity. Cell 64:751-759.
41. Noel, K.D., A. Sánchez, L. Fernández, J. Leemans and M.A. Cevallos. 1984. *Rhizobium phaseoli* symbiotic mutants with transposon Tn5 insertions. J. Bacteriol. 158:148-155.
42. Norel, F. and C. Elmerich. 1987. Nucleotide sequence and functional analysis of the two *nifH* copies of *Rhizobium* ORS571. J. Gen. Microbiol. 133:1563-1576.
43. Pérez-Martín, J. and V. de Lorenzo. 1995. Integration host factor suppresses promiscuous activation of the  $\sigma^{54}$ -dependent promoter Pu of *Pseudomonas putida*. Proc. Natl. Acad. Sci. USA. 92:7277-7281.
44. Piñero, D., E. Martínez and R.K. Selander. 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. Appl. Environ. Microbiol. 54:2825-2832.
45. Prentki, P. and H.M. Krisch. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29:303-313.
46. Quinto, C., H. de la Vega, M. Flores, L. Fernández, T. Ballado, G. Soberón and R. Palacios. 1982. Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. Nature 299:724-726.
47. Quinto, C., H. de la Vega, M. Flores, J. Leemans, M.A. Cevallos, M.A. Pardo, R. Azpiroz, M.deL. Girard, E. Calva and R. Palacios. 1985. Nitrogenase reductase: A functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. USA. 82:1170-1174.
48. Roelvinck, P.W., M. Harmsen, A. van Kammen and R.C. van den Bus. 1990. The *nifH* promoter region of *Rhizobium leguminosarum*: nucleotide sequence and promoter elements controlling activation by NifA protein. Gene 87:31-36.
49. Romero, D., S. Brom, J. Martínez-Sulazar, M.L. Girard, R. Palacios and G. Dávila. 1991. Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *Rhizobium phaseoli*. J. Bacteriol. 173:2435-2441.
50. Romero, D., P.W. Singleton, I. Segovia, E. Morett, B.B. Bohlool, R. Palacios and G. Dávila. 1988. Effect of naturally occurring *nif* reiterations on symbiotic effectiveness in *Rhizobium phaseoli*. Appl. Environ. Microbiol. 54:848-850.
51. Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York.
52. Santero, E., T.R. Hoover, A.K. North, D.K. Berger, S.C. Porter and S. Kustu. 1992. Role of integration host factor in stimulating transcription from the  $\sigma^{54}$ -dependent *nifH* promoter. J. Mol. Biol. 227:602-620.
53. Segovia, I., J.P.W. Young and E. Martínez-Romero. 1993. Reclassification of american *Rhizobium leguminosarum* biovar *phaseoli* type 1 strains as *Rhizobium etli* sp. nov. Int. J. Syst. Bacteriol. 43:374-377.
54. Simon, R., U. Priefer and A. Pühler. 1983. Vector plasmids for in-vivo and in-vitro manipulations of Gram-negative bacteria, p.98. In A. Pühler (ed.), Molecular genetics of the Bacteria-Plant interaction, Springer-Verlag, Berlin.
55. Sundaresan, V., D.W. Ow and F.M. Ausubel. 1983. Activation of *Klebsiella pneumoniae* and *Rhizobium meliloti* nitrogenase promoters by *gln* (*nif*) regulatory proteins. Proc. Natl. Acad. Sci. USA. 80:4030-4034.
56. Szeto, W.W., B.T. Nixon, C.W. Ronson and F.M. Ausubel. 1987. Identification and characterization of the *Rhizobium meliloti* *nifC* gene: *R. meliloti* has separate regulatory pathways for activation of nitrogen fixation genes in free-living and symbiotic cells. J. Bacteriol. 169:1423-1432.
57. Troup, B., C. Hungerer and D. Jahn. 1995. Cloning and characterization of the *Escherichia coli* *hemN* gene encoding the oxygen-independent coproporphyrinogen III oxidase. J. Bacteriol.

177:3326-3331.

58. **Walsley, J., A. Toukdarian and C. Kennedy.** 1994. The role of regulatory genes *nifA*, *vnfA*, *anfA*, *nfrX*, *ntrC*, and *rpoN* in expression of genes encoding the three nitrogenases of *Azotobacter vinelandii*. *Arch. Microbiol.* **162**:422-429.

59. **Watson, J.M. and P.R. Schofield.** 1985. Species-specific, symbiotic plasmid-located repeated DNA

sequences in *Rhizobium trifolii*. *Mol. Gen. Genet.* **199**:279-289.

60. **Xu, K., J. Delling and T. Elliott.** 1992. The genes required for heme synthesis in *Salmonella typhimurium* include those encoding alternative functions for aerobic and anaerobic coproporphyrinogen oxidation. *J. Bacteriol.* **174**:3953-3963.

## LEGENDS TO FIGURES

TABLE 1. Bacterial strains and plasmids used in this study.

TABLE 2. Expression of *nifH* genes in *R. etli* under different oxygen concentration.

$\beta$ -galactosidase activities of different *nifH-lacZ* fusions in *R. etli* wild type and a *nifA* mutant derivative under different oxygen concentrations. Values are mean  $\beta$ -galactosidase specific activity units from two different cultures in duplicate.

TABLE 3. Heterologous expression in *E. coli* of *nifH* genes from *R. etli*, *R. meliloti* and *K. pneumoniae*.

ET8000, ET8045 and SE100 are isogenic *lacZ::IS1* *E. coli* strains. They carry different *nifH-lacZ* fusions which expression is activated by the *K. pneumoniae* NifA proteins contained either in the pMJ220 (wild type), the pMJ221 (unable to bind DNA derivative) or the pMJ220tc (pMJ220 but tetracycline resistant) plasmids. ET8045 and SE100 are derivatives of ET800 carrying *rpoN::tc* and *himD::cm* insertions respectively. For induction conditions see Materials and Methods.

Cultures were grown in minimal medium with ammonium chloride (1) or glutamine (2) as nitrogen sources. Values are mean  $\beta$ -galactosidase Miller Units from at least two independent cultures.

<sup>a</sup> Abbreviations: Re, *R. etli*; Rm, *R. meliloti*; Kp, *K. pneumoniae*.

<sup>b</sup> Regulatory proteins present: NtrC, encoded by the chromosomal *ntrC* gene; NifA, encoded in the pMJ220 plasmid; NifAY512F, encoded in the pMJ221 plasmid; IHF, encoded by the chromosomal *himA* and *himD* genes;  $\sigma$ 54, encoded by the chromosomal *rpoN* gene.

<sup>c</sup> NT, not tested.

FIGURE 1. Symbiotic expression of the *R. etli nifHa* and *nifHc* genes.

Root nodules from *P. vulgaris* inoculated with *R. etli* strains CE3 , DEM233  or DEM153 , were assayed for  $\beta$ -galactosidase activities as described in Materials and Methods. Values are mean of 10 different plants.

FIGURE 2. Nucleotide sequence of *nifHc* (A) and *nifHa* (B) upstream regions. UAS and promoter sequences are indicated by open boxes. Shaded boxes indicate the putative translation initiation codon. Putative IHF binding sites are underlined. Arrows indicate differences with the nucleotide sequences previously reported by Quinto *et al.* (47). Residues protected from cleavage in NifA-dependent *in vivo* methylation protection analyses are marked by dots.

FIGURE 3. NifA-dependent methylation protection of the *nifHc* (A) and *nifHa* (B) promoter elements. Methylation was performed *in vivo* in the presence (+) or absence (-) of NifA. The open circles indicate guanine residues protected when NifA was present in the cells

TABLE 1

Strain or plasmid	Relevant genotype and characteristics	Source or reference
<i>Rhizobium etli</i>		
CE3	Wild type CFN42, sm <sup>r</sup>	(41)
CFNX247	CE3, <i>nifAA::sp/sm</i>	(20)
DEM233	CE3, <i>nifHc-lacZ</i>	This study
DEM233-1	CFNX247, <i>nifHc-lacZ</i>	This study
DEM153	CE3, <i>nifHa-lacZ</i>	This study
DEM153-1	CFNX247, <i>nifHa-lacZ</i>	This study
<i>Escherichia coli</i>		
ET8000	<i>rbs, gyrA, hut, lacZ::IS1 Mu, cts62</i>	(31)
ET8045	<i>rpoN208::tn10</i> derivative of ET8000	(31)
SE100	<i>himD::cm<sup>r</sup></i> derivative of ET8000	(35)
Plasmids		
pSUP205	Mobilizable derivative of pBR325 (tc <sup>r</sup> )	(54)
pKOK6	<i>lacZ-km<sup>r</sup></i> interposon in pKOK4 (km <sup>r</sup> , cb <sup>r</sup> )	(28)
pDEM233	Re <i>nifHc-lacZ-km<sup>r</sup></i> in pSUP205 (km <sup>r</sup> , tc <sup>r</sup> )	This study
pDEM153	Re <i>nifHa-lacZ-km<sup>r</sup></i> in pSUP205 (km <sup>r</sup> , tc <sup>r</sup> )	This study
pMB210	<i>R. meliloti nifH-lacZ</i> in pGC926 (tc <sup>r</sup> )	(5)
pMB2101	pMB210 but in pMC1403 (cb <sup>r</sup> )	(7)
pJMW6	<i>K. pneumoniae nifH-lacZ</i> in pJEL126 (cb <sup>r</sup> )	(6)
pMJ220	<i>K. pneumoniae nifA</i> expressed from <i>plac</i> in pEMBL8 (cm <sup>r</sup> )	(36)
pMJ221	pMJ220 but tyr512 to phe (cm <sup>r</sup> )	(36)
pMJ220tc	pMJ220 but tc <sup>r</sup>	This study

TABLE 2

Strain	Relevant genotype	Oxygen concentration (B-gal)			Induction ratio 1% / 20%
		1%	5%	20%	
CE3	wild type	10 ± 0.2	10 ± 0.9	10 ± 0.5	1.00
DEM153	<i>nifHa-lacZ</i>	141 ± 3	54 ± 7	25 ± 1	5.64
DEM 153-1	<i>nifHa-lacZ</i> <i>nifAΔ::sp</i>	50 ± 9	73 ± 1	35 ± 1	1.42
DEM233	<i>nifHc-lacZ</i>	327 ± 9	74 ± 8	45 ± 9	7.26
DEM233-1	<i>nifHc-lacZ</i> <i>nifAΔ::sp</i>	57 ± 4	76 ± 7	36 ± 6	1.58
CE3 / pMB210	Rm <i>nifH-lacZ</i>	553 ± 38	142 ± 49	134 ± 57	4.12
CFNX247 / pMB210	Rm <i>nifH-lacZ</i> <i>nifAΔ::sp</i>	122 ± 4	98 ± 4	77 ± 2	1.58

TABLE 3

$\beta$ -galactosidase activity (Miller Units)

		I	II	III	IV	V	VI	VII	VIII
		ET8000 <sup>(1)</sup>	ET8000/ <sup>(1)</sup> pMJ220	ET8000/ <sup>(1)</sup> pMJ221	ET8000 <sup>(2)</sup>	E18045/ <sup>(2)</sup> pMJ220	SE100 <sup>(1)</sup>	SE100 <sup>(2)</sup>	SE100/ <sup>(2)</sup> pMJ220tc
	NuC <sup>+</sup>	-	-	-	+	+	-	+	-
	NifA	-	+	-	-	+	-	-	+
	NifAY512F	-	-	+	-	-	-	-	-
	IHF	+	+	+	+	+	-	-	-
	$\sigma^{54}$	+	+	+	+	-	+	+	+
Plasmid	Genotype								
pDEM133	Re <i>nifH-lacZ</i>	22( $\pm$ 0.7)	16622( $\pm$ 2220)	52( $\pm$ 4)	30( $\pm$ 2)	21( $\pm$ 3)	40( $\pm$ 4)	37( $\pm$ 9)	624( $\pm$ 7)
pDEM233	Re <i>nifH-lacZ</i>	94( $\pm$ 10)	10713( $\pm$ 872)	166( $\pm$ 17)	57( $\pm$ 3)	101( $\pm$ 40)	203( $\pm$ 44)	159( $\pm$ 45)	3543( $\pm$ 746)
pJMW6	Kp <i>nifH-lacZ</i>	170( $\pm$ 6)	22361( $\pm$ 2281)	188( $\pm$ 17)	142( $\pm$ 31)	NT <sup>a</sup>	104( $\pm$ 39)	104( $\pm$ 63)	543( $\pm$ 142)
eMB2101	Rm <i>nifH-lacZ</i>	215( $\pm$ 48)	19199( $\pm$ 1679)	1317( $\pm$ 270)	1268( $\pm$ 121)	317( $\pm$ 126)	93( $\pm$ 32)	3624( $\pm$ 546)	3844( $\pm$ 754)

FIGURE 1

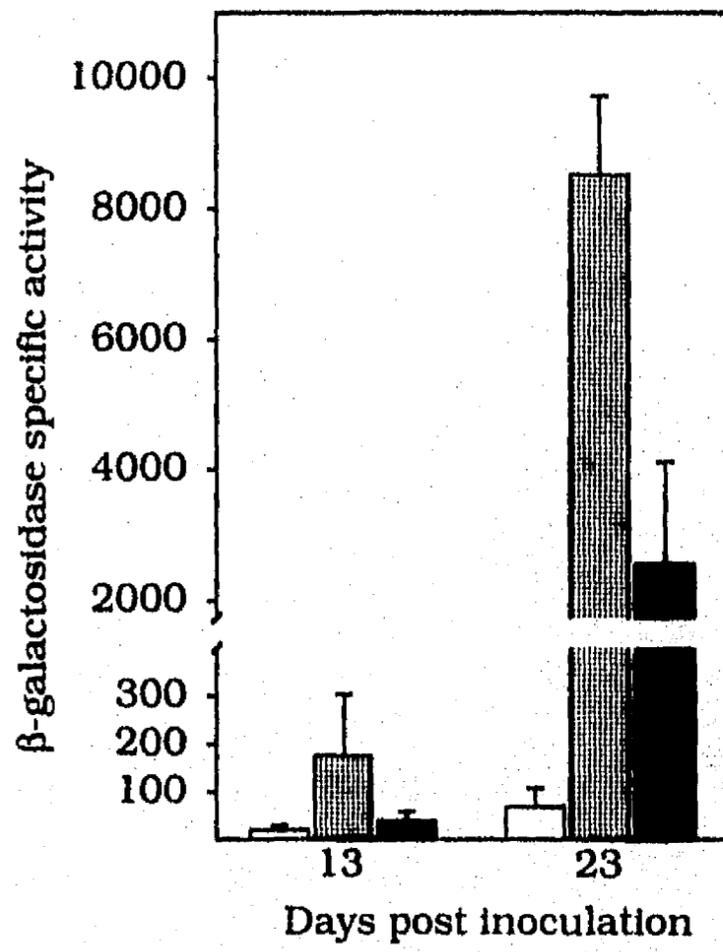


FIGURE 2

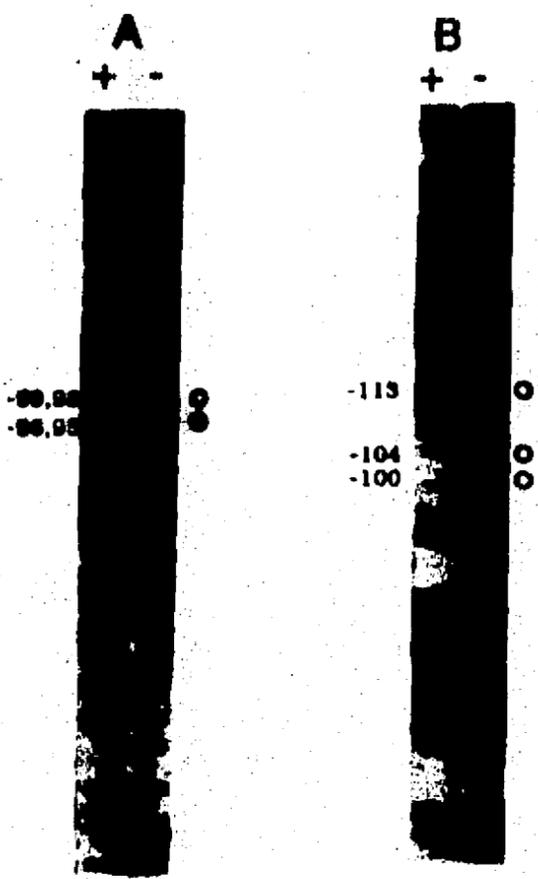
A

TTCNTCCTC GCTGAGAG AGGAGG GATTT ATTGATCTT GAGGCAAGG ATTCCAAC GAGGAGACT CCGAGGCGT TCGTCCAGT  
\* \* \*  
TTTGGTTTC GAG AGAG TTTCTGAC CCGCTTCA AAGTGGAA CTAATAGCT GAGATCAG AAACAGCTG TTTTATATG TCGTCCCAC  
↑ ↑ ↑ ↑ ↑ ↑ ↑  
ATGACAGG TTTTCAG AT TCGATCGA GCGGCGCA GCTGCCGC TTTTACTCAG CGATGGATG AAGAAAGAA GGAAGCGAT TCAGTT

B

ACCGGTCCTG TCGGCTT TTT CCGGCGCG CATTAAGTT TGTTTGGAG TTTCTGTG TGTTGGAG TAAGTTCTG AAACCAACA AAAGGCTCT  
\* \* \*  
TCTNTGCTT CGATCGGCG A ATGACAG TTTTCAG ATTGCCATG GAGGCGCG GAGCTGCTG CTTTACTC AGCGATGGT GGAAGGAAG  
AAGAAAGCG AT TCAGA

FIGURE 3



*Perspectivas en la aplicación de este trabajo*

La expresión controlada de algunos genes clave permite, en principio, modificar el metabolismo de un organismo. Existe interés general en esta estrategia y se dispone comercialmente de algunos vehículos que permiten modular el grado de expresión de un gene clonado por medio de la adición de un inductor (p. ej. IPTG con el promotor *lac*). Otros vehículos utilizan ARN polimerasas exógenas, por ejemplo la del bacteriófago T7, la cuál transcribe exclusivamente el gene insertado bajo su promotor específico, lo cuál asegura la sobre-expresión de la proteína clonada con respecto al resto del genoma. Se dispone de sofisticados vehículos que fusionan a la proteína clonada a un péptido líder que dirige su exportación y para purificarla basta concentrar el sobrenadante de un cultivo de tejidos, reduciendo al mínimo el proceso (167).

Aunque útiles, estos sistemas funcionan únicamente en cultivos, ya sea de bacterias o tejidos eucarióticos. Para controlar la expresión de genes en organismos completos se utilizan vehículos en los cuales el gene clonado se encuentre bajo el control de un promotor específico de cierto tejido o etapa del desarrollo proveniente del mismo organismo o de alguno relacionado. Como ejemplos, se han utilizado en leguminosas los promotores de proteínas específicas de nódulos (nodulinas) (40), y en mamíferos se utilizan el promotor temprano del virus SV40 o el de la timidina cinasa del virus de herpes simple, cuya expresión es tejido específica (167).

Específicamente para *Rhizobium*, se utilizan promotores inducibles por IPTG o promotores constitutivos, especialmente algunos de los que expresan genes de resistencia a antibióticos, p. ej. la cloramfenicol acetil transferasa de derivados del pBR325 (4). La expresión constitutiva de estos promotores se basa en su alta afinidad por  $E\sigma^{70}$ , por lo que son transcritos sin necesitar activadores (Ver Introducción).

Sin embargo, existen ocasiones en que el producto de cierto gene es nocivo para la célula en condiciones específicas o simplemente que se precise limitar su expresión a cierta etapa. Para lograr esto, el experimentador necesita coordinar la transcripción con alguna señal fácilmente controlable. Como ya se mencionó, cuando se trabaja con cultivos está señal puede ser tan sencilla como adicionar algún compuesto al medio, pero cuando se trabaja con el proceso de nodulación el manejo es más complicado y aunque se puede añadir el inductor a la solución de riego de las plantas, existe gran variación entre las condiciones locales de cada raíz, y sería muy difícil sincronizar todas las réplicas de un experimento.

Como una alternativa, hemos estado colaborando con el Dr. Alberto Mendoza en la construcción de un vehículo de expresión dependiente de NifA y por lo tanto inducible en condiciones de microaerobiosis y simbiosis. Este vehículo se basa en los resultados presentados en el manuscrito anexo, en el cuál se determinan los niveles de expresión de los genes reiterados para la nitrogenasa reductasa. De estos, *nifHc* se expresa a mayores niveles que *nifHa* durante el desarrollo del nódulo, siendo ambos transcritos a partir de promotores dependientes de  $\sigma^{54}$ .

El diseño básico de este vehículo se describe en la Figura 8 y consiste en la clonación de la

región promotora de *nifHc* en un vector de replicación estable en *Rhizobium* (pTR101) (188). Debajo de la región reguladora se encuentra un sitio de clonado múltiple (polilinker) proveniente del plásmido pUX19, derivado del pUK21 (180), que permita insertar en diferentes sitios. La eficiencia de la inserción puede monitorearse en cultivos microaeróbicos, midiendo la actividad de la proteína expresada en respuesta a la concentración de oxígeno, inducida en 1% y reprimida en 20%.

Esta estrategia ya ha sido utilizada para expresar el gene *gdhA*, el cual codifica para la enzima glutamato deshidrogenasa de *E. coli* en *R. etli* CE3. Se había demostrado con anterioridad que la expresión de GDH en *R. etli* bloquea la infección de raíces de frijol (110,111). Este efecto es aparentemente a través de la inhibición de los primeros pasos del establecimiento de la simbiosis, ya que la expresión del gene *nodA* se encuentra abatida en presencia de GDH, por lo que no se producen los factores de nodulación necesarios (110). Para superar este efecto y limitar la expresión de GDH a etapas posteriores del proceso, se insertó al gene *gdhA* bajo el control del promotor *nifHc* y se comprobó la fidelidad de la construcción al encontrar actividad de glutamato deshidrogenasa solamente en microaerobiósis. En este momento se están llevando a cabo los experimentos de infección de raíces de frijol correspondientes.

En una segunda línea de investigación también dentro de nuestro grupo, se ha llevado a cabo en colaboración con el Dr. Alberto Mendoza la clonación de los genes que codifican para la síntesis de polihidroxicanoatos *phaCBA* de *Alcaligenes eutrophus* en el vehículo de expresión específico de simbiosis que hemos mencionado. La presencia de este plásmido (pAM350) en *R. meliloti* permite la acumulación de PHB (polihidroxibutirato) en nódulos de alfalfa, lo cual no se observa con la cepa silvestre (120). Esta construcción permitirá el estudio de las interacciones del metabolismo de carbono con la eficiencia simbiótica en *R. meliloti*, estudios llevados a cabo con éxito en *R. etli* (50) pero que no habían podido ser extendidos a esa especie.

En conclusión, basados en la experiencia obtenida en la construcción de los dos plásmidos conteniendo los promotores *nifH* de *R. etli*, estamos trabajando en la construcción de un vehículo de expresión universal para simbiosis.

*Consideraciones sobre la evolución del sistema  
transcripcional de la nitrogenasa*

La capacidad de fijar nitrógeno atmosférico ha sido identificada tanto en eubacterias como en la rama de metanógenos/halófilos de las arqueobacterias, o sea, en la mitad de los grupos de procariotes conocidos. Dentro de estos grupos los fijadores de nitrógeno se encuentran en casi 100 géneros distribuidos sobre la mayor parte de las divisiones filogenéticas. La clasificación actual de todos los organismos está basada en la comparación de secuencias codificadoras para el ARN ribosomal 16S, un elemento universal y altamente conservado (105,138). Se ha cuestionado la validez de las inferencias filogenéticas basadas en la secuencia del 16S ribosomal con el argumento que no representa la evolución del genoma como tal y aunque en principio es cierto, la mayor parte de los estudios realizados con otras secuencias codificadoras coinciden esencialmente con el árbol de 16S, siendo las distorsiones la excepción (71).

En el caso de los genes estructurales para la nitrogenasa, la correspondencia con los árboles de 16S permitiría distinguir entre dos orígenes alternativos para la amplia distribución de esta capacidad entre los procariotes: la transferencia horizontal relativamente reciente de los genes *nif*, apoyada por la existencia de replicones movilizables portadores de estos genes en algunas especies; o bien, un ancestro común para todas las nitrogenasas, lo cual indicaría su origen como un evento previo a la separación de eubacterias y arqueobacterias.

En particular, el gene que codifica para la nitrogenasa reductasa se encuentra muy conservado y se ha utilizado para la construcción de árboles filogenéticos que corresponden fuertemente con el árbol de 16S (134,198). Adicionalmente, el contenido de GC de *nifH* refleja fielmente el del genoma que lo posee (69), una segunda evidencia de que este gene es un componente establecido del genoma que lo porta. Ambas líneas de evidencia indican que los genes para la nitrogenasa tienen un origen verdaderamente antiguo y quizá presentes en el ancestro de todos los organismos existentes.

Existe disparidad acerca de la antigüedad de la fijación biológica del nitrógeno. Algunos autores consideran que surgió hace 3 mil millones de años, correspondiendo con una atmósfera reductora (166), con lo cual explican la sensibilidad a oxígeno de la nitrogenasa, mientras que otros concluyen que su aparición es posterior a la oxidación de la atmósfera (106), y que la sensibilidad a oxígeno se debe más a razones termodinámicas que a una evidencia de su origen. Independientemente de la fecha exacta del surgimiento del proceso, las evidencias bioquímicas (21) y moleculares (198) obtenidas indican que existe solo un origen común para todos los sistemas de nitrogenasas conocidos.

La genética de la fijación de nitrógeno ha sido estudiada intensamente y sería imposible citar todos los trabajos realizados, pero se pueden consultar revisiones recientes sobre *K. pneumoniae*, considerado el paradigma (121) y sobre el género *Rhizobium*, dentro del cual se encuentra el organismo estudiado en este trabajo (53). Una de las conclusiones integradoras derivadas de todos estos estudios es la conservación del sistema transcripcional de los genes *nif*,

especialmente de *nifHDK*. En todos los casos descritos hasta ahora, la transcripción se efectúa a partir de un promotor dependiente de  $\sigma^{54}$  activado por NifA. La arquitectura del complejo activador parece estar también conservada, pues en todos los casos donde se ha buscado se ha encontrado el doblez impuesto por IHF (Fig. 7). Sin embargo, toda esta información se deriva del estudio de eubacterias gram-negativas.

No se han logrado identificar secuencias similares a promotores dependientes de  $\sigma^{54}$  o sitios de pegado de NifA en secuencias de *nifH* proveniente de firmibacterias (eubacterias gram-positivas) o arqueobacterias (32,164,165). En *Clostridium pasteurianum*, una firmibacteria, existen secuencias similares a promotores dependientes de  $\sigma^{70}$  arriba de *nifH2* (32). Esto sugiere que *nifH* se encuentra bajo el control de otro sistema transcripcional, ya que se conoce que el homólogo de  $\sigma^{54}$  (sigma L) caracterizado en *B. subtilis* (42) reconoce secuencias similares al consenso descrito en bacterias gram-negativas (23).

En el caso de las arqueobacterias, se ha descrito en *Methanococcus thermolithotrophicus* la transcripción de *nifH1* a partir de un promotor constitutivo (62). Este promotor consiste en la secuencia TATA en la posición -25 y la secuencia TTGT en la posición +1 con respecto al inicio de la transcripción. Ambos elementos se encuentran también en la secuencia reguladora del ARN ribosomal de *Sulfolobus shibatae* (148) y del ARN de transferencia de valina en *Methanococcus vanielli* (68).

Todos estos resultados indican que existe más de un mecanismo transcripcional de genes *nifH*, a pesar de su origen ancestral. Dado que obviamente se necesitan más casos para establecer la distribución de los mecanismos de transcripción, se realizó el análisis de otras secuencias reguladoras relevantes depositadas en los bancos de secuencias, buscando elementos reguladores similares a los previamente descritos. Se revisaron genes *nifH* de seis organismos diferentes, dos cianobacterias (*Anabaena azollae* y *Nostoc commune*), una thalobacteria (*Frankia*), tres arqueobacterias (*Methanobacterium thermoautotrophicum*, *Methanococcus voltae* y *Methanococcus maripaludis*), y dos proteobacterias (*Acetobacter diazotrophicus* (57) y *Rhodobacter sphaeroides*). Esta última se clasifica en la misma subdivisión de *R. etli*.

Como se esperaba, dada su cercanía con *Rhizobium*, los genes de *A. diazotrophicus* y *R. sphaeroides* presentan un promotor dependiente de  $\sigma^{54}$  y una secuencia consenso de pegado para NifA. Los mismos elementos se encuentran en los genes de las cianobacterias *A. azollae* y *N. commune* (Tabla 2).

Las thalobacterias son eubacterias gram-positivas y forman junto con las firmibacterias el grupo de las firmicutas. Dada su cercanía, esperábamos encontrar similitudes entre los elementos reguladores del gene de *Frankia* con los del gene de *C. pasteurianum*, sin embargo, no fue posible encontrar arriba de *nifH* de *Frankia* secuencias similares a promotores dependientes de  $\sigma^{54}$ , a los elementos detectados en *C. pasteurianum*, ni a un promotor de la familia de  $\sigma^{70}$  identificado en

otra cepa de *Frankia* (38), por lo que no se puede concluir nada al respecto de este organismo. La falta de similitud entre los elementos detectados en *Frankia* y en *C. pasteurianum* puede deberse a la diferencia en el contenido de GC de sus genomas.

Tal como podría esperarse, dada su cercanía filogenética, las tres secuencias provenientes de arqueobacterias presentan similitud con el elemento TATA-N25-TTGT identificado arriba de *nifH* de *M. thermolithotrophicus*. En este organismo se localiza en la posición -90 con respecto al codón de iniciación, mientras que en *M. voltae* se localiza en la posición -50, en *M. thermoautotrophicus* en la posición -45 y en *M. maripaludis* en la posición -80 (Tabla 2). El gene *nifH* de *M. thermoautotrophicus* está precedido también por una secuencia similar a un promotor dependiente de  $\sigma^{54}$  pero no por una secuencia de pegado de NifA. Esta última similitud bien pudiera ser azarosa pues el contenido de GC en este organismo es de 52% mientras que en los otros tres es de 28-38%, con lo que sería mayor la probabilidad de presentar una secuencia similar a un promotor dependiente de  $\sigma^{54}$ , rico en GC.

Esta información, tomada en conjunto, nos indica la existencia de dos diferentes mecanismos transcripcionales para los genes de la nitrogenasa, dependiendo de si el organismo que los porta es eubacteria o arqueobacteria, y que a pesar de sus diferencias obvias comparten algunas características comunes.

Se ha discutido con anterioridad la similitud entre el mecanismo de  $E\sigma^{54}$  y la ARN polimerasa eucariótica, siendo algunos de los puntos más relevantes:

1.- La capacidad del factor eubacterial  $\sigma^{54}$  (16) o la proteína TBP (TATA binding protein) eucariótica (14,158) de formar complejos cerrados estables.

2.- La interacción demostrada de TBP con el surco menor del ADN (158) y sugerida para  $\sigma^{54}$  (25,129). Virtualmente todas las otras proteínas que unen ADN lo hacen a través de contactos con el surco mayor.

3.- La estabilización de la interacción inicial  $\sigma^{54}$ -ADN por la enzima básica (27) en eubacterias y de TBP-ARNP-ADN por TFIIA y TFIID (101).

4.- El requerimiento de la hidrólisis de ATP para la isomerización del promotor a un complejo abierto. En ambos sistemas existe una correlación directa entre de este paso y un cambio conformacional de la holoenzima (158,185).

5.- La presencia de secuencias "enhancers" (101,190).

6.- La existencia de motivos estructurales típicos de activadores eucarióticos en  $\sigma^{54}$  (157).

Estos puntos han sido resaltados con anterioridad por otros autores, pero lo que no se ha hecho es correlacionar el mecanismo de  $\sigma^{54}$  con el del factor de iniciación arqueal (aTBP). Aunque no se ha estudiado tan exhaustivamente como los factores mencionados, aTBP ha demostrado ser esencialmente homólogo, tanto a nivel de secuencia como a nivel de mecanismo con el factor TBP eucariótico (5). Su similitud permite inclusive la sustitución de aTBP por TBP de levadura o de

humano en la transcripción *in vitro* de promotores arqueales (191). Es especialmente interesante que uno de los promotores probados es el de *nifH* de *M. thermolithotrophicus* (191).

De los seis puntos similares entre los mecanismos de  $\sigma^{54}$  y TBP, se ha demostrado que aTBP también forma complejos estables con el promotor y que estos son estabilizados por otros componentes de la holoenzima (154). Dado que aún no existen suficientes evidencias experimentales sobre el mecanismo de aTBP para relacionarla directamente con  $\sigma^{54}$ , se llevó a cabo una búsqueda sobre secuencias de factores  $\sigma$ , TBP y aTBP que permitieran un análisis filogenético que nos indicara sus relaciones (Tabla 3).

De las 30 secuencias obtenidas, las más corta son las aTBP con una longitud promedio de 200 residuos y las más largas las de la familia de  $\sigma^{70}$ , con una longitud promedio de 800 residuos. Del alineamiento de todas estas secuencias se seleccionó una porción de 200 residuos conteniendo la mayor parte de las TBP y aTBP, las regiones 2, 3 y 4.1 de  $\sigma^{70}$  y el dominio III de  $\sigma^{54}$ .

En base a este alineamiento, se llevó a cabo un análisis filogenético con algunos de los algoritmos más utilizados (UPGMA, Fitch, Neighbor-Joining, Máxima parsimonia). El único algoritmo que produjo un árbol congruente fue el de Neighbor-Joining (Fig. 9), en el resto de los árboles los grupos se intercalaban de manera azarosa. En este árbol se puede observar que cada una de las secuencias es bastante similar con los demás elementos de su mismo grupo, mientras que los diferentes grupos se encuentran razonablemente distanciados. La única excepción es  $\sigma^{32}$  de *C. crecentus*, que se encuentra en un grupo más cercano a las aTBP que al resto de los miembros de la familia de  $\sigma^{70}$ . Esta excepción se debe a la presencia de un par de indels en esta secuencia que no son compartidos por el resto de los miembros de la familia.

Las seis secuencias de  $\sigma^{54}$  forman un grupo discreto, a una distancia de 1.3 unidades del nodo central. Las proteínas arqueales y eucarióticas se organizan en tres ramas, las dos más cercanas contienen las provenientes de arqueobacterias, la primera las de *S. shibatae* y *H. salinarum* (0.03 unidades) y la segunda las de *P. woesei* y *T. celer* (0.45 unidades). Las secuencias eucarióticas se organizan en tercer grupo localizado a 0.45 unidades del nodo central. En resumen, si tomamos el nodo más cercano al nodo central de cada grupo, la distancia entre la rama de  $\sigma^{54}$  y la de  $\sigma^{70}$  es de 2.9 unidades, mientras que con las aTBP es de 1.33 a 1.85 unidades.

Esta distribución coincide más con las características del mecanismo transcripcional de los tres grandes grupos de organismos existentes que con sus relaciones filogenéticas reales. A pesar de las grandes diferencias entre eucariotes y arqueobacterias en la estructura de la ARN polimerasa existe suficiente similitud a nivel de secuencia de más de un componente para considerarlas homólogas. Con respecto a las eubacterias, dado el tiempo transcurrido desde su separación de las arqueobacterias, es muy significativo que todavía se reconozcan similitudes entre sus mecanismos de acción, especialmente con  $E\sigma^{54}$ , aunque no se conserve a nivel de secuencia.

Se ha tratado previamente de encontrar similitud a nivel de secuencia entre  $\sigma^{70}$  y factores eucarióticos o arqueales pero en ningún caso ha sido significativa (49). Si adicionalmente consideramos que sus mecanismos no son equivalentes y tomamos en cuenta las distancias reportadas en la Fig. 9, parecería más difícil sostener su homología. Una alternativa viable es suponer que  $\sigma^{70}$  proviene de la modificación de un  $\sigma^{54}$  primigenio, el cuál si estaría relacionado con los factores TBP y aTBP a nivel de mecanismo. Desafortunadamente la similitud a nivel de secuencia entre  $\sigma^{54}$  y  $\sigma^{70}$  es muy baja, menor de la que se considera indicadora de homología, por lo que no puede utilizarse como argumento. Sin embargo podemos apoyarnos en evidencias funcionales, específicamente una publicación reciente donde por medio de una serie de sustituciones a saturación en el dominio de activación de  $\sigma^{54}$  de *E. coli*, ésta es liberada de la necesidad de un activador, capacitandola para isomerizar espontaneamente al complejo cerrado (184). El camino inverso no ha sido demostrado, más dada la complejidad del mecanismo de activación de promotores dependientes de  $\sigma^{54}$  podría ser menos factible.

Finalmente, si el ancestro de todos los factores de iniciación conocidos fuera aTBP, podría haber sucedido que algunos promotores que no requirieran altos niveles basales de transcripción conservaran el mecanismo original (*nifHDK*, *glnA*, *flbG*, etc.), mientras que aquellos involucrados en funciones fundamentales (proteínas ribosomales, resistencia a antibióticos, división celular, replicación, etc.), adquirieran el sistema de  $\sigma^{70}$ , de mayor rendimiento.

*Un homólogo de hemN en el plásmido simbiótico de Rhizobium etli*

Como se mencionó en los antecedentes, la región de ADN directamente abajo del gene *nifHc* se encuentra altamente conservada en todos los aislados de *R. etli* probados, pero no en otras Rhizobaceas, como por ejemplo *R. loti*, *B. loti*, *R. fredii* o *R. meliloti* según se muestra en la Figura 10. La secuencia nucleotídica de esta porción del plásmido simbiótico reveló la existencia de una fase abierta de lectura de 756 nucleótidos que codifica para una proteína de 251 residuos (Fig. 11). Como puede observarse en la Fig. 12, la secuencia deducida de ORF251 es 50% similar a varias de las proteínas HemN, una enzima involucrada en la síntesis del grupo hemo.

La biogénesis bacteriana de tetrapirroles involucra el producto de al menos nueve genes diferentes. Los primeros pasos producen al precursor común del grupo hemo y de la clorofila, comenzando con la conversión de ácido glutámico en ácido  $\delta$ -aminolevulínico hasta la síntesis de protoporfirina IX. Las primeras reacciones, hasta la formación de coproporfirinógeno III son llevados a cabo por las proteínas HemL, HemA, HemB, HemC, HemD y HemE. La oxidación de éste último compuesto puede ser catalizada por las isoenzimas HemN y HemF. El último paso lo lleva a cabo la proteína HemG, y el producto final de la vía, la protoporfirina IX puede seguir dos caminos alternativos. En caso de ser asociada por HemH a una molécula de hierro, se convertirá en un grupo hemo, mientras que si es a una molécula de magnesio, dará origen al grupo clorofila (39). Evidencias genéticas y bioquímicas sobre la regulación de ésta vía sugieren la existencia de dos puntos de control. El primero en la formación del ácido  $\delta$ -aminolevulínico y el segundo en la formación de protoporfirina IX a partir del coproporfirinógeno III (39).

En *E. coli*, los genes *hem* están distribuidos en todo el cromosoma. Alrededor del minuto 25 se encuentran los genes *hemK*, *hemA* y *hemM*, entre los minutos 85 y 87 se encuentran los genes *hemC*, *hemD*, *hemX*, *hemG* y *hemN*. El gene *hemF* se encuentra en el minuto 52, *hemE* en el 89, y *hemL* en el 3, mientras que no se sabe la localización del gene *hemB*. Los genes *hemN* y *HemF* codifican para las dos isoenzimas de la coproporfirinógeno III oxidasa. Ambos genes han sido clonados en *S. typhimurium* y en *E. coli* (59,174,175,196,197). Se han identificado homólogos a *hemN* en *Haemophilus influenzae* (54), *Rhodobacter sphaeroides* (36) y *Paracoccus denitrificans* (NCBI gi 1002875). Secuencias equivalentes a *hemF* han sido identificadas en *Chlamydomonas reinhardtii* (85) y *Pseudomonas aeruginosa* (NCBI gi 695693) así como en varios eucariotes, donde se conoce como *hem13* o *hem6*: ratón (88), humano (107), soya (103) y levadura (103,201).

La decarboxilación oxidativa del coproporfirinógeno III para formar protoporfirinógeno IX requiere un receptor de electrones. HemF utiliza oxígeno molecular, mientras que HemN utiliza NAD<sup>+</sup> o NADP<sup>+</sup> (159,195). Cepas de *E. coli* interrumpidas en cualquiera de los dos genes carecen de fenotipo en condiciones de crecimiento aeróbico, sin embargo, mutaciones en *hemN* confieren auxotrofia por el grupo hemo en condiciones anaeróbicas (195). Estos datos indican que la síntesis anaeróbica del grupo hemo requiere la función de HemN, mientras que la aeróbica

puede ser completada por cualquiera de las dos isoenzimas. Se ha propuesto que la actividad de la isoforma anaeróbica de *R. sphaeroides* está coordinada con la síntesis de clorofila en condiciones de inducción de fotosíntesis (36).

Una mutante derivada de la cepa CE3 con una inserción en ORF251 no presenta fenotipo simbiótico, ni diferencias en su crecimiento en medio rico PY o en medio mínimo con cloruro de amonio y succinato en condiciones aeróbicas. No se detectó expresión de una fusión ORF251-*lacZ* en microaerobiosis, probada bajo las mismas condiciones de ensayo de los genes *nifH* (Ver manuscrito). Sin embargo, se expresa activamente en nódulos de frijol (Fig. 13).

Es de suponerse que la síntesis del grupo hemo es indispensable para los bacteroides de *Rhizobium*, ya que no solo es precursor para los intermediarios de la cadena respiratoria, sino que forma parte de la leghemoglobina, una proteína específica de simbiosis (nodulina) que enlaza oxígeno molecular y lo transporta hacia el interior de nódulo (1). La leghemoglobina es una proteína fundamental para el funcionamiento del nódulo, y durante las etapas tardías de su desarrollo se sintetiza en grandes cantidades (1). Mutantes de *B. japonicum* carentes de las proteínas HemA o HemG presentan graves deficiencias simbióticas (43,137). La falta de fenotipo simbiótico de la inserción en ORF251 podría explicarse de dos maneras, la más directa sería por complementación con HemF, aunque el gene para esta proteína no ha sido detectado aún en *Rhizobium*. La otra sería por complementación cruzada con la planta, ya que el homólogo de *hemF* en soya (*hem6*) se transcribe abundantemente en nódulos, coincidiendo con la etapa de síntesis de leghemoglobina (103).

Dadas las diferencias entre los perfiles de expresión de ORF251 con el de *hemA* de *B. japonicum* en cultivo (26), consideramos posible que no hayamos logrado encontrar las condiciones adecuadas para su inducción. La expresión de *hemA* se observa en condiciones de crecimiento anaeróbicas, mientras que las nuestras son microaeróbicas. Sin embargo, ambos genes se expresan en nódulo, lo que sugiere que ORF251 es un verdadero gene, y que está bajo regulación simbiótica, tal como *hemA*.

A partir de la secuencia del producto de ORF251 llevamos a cabo un análisis filogenético comparativo contra otras 13 secuencias disponibles de genes *hemN* o *hemF*. El resultado de este análisis se presenta en tres variantes obtenidas a partir del alineamiento de la secuencia central compartida por todos los elementos. El primer árbol contiene las 14 secuencias en un arreglo obtenido por medio del algoritmo de UPGMA, a partir de una tabla de distancias genéticas corregidas por el método de Kimura, utilizando el programa Phylip (Fig 14) (52). En esta figura se observa que todas las secuencias se organizan en dos grupos claramente definidos, el grupo I incluye las proteínas HemF de *E. coli*, *S. typhimurium* y *P. aeruginosa* y las proteínas Hem6 eucarióticas, mientras que el grupo II consta de dos subgrupos. En el subgrupo IIA se encuentran dos proteínas aún no caracterizadas fisiológicamente, una de *H. influenzae* y el producto de una

fase de lectura localizada en el minuto 66 del cromosoma de *E. coli* (NCBI gi 882484). Ambas secuencias se obtuvieron a partir de proyectos de secuenciación genómica y fueron seleccionadas de los bancos de secuencias en base a su alta similitud con HemN. El subgrupo IIB contiene el resto de las secuencias HemN. Las de *E. coli*, *S. typhimurium* y *R. sphaeroides* codifican para productos activos bioquímicamente, mientras que las de *P. denitrificans* y de *R. etli* solo son similares a nivel de secuencia. Se obtuvo un arreglo esencialmente igual al utilizar los algoritmos de Neighbor Joining y de Fitch.

Al comparar secuencias provenientes de organismos tan distantes filogenéticamente como procariotes y eucariotes es probable subestimar las distancias reales entre ambos grupos. Esto se debe a que conforme mas lejano sea el evento de divergencia es más frecuente la ocurrencia de eventos de sustitución múltiple en el mismo sitio. Para afinar las relaciones entre las secuencias procarióticas, repetimos el análisis con todos los elementos del grupo II y las secuencias de HemF de *E. coli* y *S. typhimurium*. Como puede observarse en la Fig. 15, las dos proteínas HemF se conservan en un grupo definido, distante de las anaeróbicas. Los elementos de grupo II se asocian en congruencia con las relaciones filogenéticas observadas entre ellos en base al análisis de la secuencia de la subunidad pequeña del ARN ribosomal. La única excepción la presentan la secuencias de *H. influenzae* y la fase de lectura identificada en *E. coli*, ya que esta ultima no se agrupa con las otras dos secuencias de bacterias entéricas.

Los resultados anteriores indican tres posibles linajes para la enzima coproporfirinógeno III oxidasa. En resumen, el grupo I incluye a las secuencias dependientes de oxígeno como aceptor de electrones, ya sean procarióticas o eucarióticas, mientras que el grupo II todas las secuencias relacionadas con la utilización de otros aceptores de electrones. Los tres linajes sugeridos son más fácilmente observados en el alineamiento de las secuencias provenientes de bacterias entéricas (*E. coli* y *S. typhimurium*). La similitud marginal observada entre las secuencias de las formas aeróbicas y anaeróbicas podría indicar un origen común, pero sería necesario invocar tasas de mutación completamente diferentes para cada uno de los genes como explicación de la fuerte divergencia (Fig. 16). Creemos que una alternativa más probable sería la convergencia funcional, y que proviniendo de orígenes diferentes, adquirieran la misma función. La tercera variante (codificada por el ORF encontrado en *E. coli*) es suficientemente parecida a HemN para considerarla homóloga, aunque contiene diferencias discretas significativas, lo que también ocasiona distorsión en la clasificación. Es importante recordar que aunque no existen evidencias de función para esta secuencia, es altamente similar con HemN de *H. influenzae* y dado que aparentemente no existen en este ultimo organismo otras secuencias similares a *hemN* o *hemF*, es muy probable que sea funcional (54).

Para concluir, no tenemos evidencias sobre la regulación de ORF251 de *R. etli*. Sabemos que se expresa en simbiosis, pero que su patrón de expresión es diferente al del gen *nifHc*, al que

se encuentra asociado. Esto indica que su expresión es a partir de un promotor independiente de NifA pero que se expresa específicamente en simbiosis. En concordancia, el análisis de la secuencia reguladora de ORF251 no muestra ningún probable promotor dependiente de  $\sigma^{54}$  o sitio de unión de NifA. Sin embargo, en la posición 131 antes del probable inicio de la traducción se encuentra la siguiente secuencia, GAACT-N16-TCT, recientemente reportada como la secuencia de promotores dependientes del factor  $\sigma^E$  (70,153). Este factor ha sido caracterizado en la respuesta de *E. coli* a choques de calor, aunque también existen evidencias de su papel como regulador transcripcional en respuesta a estímulos extracelulares (153). En *P. aeruginosa* participa en la síntesis de alginatos y su actividad es uno de los puntos de control de la adquisición de virulencia en este organismo (70).

Para finalizar, quisiera hacer énfasis en el significado de la conservación de ORF251 en *R. etli* y su estrecha asociación con el gene *nifHc* en el plásmido simbiótico. Estudios recientes sobre la estructura de la cadena respiratoria, tanto de *R. etli* como de *B. japonicum*, demuestran la existencia de un control estricto en la expresión de algunos elementos específicos para simbiosis (136,171). Quizá la biosíntesis del grupo hemo también se encuentra regulada simbioticamente, y la localización de ORF251 en el plásmido simbiótico y su expresión son parte de este sistema. Recientemente han sido clonado por Adriana Castillo en nuestro grupo otros elementos del regulón de genes *hem*, por lo que en un futuro próximo se podrán llevar a cabo experimentos integradores de la información presentada aquí, con respecto a la regulación genética global de estos genes en *R. etli*.

# **БІБЛІОГРАФІЯ**

1. **Appleby, C.A.** 1984. Leghemoglobin and *Rhizobium* respiration. *Ann.Rev.Plant Physiol.* **35**:443-478.
2. **Auble, D.T. and P.L. de Haseth.** 1988. Promoter recognition by *Escherichia coli* RNA polymerase: influence of DNA structure in the spacer separating the -10 and -35 regions. *J.Mol.Biol.* **202**:471-482.
3. **Austin, S. and R. Dixon.** 1992. The prokaryotic enhancer binding protein NTRC has an ATPase activity which is phosphorylation and DNA dependent. *EMBO J* **11**:2219-2228.
4. **Balbás, P. and F. Bolívar.** 1990. Design and construction of expression plasmid vectors in *Escherichia coli*. *Meth.Enzymol.* **185**:14-37.
5. **Baumann, P., S.A. Qureshi and S.P. Jackson.** 1995. Transcription: new insights from studies on Archaea. *TIG* **11**:279-283.
6. **Berger, D.K., F. Narberhaus and S. Kustu.** 1994. The isolated catalytic domain of NIFA, a bacterial enhancer-binding protein, activates transcription *in vitro*: Activation is inhibited by NIFL. *Proc.Natl.Acad.Sci.USA.* **91**:103-107.
7. **Beynon, J.L., J.E. Beringer and A.W.B. Johnston.** 1980. Plasmids and host-range in *Rhizobium leguminosarum* and *Rhizobium phaseoli*. *J.Gen.Microbiol.* **120**:421-429.
8. **Borowiec, J.A. and J.D. Gralla.** 1986. High-resolution analysis of transcription complexes inside cells. *Biochemistry* **25**:5051-5057.
9. **Borthakur, D., J.A. Downie, A.W.B. Johnston and J.W. Lamb.** 1985. *psi*, a plasmid-linked *Rhizobium phaseoli* gene that inhibits exopolysaccharide production and which is required for symbiotic nitrogen fixation. *Mol.Gen.Genet.* **200**:278-282.
10. **Borthakur, D. and A.W.B. Johnston.** 1987. Sequence of *psi*, a gene on the symbiotic plasmid of *Rhizobium phaseoli* which inhibits exopolysaccharide synthesis and nodulation and demonstration that its transcription is inhibited by *psr*, another gene on the symbiotic plasmid. *Mol.Gen.Genet.* **207**:149-154.
11. **Borthakur, D., J.W. Lamb and A.W.B. Johnston.** 1987. Identification of two classes of *Rhizobium phaseoli* genes required for melanin synthesis, one of which is required for nitrogen fixation and activates the transcription of the other. *Mol.Gen.Genet.* **207**:155-160.
12. **Borukhov, S., V. Sagitov and A. Goldfarb.** 1993. Transcript cleavage factor from *E. coli*. *Cell* **72**:459-466.
13. **Brodolin, K.L., V.M. Studitsky and A.D. Mirzabekov.** 1993. Conformational changes in *E. coli* RNA polymerase during promoter recognition. *Proc.Natl.Acad.Sci.USA.* **21**:5748-5753.
14. **Brown, D.** 1984. The role of stable complexes that repress and activate eucaryotic genes. *Cell* **37**:359-365.
15. **Buchanan-Wollaston, V., M.C. Cannon, J.L. Beynon and F.C. Cannon.** 1981. Role of the *nifA* gene product in the regulation of *nif* expression in *Klebsiella pneumoniae*. *Nature* **294**:776-778.

16. **Buck, M. and W. Cannon.** 1992. Specific binding of the transcription factor sigma-54 to promoter DNA. *Nature* **358**:422-424.
17. **Buck, M. and W. Cannon.** 1992. Activator-independent formation of a closed complex between  $\sigma^{54}$ -holoenzyme and *nifH* and *nifU* promoters of *Klebsiella pneumoniae*. *Mol.Microbiol.* **6**:1625-1630.
18. **Buck, M., W. Cannon and J. Woodcock.** 1987. Transcriptional activation of the *Klebsiella pneumoniae* nitrogenase promoter may involve DNA loop formation. *Mol.Microbiol.* **1**:243-249.
19. **Buck, M. and E. Morett.** 1988. Transcriptional activation of the *Klebsiella pneumoniae nifH* promoter by upstream bound NifA protein. *J Cell Biochem* **32**:137-139.
20. **Burgess, R.R., A.A. Travers, J.J. Dunn and E.W. Bourne.** 1969. Factor stimulating transcription by RNA polymerase. *Nature* **221**:45-47.
21. **Burris, R.H.** 1991. Nitrogenases. *J.Biol.Chem.* **266**:9370-9373.
22. **Busby, S. and R.H. Ebright.** 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**:743-746.
23. **Calogero, S., R. Gardan, P. Glaser, J. Schweitzer, G. Kapermort and M. Debarbouille.** 1994. RocR, a novel regulatory protein controlling antibiotic utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. *J.Bacteriol.* **176**:1234-1241.
24. **Cannon, W., S. Austin, M. Moore and M. Buck.** 1993. Identification of close contacts between the  $\sigma^N$  ( $\sigma_{54}$ ) protein and promoter DNA in closed promoter complexes. *Nucl.Acids Res.* **23**:351-356.
25. **Cannon, W. and M. Buck.** 1992. Central domain of the positive control protein NifA and its role in transcriptional activation. *J.Mol.Biol.* **225**:271-286.
26. **Cannon, W., W. Charlton and M. Buck.** 1991. Organization and function of binding sites for the transcriptional activator NifA in the *Klebsiella pneumoniae nifH* and *nifU* promoters. *J.Mol.Biol.* **220**:915-931.
27. **Cannon, W., F. Claverie-Martin, S. Austin and M. Buck.** 1993. Core RNA polymerase assists binding of the transcription factor  $\sigma^N$  to promoter DNA. *Mol.Microbiol.* **8**:287-298.
28. **Cannon, W., F. Claverie-Martin, S. Austin and M. Buck.** 1994. Identification of a DNA-contacting surface in the transcription factor sigma 54. *Mol.Microbiol.* **11**:227-236.
29. **Cannon, W., S. Missailidis, C. Smith, A. Cullier, S. Austin, M. Moore and M. Buck.** 1995. Core RNA polymerase and promoter DNA interactions of purified domains of  $\sigma^N$ : bipartite functions. *J.Mol.Biol.* **248**:781-803.
30. **Cannon, W.V., R. Kreutzer, H.M. Kent, E. Morett and M. Buck.** 1990. Activation of the *Klebsiella pneumoniae nifU* promoter: identification of multiple and overlapping upstream NifA binding sites. *Nucl.Acids Res.* **18**:1693-1701.

31. Cava, J.R., P.M. Elias, D.A. Turowski and K.D. Noel. 1989. *Rhizobium leguminosarum* CFN42 genetic regions encoding lipopolysaccharide structures essential for complete nodule development on bean plants. *J.Bacteriol.* **171**:8-15.
32. Chen, K.C.K., J.S. Chen and J.L. Johnson. 1986. Structural features of multiple *nifH*-like sequences and very biased codon usage in nitrogenase genes of *Clostridium pasteurianum*. *J.Bacteriol.* **166**:162-172.
33. Chen, P. and L.J. Reitzer. 1995. Active contribution of two domains to cooperative DNA binding of the enhancer-binding protein nitrogen regulator I (NtrC) of *Escherichia coli*: stimulation by phosphorylation and the binding of ATP. *J.Bacteriol.* **177**:2490-2496.
34. Claverie-Martin, F. and B. Magasanik. 1991. Role of integration host factor in the regulation of the *glnH*<sub>2</sub> promoter of *Escherichia coli*. *Proc.Natl.Acad.Sci.USA.* **88**:1631-1635.
35. Claverie-Martin, F. and B. Magasanik. 1992. Positive and negative effects of DNA bending on activation of transcription from a distant site. *J.Mol.Biol.* **227**:996-1008.
36. Coomber, S.A., R.M. Jones, P.M. Jordan and C.N. Hunter. 1992. A putative anaerobic coproporphyrinogen III oxidase in *Rhodobacter sphaeroides*. I. Molecular cloning, transposon mutagenesis and sequence analysis of the gene. *Mol.Microbiol.* **6**:3159-3169.
37. Coppard, J.R. and M.J. Merrick. 1991. Cassette mutagenesis implicates a helix-turn-helix motif in promoter recognition by the novel RNA polymerase sigma factor  $\sigma^{54}$ . *Mol.Microbiol.* **5**:1309-1317.
38. Cournoyer, B. and P. Normand. 1994. Gene expression in Frankia: characterization of promoters. *Microbios* **78**:229-236.
39. Dailey, H.A.(ed) 1990. Biosynthesis of heme and chlorophylls, McGraw Hill Publishing Co., Highstown, N.J.,
40. de Bruijn, F.J., L. Szabados and J. Schell. 1990. Chimeric genes and transgenic plants are used to study the regulation of genes involved in symbiotic plant-microbe interactions (nodulin genes). *Dev.Genet.* **11**:182-196.
41. de Haseth, P.L. and J.D. Helmann. 1995. Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol.Microbiol.* **16**:817-824.
42. Debarhoulle, M., I. Martin-Verstraete, F. Kunst and G. Rapoport. 1991. The *Bacillus subtilis sigL* gene encodes an equivalent of sigma 54 from gram-negative bacteria. *Proc.Natl.Acad.Sci.USA.* **88**:9092-9096.
43. Dickstein, R., D.C. Scheirer, W.H. Fowle and F.M. Ausubel. 1991. Nodules elicited by *Rhizobium meliloti* heme mutants are arrested at an early stage of development. *Mol.Gen.Genet.* **230**:423-432.
44. Doolittle, R.F., D-F. Feng, S. Tsang, G. Cho and E. Little. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**:470-477.
45. Downie, J.A., M.a. Qing-sheng, C.D. Knight, G. Hombrecher and A.W.B.

- Johnston.** 1983. Cloning of the symbiotic region of *Rhizobium leguminosarum*: the nodulation genes are between the nitrogenase genes and a *nifA*-like gene. *EMBO J.* 2:947-952.
46. **Drummond, M., P. Whitty and J. Wootton.** 1986. Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*: homologies to other regulatory proteins. *EMBO J.* 5:441-447.
47. **Drummond, M.H., A. Contreras and L.A. Mitchenall.** 1990. The function of isolated domains and chimaeric proteins constructed from the transcriptional activators NifA and NtrC of *Klebsiella pneumoniae*. *Mol.Microbiol.* 4:29-37.
48. **Eardly, B.D., F.S. Wang, .T.S. Whittam and R.K. Selander.** 1995. Species limits in *Rhizobium* populations that nodulate the common bean (*Phaseolus vulgaris*). *Appl.Environ.Microbiol.* 61:507-512.
49. **Elck, D., A. Wedel and H. Heumann.** 1994. From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. *TIG* 10:292-296.
50. **Encarnación, S., M. Dunn, K. Willms and J. Mora.** 1995. Fermentative and aerobic metabolism in *Rhizobium etli*. *J.Bacteriol.* 177:3058-3066.
51. **Eydmann, T., E. Söderbäck, T. Jones, S. Hill, S. Austin and R. Dixon.** 1995. Transcriptional activation of the nitrogenase promoter in vitro: adenosine nucleotides are required for inhibition of NIFA activity by NIFL. *J.Bacteriol.* 177:1186-1195.
52. **Felsenstein, J.** 1993. PHYLIP, Phylogeny Inference Package (Dept. of Genetics, Univ. of Washington, Seattle), Version 3.5C.
53. **Fischer, H.M.** 1994. Genetic regulation of nitrogen fixation in Rhizobia. *Microbiol.Revs.* 58:352-386.
54. **Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C.A. Fields, J.D. Gocavne, J.D. Scott, R. Shirley, L.I. Liu, A. Glodeck, J.M. Kelley, J.F. Weldman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith and J.C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512.
55. **Foster-Harnett, D. and R.G. Kranz.** 1992. Analysis of the promoters and upstream sequences of *nifA1* and *nifA2* in *Rhodobacter capsulatus*: activation requires *ntrC* but not *rpoN*. *Mol.Microbiol.* 6:1049-1060.
56. **Frank-Kamenetskii, M.** 1987. DNA chemistry. How the double helix breathes. *Nature* 328:17-18.
57. **Fuentes-Ramírez, L.E. and E. Martínez-Romero.** 1995. Secuencia de la región regulatoria de la transcripción del gene *nifH* de *Acetobacter diazotrophicus*, p.58. Procedimientos del IV Congreso Nacional de la Fijación Biológica del Nitrógeno.
58. **Giblin, L., B. Boesten, S. Turk, P. Hooykas and F. O'gara.** 1995. Signal

transduction in the *Rhizobium meliloti* dicarboxylic acid transport system. FEMS Microbiol.Lett. 126:25-30.

59. Gibson, L.C., P. McGlynn, M. Chaudhri and C.N. Hunter. 1992. A putative anaerobic coproporphyrinogen III oxidase in *Rhodobacter sphaeroides*. II. Analysis of a region of the genome encoding *hemF* and the *puc* operon. Mol.Microbiol. 6:3171-3186.

60. Girard, M.deL., M. Flores, S. Brom, D. Romero, R. Palacios and G. Dávila. 1991. Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. phaseoli. J.Bacteriol. 173:2411-2419.

61. Girard, M.deL., B. Valderrama, S. Brom, R. Palacios, D. Romero and G. Dávila. 1993. Identification and specific amplification of *nifA* gene of *Rhizobium leguminosarum* bv. phaseoli, p.614. In R. Palacios, J. Mora and W.E. Newton (ed.), New horizons in Nitrogen Fixation, Dordrecht:Kluwer Academic Publishers,

62. Gohl, H.P., W. Hausner and M. Thomm. 1992. Cell-free transcription of the *nifH1* gene of *Methanococcus thermolithotrophicus* indicates that promoters of archaeal *nif* genes share basic features with the methanogen consensus promoter. Mol.Gen.Genet. 231:286-295.

63. Goosen, N. and P. van de Putte. 1995. The regulation of transcription initiation by integration host factor. Mol.Microbiol. 16:1-7.

64. Greenblatt, J. 1992. Protein-protein interactions as critical determinants of regulated initiation and termination of transcription, p.203. Transcriptional regulation, Cold Spring Harbor Laboratory Press, New York.

65. Gross, C.A., M. Loneto and R. Losick. 1992. Bacterial sigma factors, p.129. In S.L. McKnight and K.R. Yamamoto (ed.), Transcriptional regulation, Cold Spring Harbor Laboratory Press, New York.

66. Gu, B., J.H. Lee, T. Hoover, D. Scholl and B.T. Nixon. 1994. *Rhizobium meliloti* DctD, a  $\sigma^{54}$ -dependent transcriptional activator, may be negatively controlled by a subdomain in the C-terminal end of its two-component receiver module. Mol.Microbiol. 13:51-66.

67. Gubler, M. 1989. Fine-tuning of *nif* and *fix* genes expression by upstream activator sequences in *Bradyrhizobium japonicum*. Mol.Microbiol. 3:149-159.

68. Hausner, W., G. Frey and M. Thomm. 1991. Control regions of an archaea gene. A TATA box and initiator element promote cell free transcription of the tRNA Val gene of *Methanococcus vanniellii*. J.Mol.Biol. 222:495-508.

69. Hennecke, H., K. Kaluza, B. Thöny, M. Fuhrmann, W. Ludwig and E. Stackebrandt. 1985. Concurrent evolution of nitrogenase genes and 16s rRNA in *Rhizobium* species and other nitrogen fixing bacteria. Arch.Microbiol 142:342-348.

70. Hershberger, C.D., R.W. Ye, M.R. Parsek, Z.D. Xie and A.M. Chakrabarty. 1995. The *algT* (*algU*) gene of *Pseudomonas aeruginosa*, a key regulator involved in alginate biosynthesis, encodes an alternative  $\sigma$  factor ( $\sigma^E$ ). Proc.Natl.Acad.Sci.USA. 92:7941-7945.

71. Hirsch, A.M., H.I. McKhann, A. Reddy, J. Liao, Y. Fang and C.R. Marshall. 1995. Assessing horizontal transfer of *nifHDK* genes in eubacteria: nucleotide

sequence of *nifK* from Frankia strain HPFCc13. *Mol.Biol.Evol.* **12**:16-27.

72. Hirschman, J., P.K. Wong, K. Sei, J. Keener and S. Kustu. 1985. Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc.Natl.Acad.Sci.USA.* **82**:7525-7529.

73. Hoover, T.R., E. Santero, S. Porter and S. Kustu. 1990. The integration host factor stimulates interaction of RNA polymerase with NifA, the transcriptional activator for nitrogen fixation operons. *Cell* **63**:11-22.

74. Hsieh, M. and J.D. Gralla. 1994. Analysis of the N-terminal leucine heptad and hexad repeats of Sigma 54. *J.Mol.Biol.* **239**:15-24.

75. Hsieh, M., Y. Tintut and J.D. Gralla. 1994. Functional roles for the glutamines within the glutamine-rich region of the transcription factor  $\sigma^{54}$ . *J.Biol.Chem.* **269**:373-378.

76. Huala, E. and F.M. Ausubel. 1989. The central domain of *Rhizobium meliloti* NifA is sufficient to activate transcription from the *R. meliloti nifH* promoter. *J.Bacteriol.* **171**:3354-3365.

77. Huala, E., A.L. Moon and F.M. Ausubel. 1991. Aerobic inactivation of *Rhizobium meliloti* NifA in *Escherichia coli* is mediated by *lon* and two newly identified genes, *snoB* and *snoC*. *J.Bacteriol.* **173**:382-390.

78. Huala, E., J. Stigter and F.M. Ausubel. 1992. The central domain of *Rhizobium leguminosarum* DctD functions independently to activate transcription. *J.Bacteriol.* **174**:1428-1431.

79. Hudson, G.S. and B.E. Davidson. 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K12. *J.Mol.Biol.* **180**:1023-1051.

80. Hunt, T.P. and B. Magasanik. 1985. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. *Proc.Natl.Acad.Sci.USA.* **82**:8453-8457.

81. Ishihama, A. 1993. Protein-protein communications within the transcription apparatus. *J.Bacteriol.* **175**:2483-2489.

82. Jiang, J., B.H. Gu, L.M. Albright and B.T. Nixon. 1989. Conservation between coding and regulatory elements of *Rhizobium meliloti* and *Rhizobium leguminosarum* *dct* genes. *J.Bacteriol.* **171**:5244-5253.

83. Jones, R. and R. Haselkorn. 1989. The DNA sequence of the *Rhodobacter capsulatus* *ntrA*, *ntrB* and *ntrC* gene analogues required for nitrogen fixation. *Mol.Gen.Genet.* **215**:507-516.

84. Keener, J. and S. Kustu. 1988. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: Roles of the conserved amino-terminal domain of NTRC. *Proc.Natl.Acad.Sci.USA.* **85**:4976-4980.

85. Kill, K.L. and S. Merchant. 1995. Coordinate expression of coproporphyrinogen oxidase and cytochrome c6 in the green alga *Chlamydomonas reinhardtii* in response to changes in copper availability. *EMBO J* **14**:857-865.

86. Klose, K.E., A.K. North, K.M. Stedman and S. Kustu. 1994. The mayor dimerization determinants of the nitrogen regulatory protein NTRC from enteric bacteria lie in its carboxy-terminal domain. *J.Mol.Biol.* **241**:233-245.
87. Klose, K.E., D.S. Weiss and S. Kustu. 1993. Glutamate at the site of phosphorylation of nitrogen-regulatory protein NTRC mimics aspartyl-phosphate and activates the protein. *J.Mol.Biol.* **232**:67-78.
88. Kohno, J., T. Furukawa, T. Yoshinaga, R. Tokunaga and S. Taketani. 1993. Coproporphyrinogen oxidase. Purification, molecular cloning, and induction of mRNA during erythroid differentiation. *J.Biol.Chem.* **268**:21359-21363.
89. Krummel, B. and M.J. Chamberlin. 1992. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Individual complexes halted along different transcription units have distinct and unexpected biochemical properties. *J.Mol.Biol.* **225**:221-237.
90. Kullik, I., H. Hennecke and H.M. Fischer. 1989. Inhibition of *Bradyrhizobium japonicum* *nifA*-dependent *nif* gene activation by oxygen occurs at the NifA protein level and is irreversible. *Arch.Biochem.* **151**:191-197.
91. Kustu, S., E. Santero, J. Keener, D. Popham and D. Weiss. 1989. Expression of  $\sigma^{54}$  (*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiol.Revs.* **53**:367-376.
92. Labes, M. and T. Finan. 1993. Negative regulation of  $\sigma^{54}$ -dependent *dctA* expression by the transcriptional activator DctD. *J.Bacteriol.* **175**:2674-2681.
93. Landshulz, W.H., P.F. Johnson and S.L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759-1764.
94. Laybourn, P.J. and M.E. Dahmus. 1990. Phosphorylation of RNAP polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J.Biol.Chem.* **265**:13165-13173.
95. Lederbur, H., B. Gu, J.III Sojda and B.T. Nixon. 1990. *Rhizobium meliloti* and *Rhizobium leguminosarum* *dctD* gene products bind to tandem sites in an activation sequence located upstream of  $\sigma^{54}$ -dependent *dctA* promoters. *J.Bacteriol.* **172**:3888-3889.
96. Lederbur, H. and B.T. Nixon. 1992. Tandem DctD-binding sites of the *Rhizobium meliloti* *dctA* upstream activating sequence are essential for optimal function despite a 50- to 100-fold difference in affinity for DctD. *Mol.Microbiol.* **6**:3479-3492.
97. Lee, H.S., A. Ishihama and S. Kustu. 1993. The C terminus of the  $\alpha$  subunit of RNA polymerase is not essential for transcriptional activation of  $\sigma^{54}$  holoenzyme. *J.Bacteriol.* **175**:2479-2482.
98. Lee, H.S., F. Narberhaus and S. Kustu. 1993. In vitro activity of NifL, a signal transduction protein for biological nitrogen fixation. *J.Bacteriol.* **175**:7683-7688.
99. Lee, J.H. and T.R. Hoover. 1995. Protein crosslinks studies suggest that *Rhizobium meliloti* C4-dicarboxylic acid transport protein D, a  $\sigma^{54}$  dependent transcriptional activator, interact with  $\sigma^{54}$  and the  $\beta$  subunit of RNA polymerase. *Proc.Natl.Acad.Sci.USA.* **92**:9702-9706.

100. Lee, J.H., D. Scholl, B.T. Nixon and T.R. Hoover. 1994. Constitutive ATP hydrolysis and transcription activation by a stable, truncated form of *Rhizobium meliloti* DCTD, a  $\sigma^{54}$ -dependent transcriptional activator. *J.Biol.Chem.* **269**:20401-20409.
101. Lewin, B. 1994. Building the transcription complex: promoters, factors, and RNA polymerases, p.847. *Genes V*, Oxford University Press, New York.
102. Linn, S.C. and D.S. Luse. 1991. RNAP polymerase II elongation complexes paused after the synthesis of 15- of 35- base transcripts have different structures. *Mol Cell Biol* **11**:1508-1522.
103. Madsen, O., L. Sandal, N.N. Sandal and K.A. Marcker. 1993. A soybean coproporphyrinogen oxidase gene is highly expressed in root nodules. *Plant Mol.Biol.* **23**:35-43.
104. Magasanik, B. 1988. Reversible phosphorylation of an enhancer binding protein regulates the transcription of bacterial nitrogen utilization genes. *Trends Biochem.Sci.* **13**:475-479.
105. Maldak, B.L., N. Larsen, M.J. McCaughey, R. Overbeek, G.J. Olsen, K. Fogel, J. Blandy and C.R. Woese. 1994. The Ribosomal Database Project. *Nucl.Acids Res.* **17**:3485-3487.
106. Mancinelli, R.L. and C.P. McKay. 1988. The evolution of nitrogen cycling. *Origins of Life* **18**:311-325.
107. Martasek, P., J.M. Camadro, M.H. Delfau-Larue, J.B. Dumas, J.J. Motagne, H. de Verneuil, P. Labbe and B. Grandchamp. 1994. Molecular cloning, sequencing, and functional expression of a cDNA encoding human coproporphyrinogen oxidase. *Proc.Natl.Acad.Sci.USA.* **91**:3024-3028.
108. Martínez, E., M.A. Pardo, R. Palacios and M.A. Cevallos. 1985. Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. *J.Gen.Microbiol.* **131**:1779-1786.
109. McClure, W.R., D.K. Hawley, P. Youderian and M.M. Susskind. 1983. DNA determinants of promoter selectivity in *Escherichia coli*. *Cold Spring Harbor Symp.Quant.Biol.* **47**:477-481.
110. Mendoza, A., A. Leija, E. Martínez-Romero, G. Hernández and J. Mora. 1995. The enhancement of ammonium assimilation in *Rhizobium etli* prevents nodulation of *Phaseolus vulgaris*. *Mol.Plant Microbe Interac.* **4**:584-592.
111. Mendoza, A., A. Leija, E. Martínez-Romero, G. Hernández and J. Mora. 1995. The enhancement of ammonium assimilation in *Rhizobium etli* prevents nodulation of *Phaseolus vulgaris*. *Proceedings of the 15th North American Conference on Symbiotic Nitrogen Fixation.*
112. Merrick, M. 1993. In a class of its own - the RNA polymerase sigma factor  $\sigma^{54}$  ( $\sigma^N$ ). *Mol.Microbiol.* **10**:903-909.
113. Merrick, M. and S. Chambers. 1992. The helix-turn-helix motif of  $\sigma^{54}$  is involved in recognition of the -13 promoter region. *J.Bacteriol.* **174**:7221-7226.

114. Merrick, M.J. 1983. Nitrogen control of the *nif* regulon in *Klebsiella pneumoniae*: involvement of the *nta* gene and analogies between *ntc* and *nifA*. EMBO J. 2:39-44.
115. Mettke, I., U. Fiedler and V. Weiss. 1995. Mechanism of activation of a response regulator: interaction of NtrC-P dimers induces ATPase activity. J.Bacteriol. 177:5056-5061.
116. Meyer-Almes, F.J., J. Heumann and D. Porsche. 1994. The structure of the RNA polymerase-promoter complex. DNA-bending-angle by quantitative electrooptics. J.Mol.Biol. 236:1-6.
117. Michiels, J., I. D'hooghe, C. Verreth, H. Pelemans and J. Vanderleyden. 1994. Characterization of the *Rhizobium leguminosarum* biovar phaseoli *nifA* gene, a positive regulator of *nif* gene expression. Arch.Microbiol 161:404-408.
118. Minchin, S.D., S. Austin and R.A. Dixon. 1989. Transcriptional activation of the *Klebsiella pneumoniae nifLA* promoter by NTRC is face-of-the-helix dependent and the activator stabilizes the interaction of sigma 54-RNA polymerase with the promoter. EMBO J. 8:3491-3499.
119. Moore, J.B., S.P. Shlau and L.J. Reitzer. 1993. Alterations of highly conserved residues in the regulatory domain of nitrogen regulator I (NtrC) of *Escherichia coli*. J.Bacteriol. 175:2692-2701.
120. Mora, J. 1995. El metabolismo de carbono en *Rhizobium* y su relación con la fijación de nitrógeno, p.16.Procedimientos del IV Congreso Nacional de la Fijación Biológica del Nitrógeno.
121. Morett, E. 1994. Genetic control of ammonium assimilation and nitrogen fixation in *Klebsiella pneumoniae*: Implications for the study of nitrogen fixation in other organisms. Curr.Topics in Mol.Genet. 2:245-261.
122. Morett, E. and M. Buck. 1988. NifA-dependent in vivo protection demonstrates that the upstream activator sequence of *nif* promoters is a protein binding site. Proc.Natl.Acad.Sci.USA. 85:9401-9405.
123. Morett, E. and M. Buck. 1989. *In vivo* studies on the interaction of RNA polymerase- $\sigma^{54}$  with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. J.Mol.Biol. 210:65-77.
124. Morett, E., W. Cannon and M. Buck. 1988. The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognises the upstream activator sequences of *nif* promoters and can be separated from the positive control function of NifA. Nucl.Acids Res. 16:11469-11488.
125. Morett, E., H.M. Fischer and H. Hennecke. 1991. Influence of oxygen on DNA binding, positive control, and stability of the *Bradyrhizobium japonicum* NifA regulatory protein. J.Bacteriol. 173:3478-3487.
126. Morett, E., R. Kreutzer, W. Cannon and M. Buck. 1990. The influence of the *Klebsiella pneumoniae* regulatory gene *nifL* upon the transcriptional activator protein NifA. Mol.Microbiol. 4:1253-1258.
127. Morett, E., S. Moreno and G. Espín. 1988. Transcription analysis of the three *nifH* genes of *Rhizobium phaseoli* with gene fusions. Mol.Gen.Genet. 213:499-504.

128. **Morett, E. and L. Segovia.** 1993. The  $\sigma^{54}$  bacterial enhancer-binding protein family: Mechanism of action and phylogenetic relationship of their functional domains. *J.Bacteriol.* **175**:6067-6074.
129. **Morris, L., W. Cannon, F. Claverie-Martin, S. Austin and M. Buck.** 1994. DNA distortion and nucleation of local DNA unwinding within sigma-54 ( $\sigma^N$ ) holoenzyme closed promoter complexes. *J.Biol.Chem.* **269**:11563-11571.
130. **Narberhaus, F., H.S. Lee, R.A. Schmitz, L. He and S. Kustu.** 1995. The C-terminal domain of NifL is sufficient to inhibit NifA activity. *J.Bacteriol.* **177**:5078-5087.
131. **Ninfa, A.J. and B. Magasanik.** 1986. Covalent modification of the *glnG* product, NR<sub>I</sub>, by the *glnL* product, NR<sub>II</sub>, regulates the transcription of the *glnALG* operon in *Escherichia coli*. *Proc.Natl.Acad.Sci.USA.* **83**:5909-5913.
132. **Ninfa, A.J., L.J. Reitzer and B. Magasanik.** 1987. Initiation of transcription at the bacterial *glnAp2* promoter by purified *E. coli* components is facilitated by enhancers. *Cell* **50**:1039-1046.
133. **Norel, F. and C. Elmerich.** 1987. Nucleotide sequence and functional analysis of the two *nifH* copies of *Rhizobium* ORS571. *J.Gen.Microbiol.* **133**:1563-1576.
134. **Normand, P. and J. Bousquet.** 1989. Phylogeny of nitrogenase sequences in Frankia and other nitrogen-fixing microorganisms. *J.Mol.Evol.* **29**:436-447.
135. **North, A.K., K.E. Klose, K.M. Stedman and S. Kustu.** 1993. Prokaryotic enhancer-binding proteins reflect eukaryote-like modularity: the puzzle of nitrogen regulatory protein C. *J.Bacteriol.* **175**:4267-4273.
136. **Nunn, C.M. and S. Neidle.** 1995. The structure of chain-linked DNA as a possible model for the transcription bubble. *Nature Struct.Biol.* **2**:1060-1061.
137. **O'brian, M.R., P.M. Kirshbom and R.J. Maler.** 1987. Bacterial heme synthesis is required for expression of the leghemoglobin holoprotein but not the apoprotein in soybean root nodules. *Proc.Natl.Acad.Sci.USA.* **84**:8390-8393.
138. **Olsen, G.J., K.R. Woese and R. Overbeek.** 1994. The winds of (evolutionary) change: Breathing new life into microbiology. *J.Bacteriol.* **176**:1-6.
139. **Ow, D.W. and F. Ausubel.** 1983. Regulation of nitrogen metabolism genes by *nifA* gene product in *Klebsiella pneumoniae*. *Nature* **301**:307-313.
140. **Parkhill, J., A.Z. Ansari, J.G. Wright, N.L. Brown and T.V. O'Halloran.** 1993. Construction and characterization of a mercury-independent MerR activator (MerRAC): transcriptional activation in the absence of Hg(II) is accompanied by DNA distortion. *EMBO J* **12**:413-421.
141. **Parkinson, J.S. and E.C. Kofoid.** 1992. Communication modules in bacterial signaling proteins. *Ann.Rev.Genet.* **26**:71-112.
142. **Payne, J.M., P.J. Laybourn and M.E. Dahmus.** 1989. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit II $\alpha$ . *J.Biol.Chem.* **264**:19621-19629.

143. Piñero, D., E. Martínez and R.K. Selander. 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar phaseoli. *Appl. Environ. Microbiol.* **54**:2825-2832.
144. Popham, D., D. Szeto, J. Keener and S. Kustu. 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**:629-635.
145. Porter, S.C., A.K. North, A.B. Wedel and S. Kustu. 1993. Oligomerization of NTRC at the *glnA* enhancer is required for transcriptional activation. *Genes & Devel.* **7**:2258-2273.
146. Quinto, C., H. de la Vega, M. Flores, J. Leemans, M.A. Cevallos, M.A. Pardo, R. Azpiroz, M.deL. Girard, E. Calva and R. Palacios. 1985. Nitrogenase reductase: A functional multigene family in *Rhizobium phaseoli*. *Proc.Natl.Acad.Sci.USA.* **82**:1170-1174.
147. Rees, W.A., R.W. Keller, J.P. Vesenka, G. Yang and C. Bustamante. 1993. Evidence of DNA bending in transcription complexed imaged by scanning force microscopy. *Science* **260**:1646-1649.
148. Reiter, W.D., U. Hudepohl and W. Zillig. 1990. Mutational analysis of an archaeobacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. *Proc.Natl.Acad.Sci.USA.* **87**:9509-9513.
149. Reitzer, L.J. and B. Magasanik. 1986. Transcription of *glnA* in *E.coli* is stimulated by activator bound to sites far from the promoter. *Cell* **45**:785-792.
150. Roberts, G.P., W.T. Leps, L.E. Silver and W.J. Brill. 1980. Use of two-dimensional polyacrylamide gel electrophoresis to identify and classify *Rhizobium* strains. *Appl.Environ.Microbiol.* **39**:414-422.
151. Romero, D., P.W. Singleton, L. Segovia, E. Morett, B.B. Bohlool, R. Palacios and G. Dávila. 1988. Effect of naturally occurring *nif* reiterations on symbiotic effectiveness in *Rhizobium phaseoli*. *Appl.Environ.Microbiol.* **54**:848-850.
152. Ronson, C.W., P.M. Astwood, B.T. Nixon and F.M. Ausubel. 1987. Deduced products of C4-dicarboxylate transport regulatory genes of *Rhizobium leguminosarum* are homologous to nitrogen regulatory gene products. *Nucl.Acids Res.* **15**:7921-7934.
153. Rouvière, P.E., A. de las Peñas, J. Mencias, C.Z. Lu, K. Rudd and C.A. Gross. 1995. *rpoE*, the gene encoding the second heat-shock sigma factor,  $\sigma^E$ , in *Escherichia coli*. *EMBO J* **14**:1032-1042.
154. Rowlands, T., P. Baumann and S.P. Jackson. 1994. The TATA-binding protein: A general transcription factor in eukaryotes and archaeobacteria. *Science* **264**:1326-1329.
155. Sanders, D.A., B.L. Gillece-Castro, A.L. Burlingame and D.E. Koshland. 1992. Phosphorylation site of NtrC, a protein phosphatase whose covalent intermediate activates transcription. *J.Bacteriol.* **174**:5117-5122.
156. Santero, E., T.R. Hoover, A.K. North, D.K. Berger, S.C. Porter and S. Kustu. 1992. Role of integration host factor in stimulating transcription from the sigma

- 54-dependent *nifH* promoter. *J.Mol.Biol.* **227**:602-620.
157. Sasse-Dwight, S. and J.D. Gralla. 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor  $\sigma^{54}$ . *Cell* **62**:945-954.
158. Sawadogo, M. and A. Sentenac. 1990. RNA polymerase B (II) and general transcription factors. *Annu.Rev.Biochem.* **59**:711-754.
159. Seehra, J.S., P.M. Jordan and M. Akhtar. 1982. Anaerobic and aerobic coproporphyrinogen III oxidases of *Rhodopseudomonas sphaeroides*. *Biochem.J.* **269**:709-718.
160. Segovia, L., J.P.W. Young and E. Martínez-Romero. 1993. Reclassification of american *Rhizobium leguminosarum* biovar phaseoli type I strains as *Rhizobium etli* sp. nov. *Int.J.Syst.Bacteriol.* **43**:374-377.
161. Shiau, S.P., P. Chen and L.J. Reitzer. 1993. Effect of insertions and deletions in *glnG* (*ntrC*) of *Escherichia coli* on nitrogen regulator I-dependent DNA binding and transcriptional activation. *J.Bacteriol.* **175**:190-199.
162. Simpson, R.B. 1979. The molecular topography of RNA polymerase-promoter interaction. *Cell* **18**:277-285.
163. Soberón, M., G.R. Aguilar and F. Sánchez. 1993. *Rhizobium phaseoli* cytochrome c-deficient mutant induces empty nodules on *Phaseolus vulgaris* L. *Mol.Microbiol.* **8**:159-166.
164. Souillard, N., M. Magot, O. Possot and L. Sibold. 1988. Nucleotide sequence of regions homologous to *nifH* (nitrogenase Fe protein) from the nitrogen-fixing Archaeobacteria *Methanococcus thermolithotropus* and *Methanobacterium ivanovii*. Evolutionary implications. *J.Mol.Evol.* **27**:65-76.
165. Souillard, N. and L. Sibold. 1986. Primary structure and expression of a gene homologous to *nifH* (nitrogenase Fe protein) from the archaeobacterium *Methanococcus voltae*. *Mol.Gen.Genet.* **203**:21-28.
166. Sprent, J.I. and J.A. Raven. 1992. Evolution of nitrogen fixing symbiosis, p.461. In G. Stacey, H.J. Evans and R.H. Burris (ed.), *Biological Nitrogen Fixation*, Chapman and Hall, New York.
167. Stratagene, 1994. Catálogo de Productos,
168. Su, T.T. and W.R. McClure. 1994. Selective binding of *Escherichia coli* RNA polymerase to topoisomers of minicircles carrying the TAC16 and TAC17 promoters. *J.Biol.Chem.* **269**:13511-13521.
169. Su, W., S. Porter, S. Kustu and H. Echols. 1990. DNAP-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc.Natl.Acad.Sci.USA.* **87**:5504-5508.
170. Thöny, B. and H. Hennecke. 1989. The -24/-12 promoter comes of age. *FEMS Microbiol.Rev.* **63**:341-358.
171. Thöny-Meyer, L., D. Ritz and H. Hennecke. 1994. Cytochrome c biogenesis in bacteria: A possible pathway begins to emerge. *Mol.Microbiol.* **12**:1-9.

172. Tintut, Y., C. Wong, Y. Jiang, M. Hsieh and J.D. Gralla. 1994. RNA polymerase binding using a strongly acidic hydrophobic-repeat region of  $\sigma^{54}$ . Proc. Natl. Acad. Sci. USA. **91**:2120-2124.
173. Tripatara, A. and P.L. de Haseth. 1993. A new start site for *Escherichia coli* RNA polymerase at an engineered short region of non-complementarity in double-stranded DNA. J.Mol.Biol. **233**:349-358.
174. Troup, B., C. Hungerer and D. Jahn. 1995. Cloning and characterization of the *Escherichia coli hemN* gene encoding the oxygen-independent coproporphyrinogen III oxidase. J.Bacteriol. **177**:3326-3331.
175. Troup, B., M. Jahn, C. Hungerer and D. Jahn. 1994. Isolation of the *hemF* operon containing the gene for the *Escherichia coli* aerobic coproporphyrinogen III oxidase by *in vivo* complementation of a yeast HEM13 mutant. J.Bacteriol. **176**:673-680.
176. Valderrama, B. 1989. Búsqueda y caracterización genética de sistemas alternativos para la nitrogenasa en Rhizobaceas. Tesis de Licenciatura en Investigación Biomédica Básica.
177. Valderrama, B. 1993. Análisis estructural y de expresión de los genes reiterados para la nitrogenasa en *Rhizobium phaseoli*. Tesis de Maestría en Investigación Biomédica Básica.
178. Vázquez, M., A. Dávalos, A. de las Peñas, F. Sánchez and C. Quinto. 1991. Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv. phaseoli strains. J.Bacteriol. **173**:1250-1258.
179. Vázquez, M., O. Santana and C. Quinto. 1993. The NodI and NodJ proteins from *Rhizobium* and *Bradyrhizobium* strains are similar to capsular polysaccharide secretion proteins from gram-negative bacteria. Mol.Microbiol. **8**:369-377.
180. Vieira, J. and J. Messing. 1991. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. Gene **100**:189-194.
181. Villalobos, M.A., N. Nava, M. Vázquez and C. Quinto. 1994. Nucleotide sequence of the *Rhizobium etli nodS* gene. Gene **150**:201-202.
182. Vinson, C.R., P.B. Sigler and S.L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246**:911-916.
183. von Hippel, P.H., T.D. Yager and S.C. Gill. 1992. Quantitative aspects of the transcription cycle in *Escherichia coli*, p.179. In S.L. McKnight and K.R. Yamamoto (ed.), Transcriptional regulation, Cold Spring Harbor Laboratory Press, New York.
184. Wang, J.T., A. Syed, M. Hsieh and J.D. Gralla. 1995. Converting *Escherichia coli* RNA polymerase into an enhancer-responsive enzyme: Role of an NH<sub>2</sub>-terminal leucine patch in  $\sigma^{54}$ . Science **270**:992-994.
185. Wedel, A. and S. Kustu. 1995. The bacterial enhancer-binding protein NTRC is a molecular machine: ATP hydrolysis is coupled to transcriptional activation. Genes & Devel. **9**:2042-2052.
186. Wedel, A., D.S. Weiss, D. Popham, P. Droge and S. Kustu. 1990. A bacterial

enhancer functions to tether a transcriptional activator near a promoter. *Science* **248**:486-490.

187. **Weeks, J.R., S.E. Hardin, J. Shen, J.M. Lee and A.L. Greenleaf.** 1993. Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. *Genes & Devel.* **7**:2329-2344.

188. **Weinstein, M., R.C. Roberts and D.R. Helinski.** 1992. A region of the broad-host-range plasmid RK2 causes stable in planta inheritance of plasmids in *Rhizobium meliloti* cells isolated from alfalfa root nodules. *J.Bacteriol.* **174**:7486-7489.

189. **Weiss, D.S., J. Batut, K.E. Klose, J. Keener and S. Kustu.** 1991. The phosphorylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. *Cell* **67**:155-167.

190. **Weiss, D.S., K.E. Klose, T.R. Hoover, et al.** 1992. Prokaryotic transcriptional enhancers, p.667. In S.L. McKnight and K.R. Yamamoto (ed.), *Transcriptional regulation*, Cold Spring Harbor Laboratory Press, New York.

191. **Wettach, J., H.P. Gohl, H. Tschochner and M. Thomm.** 1995. Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. *Proc.Natl.Acad.Sci.USA.* **92**:472-476.

192. **Wong, C. and J.D. Gralla.** 1992. A role for the acidic trimer repeat region of transcription factor  $\sigma^{54}$  in setting the rate and temperature dependence of promoter melting in vivo. *J.Biol.Chem.* **267**:24762-24768.

193. **Wong, C., Y. Tintut and D. Gralla.** 1994. The domain structure of sigma 54 as determined by analysis of a set of deletion mutants. *J.Mol.Biol.* **236**:81-90.

194. **Wootton, J.C. and M. Drummond.** 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Prot.Eng.* **2**:535-543.

195. **Xu, K., J. Delling and T. Elliott.** 1992. The genes required for heme synthesis in *Salmonella typhimurium* include those encoding alternative functions for aerobic and anaerobic coproporphyrinogen oxidation. *J.Bacteriol.* **174**:3953-3963.

196. **Xu, K. and T. Elliott.** 1993. An oxygen-independent coproporphyrinogen oxidase encoded by the *hemF* gene of *Salmonella typhimurium*. *J.Bacteriol.* **175**:4990-4999.

197. **Xu, K. and T. Elliott.** 1994. Cloning, DNA sequence, and complementation analysis of the *Salmonella typhimurium hemN* gene encoding a putative oxygen-independent coproporphyrinogen III oxidase. *J.Bacteriol.* **176**:3196-3203.

198. **Young, J.P.W.** 1992. Phylogenetic classification of nitrogen-fixing organisms, p.43. In G. Stacey, H.J. Evans and R.H. Burris (ed.), *Biological Nitrogen Fixation*, Chapman and Hall, New York.

199. **Young, J.P.W., H.L. Downer and B.D. Eardly.** 1991. Phylogeny of the phototrophic *Rhizobium* strain BTA11 by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. *J.Bacteriol.* **173**:2271-2277.

200. **Young, R.A.** 1991. RNA polymerase II. *Annu.Rev.Biochem.* **60**:689-715.

201. Zagorec, M., J.M. Buhler, I. Treich, T. Keng, L. Guarente and R. Rabbe-Bois. 1988. Isolation, sequence, and regulation by oxygen of the yeast HEM13 gene encoding for coproporphyrinogen oxidase. *J.Biol.Chem.* **263**:9718-9724.

# **TABLAS Y FIGURAS**

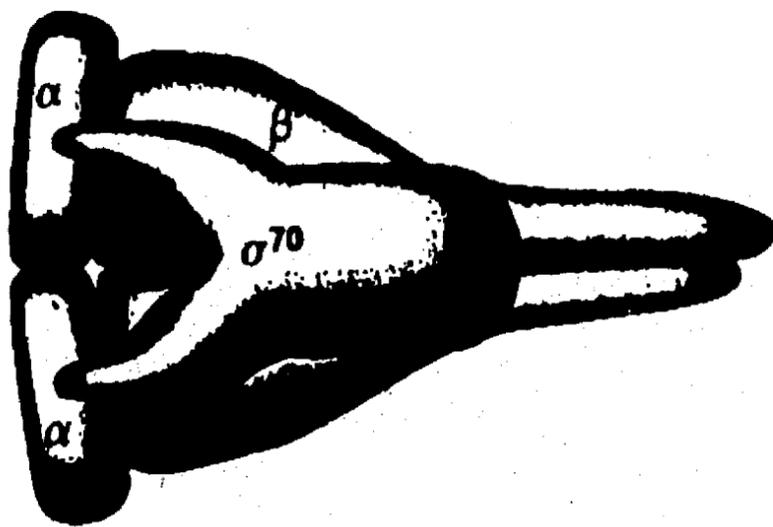


Figura 1 - Estructura propuesta por Ishihama (81) para la ARN polimerasa asociada al factor  $\sigma^{70}$

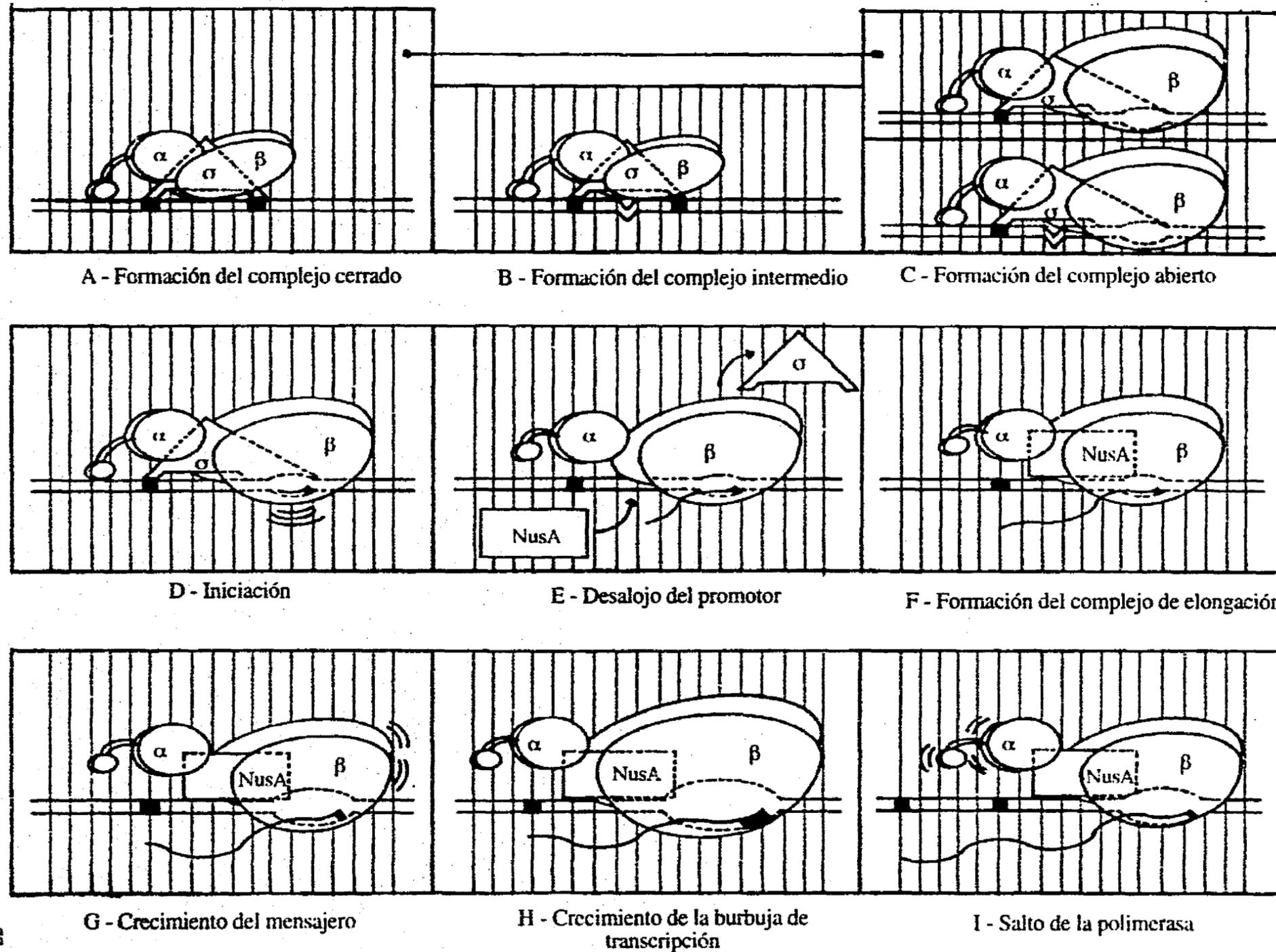


Figura 2 - Etapas del inicio de la transcripción por  $E\sigma^{70}$  en *E. coli*

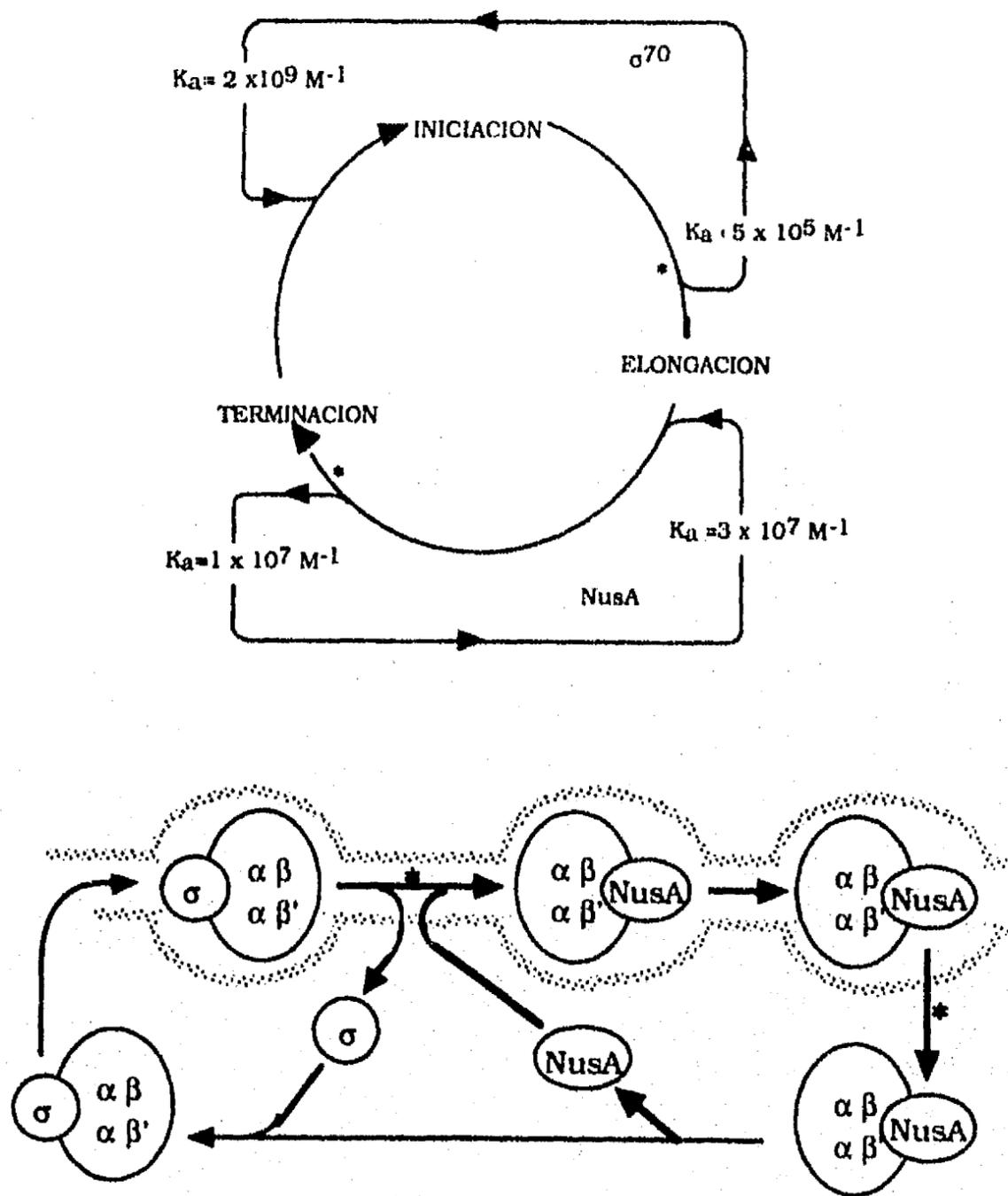


Figura 3 - Dos representaciones del ciclo de la transcripción para  $E\sigma^{70}$

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA

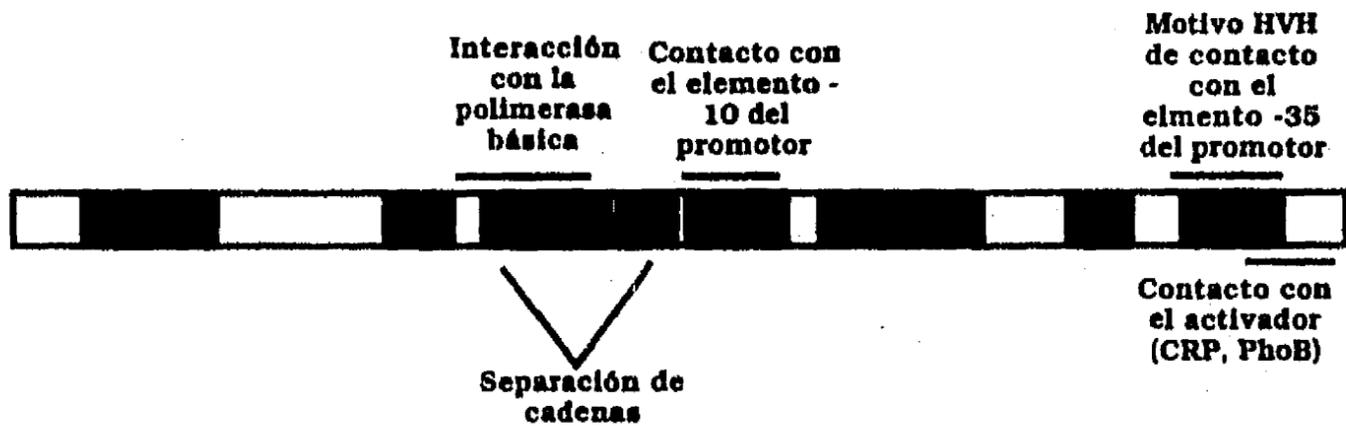


Figura 4 - Estructura de las proteínas pertenecientes a la familia de  $\sigma 70$  y algunos determinantes de función identificados

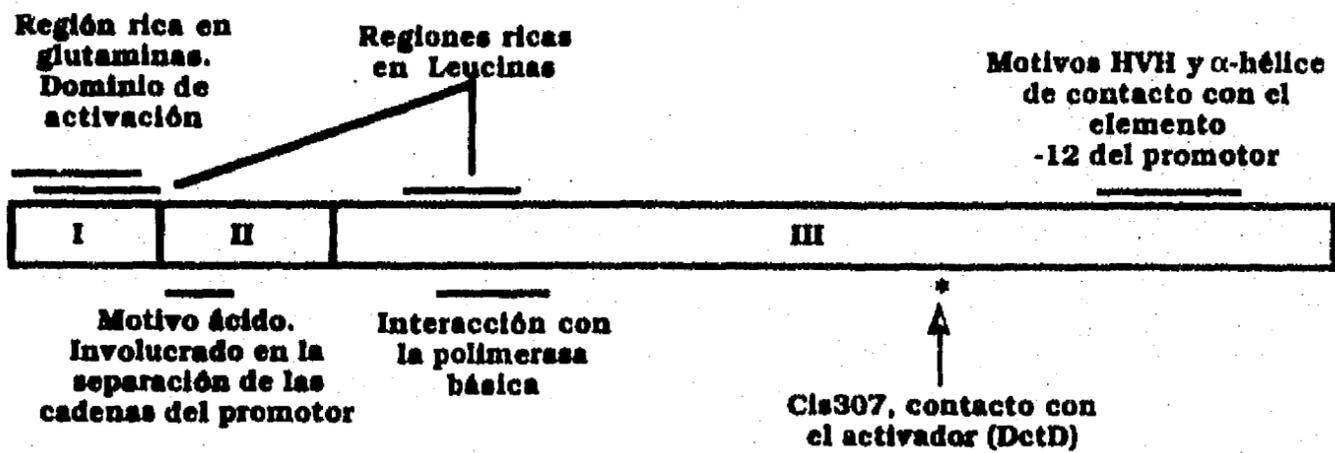


Figura 5 - Estructura de los dominios del factor de iniciación  $\sigma 54$  y algunos determinantes de función identificados.

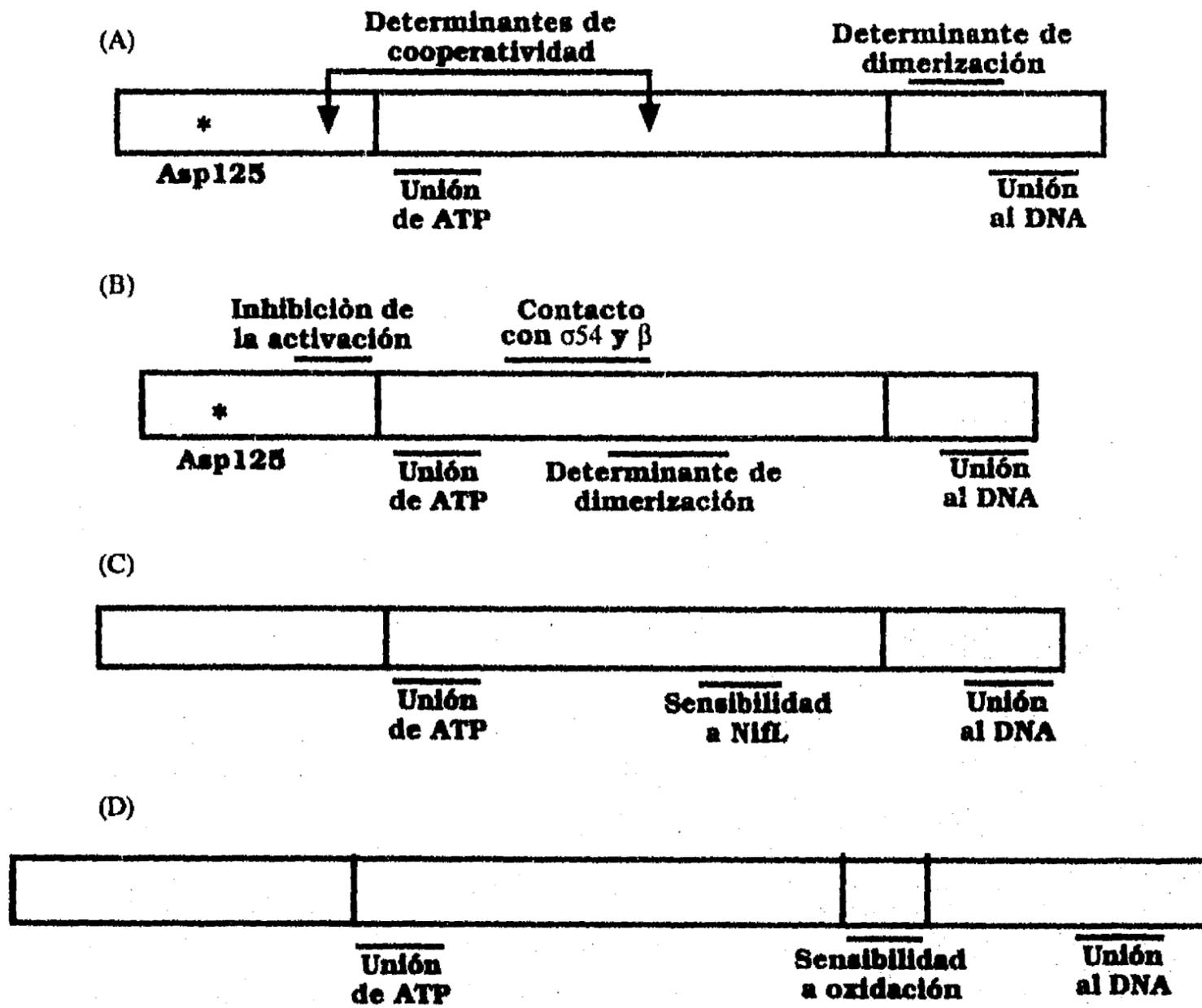
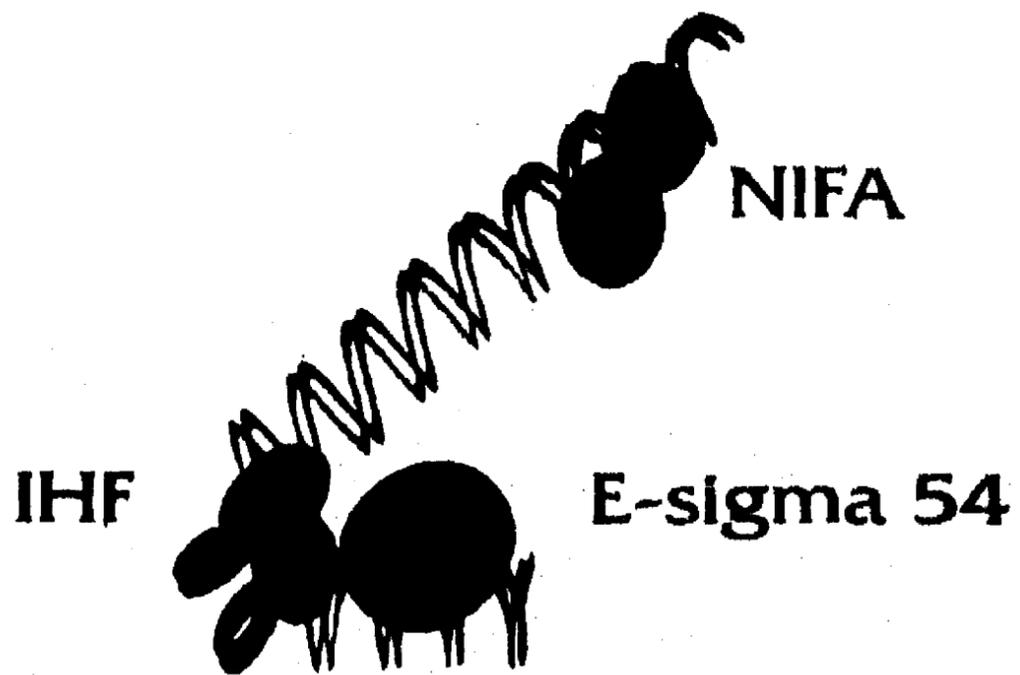


Figura 6 - Estructura modular de las proteínas activadoras de promotores dependientes de  $E\sigma_{54}$ . (A) Ntrc, (B) DctD, (C) NifA de *K. pneumoniae* y (D) NifA de *Rhizobium*. Todas las proteínas muestran la localización de algunos determinantes de función identificados.



**Figura 7 - Modelo actual que representa la arquitectura del complejo de iniciación formado por  $E\sigma 54$  unida al promotor y la proteína activadora NifA unida a sus secuencias específicas. La proteína IHF impone un dobléz al DNA interno que facilita la activación a distancia del promotor.**

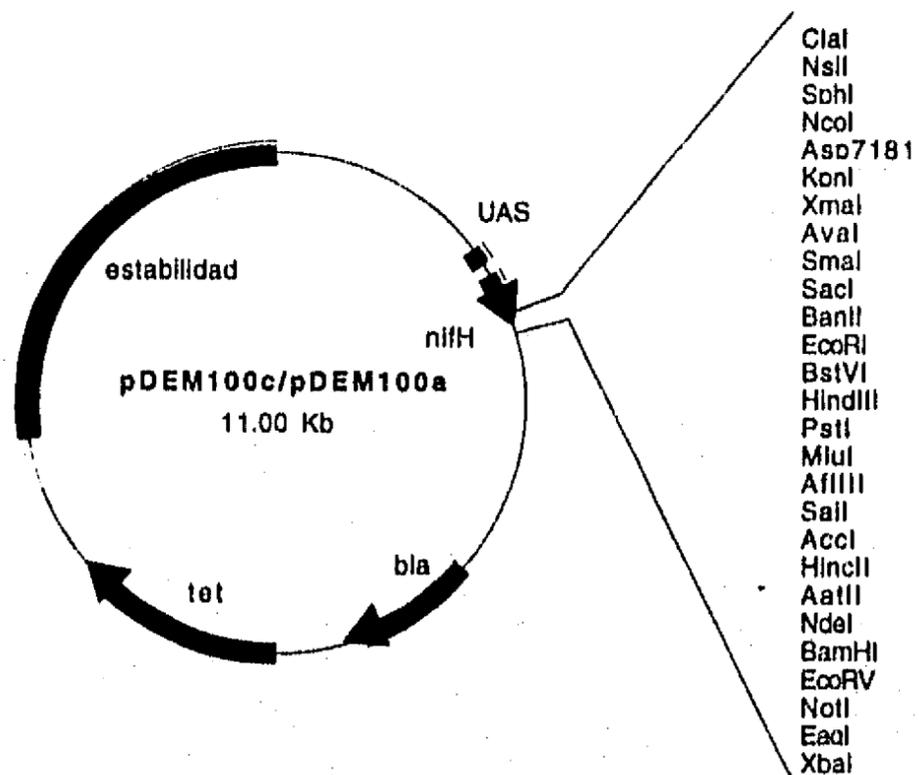
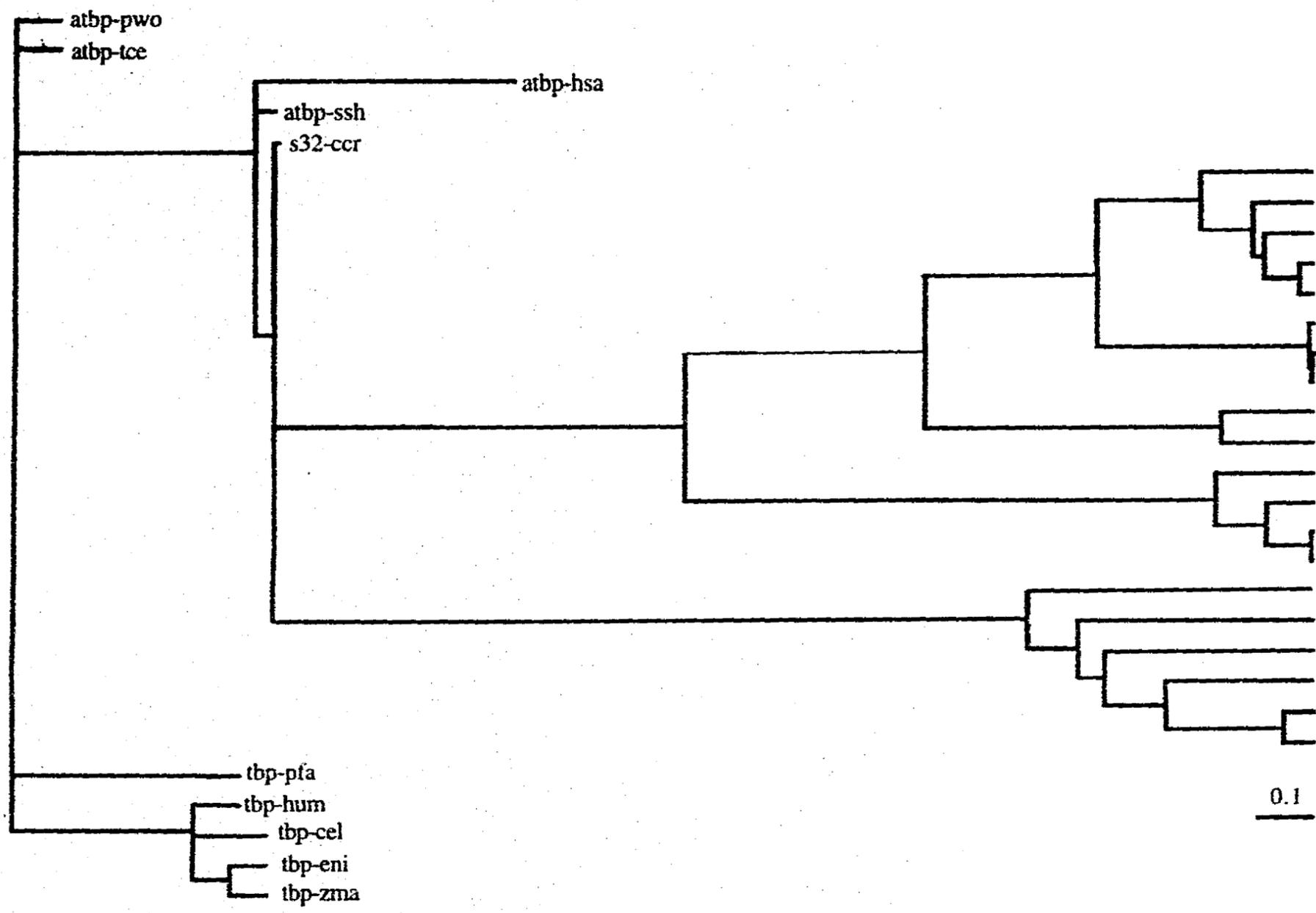


Figura 8 - Diseño de los vehículos de expresión específicos de simbiosis y microaerobiósisis derivados de las regiones promotoras de los genes *nifHa* y *nifHc* de *Rhizobium etli*.



83

FIGURA 9 - Relaciones filogenéticas entre los factores de iniciación transcripcional eubacterianos, arqueales y eucarióticos descritos en la Tabla 3. El árbol se obtuvo con el algoritmo de Neighbor-Joining.

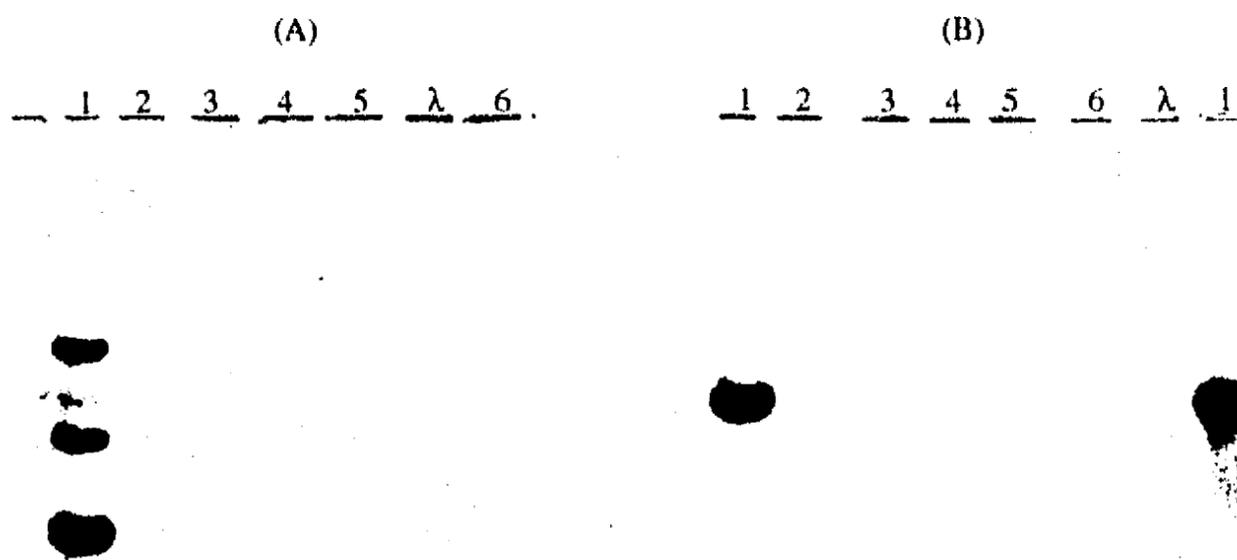


Figura 10 - Perfil de hibridización de ADN genómico obtenido de las siguientes especies y cortado con la enzima de restricción BamHI: 1 *R. etli* CE3, 2 *R. loti* NZP2230, 3 *B. loti* NZP2037, 4 *R. fredii* USDA193, 5 *R. fredii* USDA194, 6 *R. meliloti* Rm1021. Detectores utilizados: Panel (A) gene *nifD\** y extremo amino terminal de ORF251. Panel (B) ORF251, excepto el extremo amino terminal.

agaaatcggcgcccacacttatgagcaatggctgagaagccagggaactgatgacgctgt  
 ctgctctatctccacattccattctgccgctcgatgtgctggtattgtgggtttccgac  
 tagcatcactcgtcgggacactttgataaccgattatctagcaatcgtgcgcgaggaaat  
 ccgcctggtcgccgagcaatgccgcaggcactctccgtgggtgacgtgcactttggcggc  
 MetProGlnAlaLeuSerValGlyAspValHisPheGlyGly  
 ggatccccgaccattatgccatcggcagacttcctgtcgtaatggaactcctgcccggc  
 GlySerProThrIleMetProSerAlaAspPheLeuSerLeuMetGluLeuLeuArgGly  
 cgttttgcgcttgagcggaggtgcaaccattgccggttgaggtcgacccgcgcacgttcacc  
 ArgPheAlaLeuGluArgGlyAlaThrIleAlaValGluValAspProArgThrPheThr  
 accgatatggccgaagccctagaagaaccgggtgtgaatcgcgcaagcgtcggtgtgcag  
 ThrAspMetAlaGluAlaLeuGluArgThrGlyValAsnArgAlaSerValGlyValGln  
 agcttcgatcccgtcgtacaaaaagcgatcaaccggattcagagcgaggcgaagtgatg  
 SerPheAspProValValGlnLysAlaIleAsnArgIleGlnSerGluAlaGlnValMet  
 accgccgtcgaaaacctccgcctgtatggggctcaggcgtatcaatttcgacttgatggtc  
 ThrAlaValGluAsnLeuArgLeuTyrGlyValArgArgIleAsnPheAspLeuMetPhe  
 ggtttgccgaaccaaaccgtccagtcctgtctcgagagcgtatgtagctattgcgatg  
 GlyLeuProAsnGlnThrValGlnSerCysLeuGluSerAlaMetLeuAlaIleAlaMet  
 cgtcccgaaccgctcgcggttttcggttattcccacgttccatcttttcgaaaaaatcag  
 ArgProAspArgLeuAlaValPheGlyTyrSerHisValProSerPheArgLysAsnGln  
 cgcttgattgacacagcagcaactgcccgatatagccgcgagctgagcaggcctcagcc  
 ArgLeuIleAspThrAlaAlaLeuProAspIleAlaAlaArgAlaGluGlnAlaSerAla  
 atggccgatacgttagttgcagcaggctatctgcaaatgggtctcgaccattttgctttg  
 MetAlaAspThrLeuValAlaAlaGlyTyrLeuGlnIleGlyLeuAspHisPheAlaLeu  
 ccggatgacgagcttgcaatagcacagagagctgggtcgtctgcgccgaaattcgcttggt  
 ProAspAspGluLeuAlaIleAlaGlnArgAlaGlyArgLeuArgArgAsnSerLeuGly  
 tactctgccgaaacctgctcaactgtcatcggtttgggcccgtccgccattggctcgttgcg  
 TyrSerAlaGluThrCysSerThrValIleGlyLeuGlyArgProProLeuValValAla  
 gccacgggttacggtcaaaacgatctcacgcaaagctgttataaccggcacatag  
 AlaThrValThrPheLysThrIleSerArgLysAlaValIleThrGlyThrEnd

Figura 11 - Secuencia nucleotídica y secuencia de aminoácidos deducida del ORF251.  
 La porción de la región regulatoria similar a un promotor dependiente de  $\sigma^F$  se  
 indica con un marco.

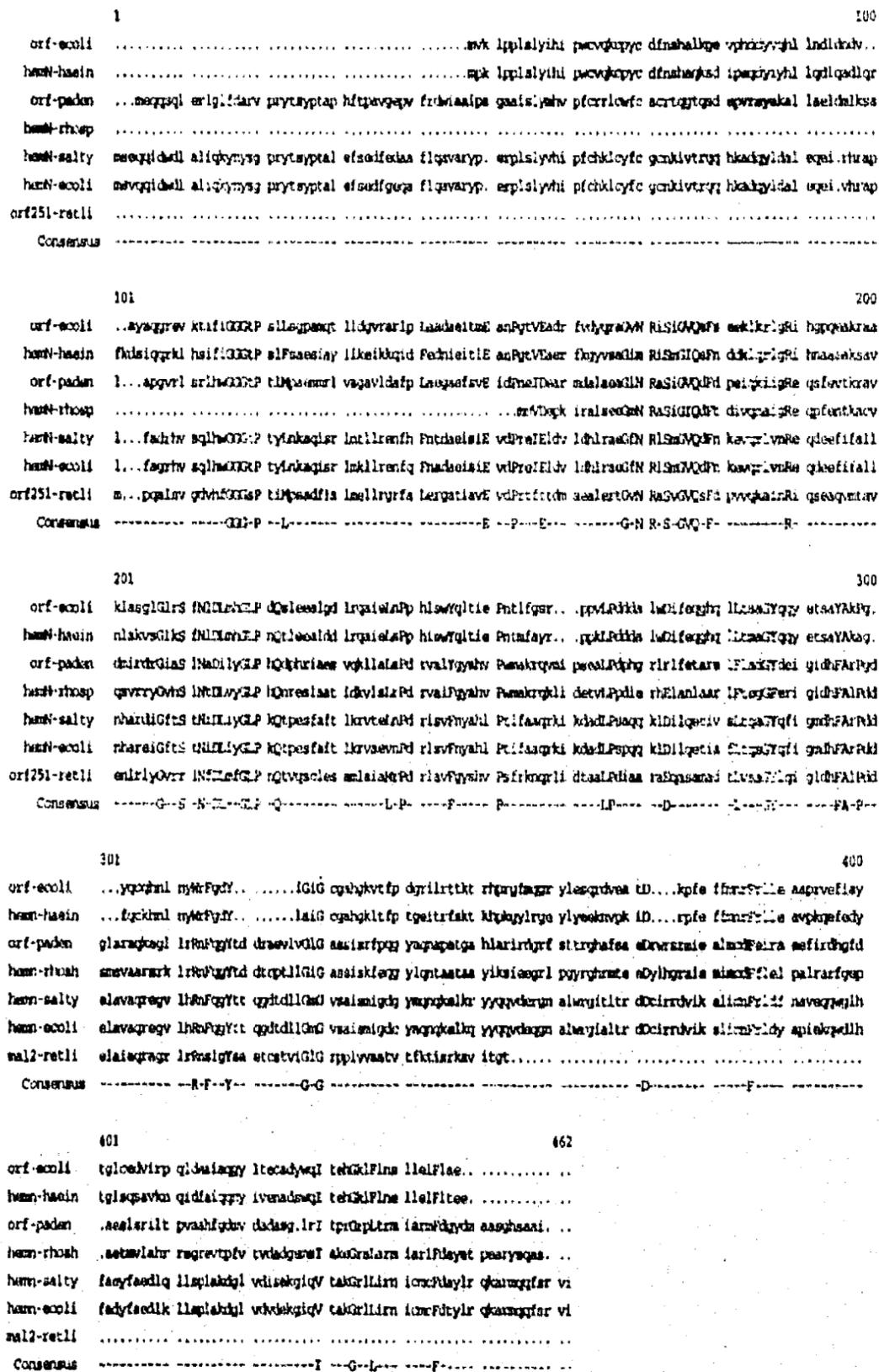


Figura 12 - Alineamiento de secuencias similares a *hemN* provenientes de los siguientes organismos: *E. coli*, *H. influenzae*, *P. denitrificans*, *R. sphaeroides*, *S. thyphimurium*, y *R. etli*.

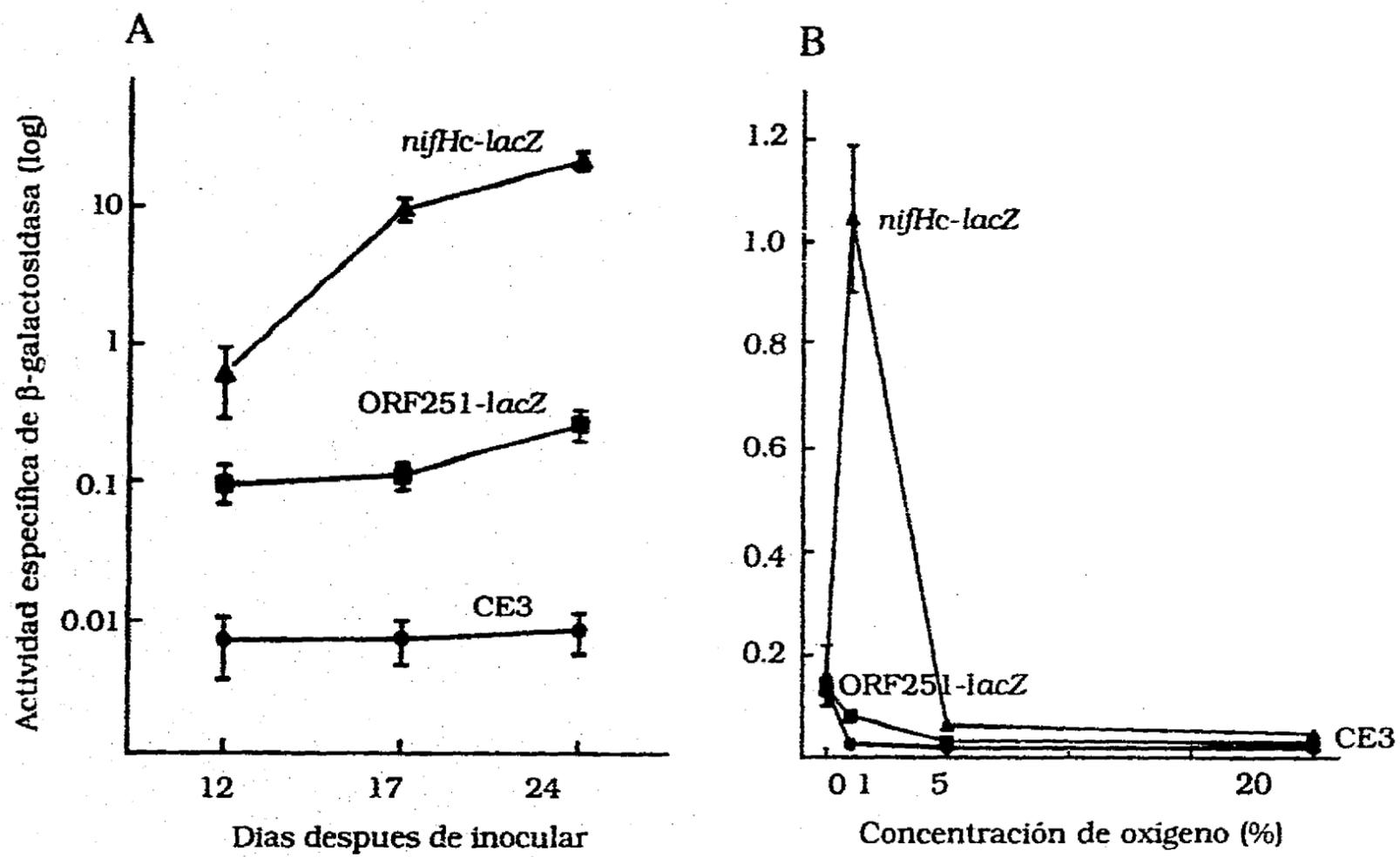


Figura 13 - Expresión del ORF251 en nódulos de frijol (Panel A) y en cultivo microaeróbico (Panel B). La actividad específica de  $\beta$ -galactosidasa está expresada en nmol ONPG/min mg de proteína.

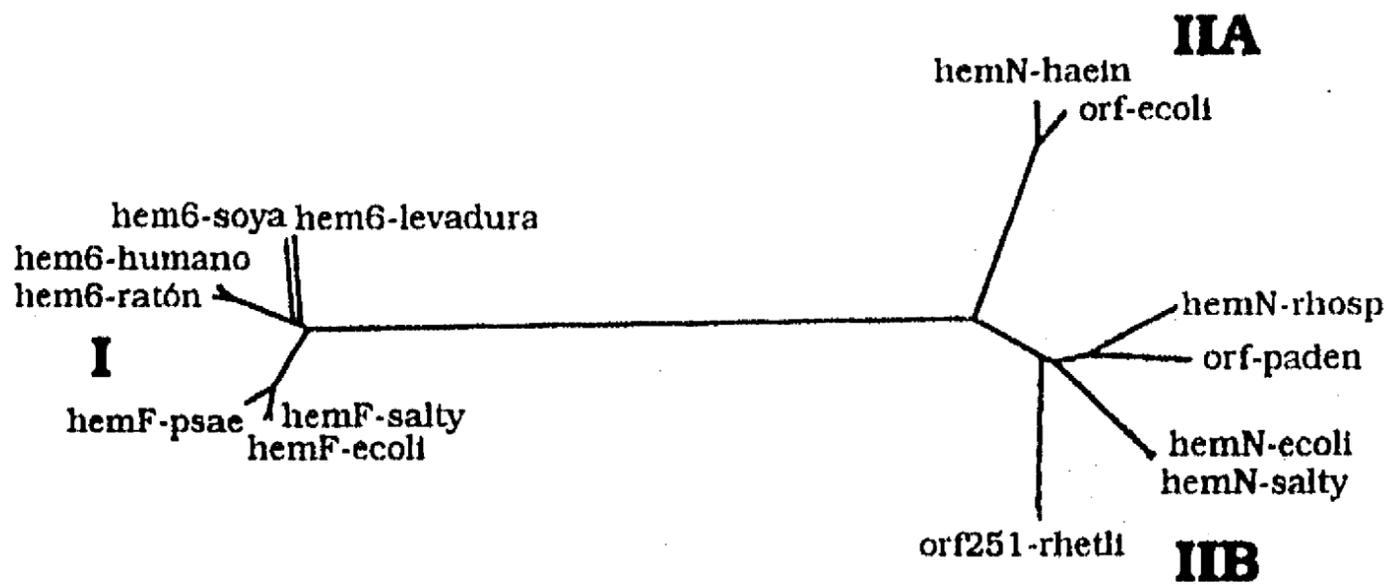


Figura 14 - Relaciones filogenéticas inferidas para 14 secuencias similares a los genes *hemN* y *hemF*. El análisis se realizó utilizando el algoritmo de UPGMA.

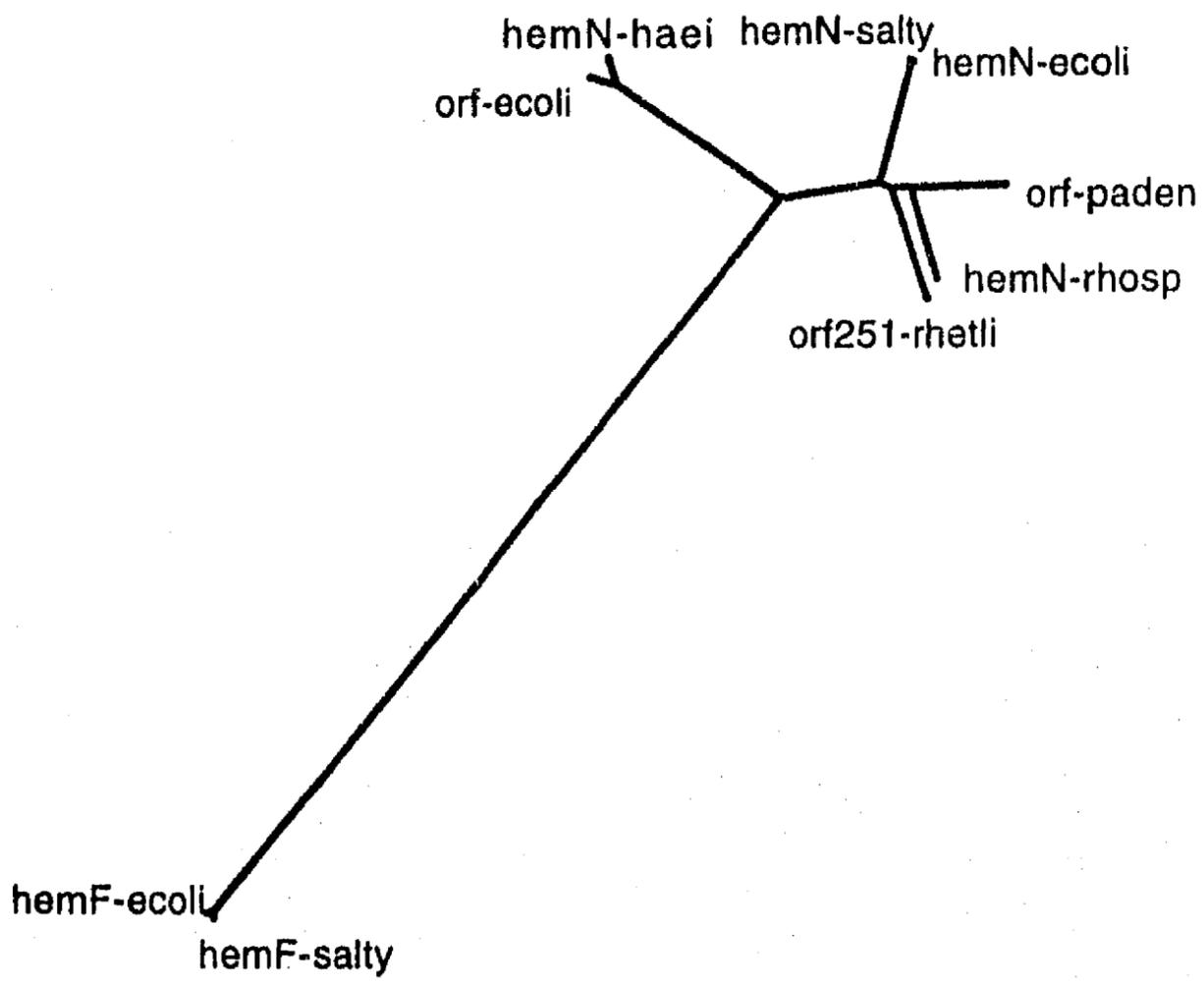


Figura 15 - Relaciones filogenéticas inferidas para 9 secuencias similares a los genes *hemN* y *hemF* en bacterias. El algoritmo utilizado fué UPGMA.

```

1
hemF-salty .....mkpdah hvkqflir ddiqqlsav
hemF-ecoli .....mkpdah qvkeflin dticqqltav
hemN-salty mseqqidwdl aliqkynysg prytsyptal efsedfedaa flqavarype rplslyvhip fchklyfcg cnkivtrqgh kadqldale qeirhraplif
hemN-ecoli msvqqidwdl aliqkynysg prytsyptal efsedfgeqa flqavarype rplslyvhip fchklyfcg cnkivtrqgh kadqldale qeivhraplif
orf-ecoli .....mvkl pplslyhip wcvqkpcygd fnshalkgev phdcvqh ndldndvaya
Consensus -----S-----

101
hemF-salty dganfvedsw rreaqSGGrs rvlmggife qagvnfshvh gdampa.sat ahrpelagrs beamg.....vslvnpfn pyip...tsh anvrffiaek
hemF-ecoli dgaefvedsw qreaqSGGrs rvlmggufe qagvnfshvh geampa.sat ahrpelagrs beamg.....vslvnpfn pyvp...tsh anvrffiaek
hemN-salty adrhvsqllhw .....SGGp tylnkaqisr lmtllrenfh fntdaeisie vdpreieldv dlhraegfn rlnagvqdfn kevqrlvnre qdeefifall
hemN-ecoli agrhvsqllhw .....SGGp tylnkaqisr lmkllrenfq fnadaeisie vdpreieldv dlhraegfn rlnagvqdfn kevqrlvnre qdeefifall
orf-ecoli qgrevktifi .....SGGp slsgpamt lldgvrarlp laadaeitme anpgtveadr vdyqragvn rlsigvysfs eeklkrigri hgpqeakraa
Consensus -----SGG-----S-----

201
hemF-salty pgnlpvwwfg ggfDLpypf feedavwhr tardlcqpfq ddvypkkw cddyffkhr neqrgvgglf fddlntbfd hcFdfMqavg ngyraylpi
hemF-ecoli pgnlpvwwfg ggfDLpfpf feedaihwhr tardlclpfg edvypkkw cdeyffkhr neqrgiggif fddlntbfd rcFafMqavg kgyraylpi
hemN-salty nhrdigfts tniDL.iyglkqtpesfa ftlkrvteln pdrlsvnya hlp.tlfaaq rk..... ikdadlfaaq qkLdiLqeti vslragyqf
hemN-ecoli nhrdigfts tniDL.iyglkqtpesfa ftlkrvaevn pdrlsvnya hlp.tlfaaq rk..... ikdadlspq qkLdiLqeti aflragyqf
orf-ecoli klsglgrs fniDL.mhglodqsleaal gdlrqaieln pphlswilt iepntlpsr pp.....v.....ladd alwdfeggh qlragyqg
Consensus -----DL-----S-----

301
hemF-salty verrkamwvq erernfqlyr rgryveFnd .....vwdrgtl flqtggrte silmsmpplv rweydwqpea gspeaalsef iqvrdwi.....
hemF-ecoli verrkamaYg erernfqlyr rgryveFnd .....vwdrgtl flqtggrte silmsmpplv rweydyqpkd gspeaalsef ikvrdv.....
hemN-salty igmd...hFa rpddehvaq regvlhRng gyttqgdtl lmgvsaism igdgyaqnqk elkryyqqvd ergnalwrgi tltRdcirr dvikalicnf
hemN-ecoli igmd...hFa rpddehvaq regvlhRng gyttqgdtl lmgvsaism igdcyagnqk elkryyqqvd eqgnalwrgi altRdcirr dvikslicnf
orf-ecoli yets...aYa kpgyq....c qhnllywfp dy.....ifggahgk vtfpdgrilr tktrhrprf mqq....ryl esqrdveatd kpfefmrf
Consensus -----S-----

401
hemF-salty .....
hemF-ecoli .....
hemN-salty rldfnaveqg wglhfaeyfa edl...qlls plakdglvdi sekgiqvtak grillirnicm cfdaylrqka rmqqfsrvi
hemN-ecoli rldyapieqg wdlhfadyfa edl...klla plakdglvdv dekgiqvtak grillirnicm cfdaylrqka rmqqfsrvi
orf-ecoli rllaaprve fiaytg..lc edvirpqlde aiaqgyltec ady.wqiteh gklflnsl..leiflae..
Consensus -----

479

```

Figura 16- Alineamiento de secuencias similares a *hemN* y *hemF* provenientes de *S. typhimurium* y *E. coli*.

Tabla 1 - Factores de iniciación de la transcripción pertenecientes a la familia de  $\sigma^{70}$  y las secuencias de ADN a las que se unen específicamente.

Sigma	Proteína	Organismo	Región -35	Región -10
$\sigma^{70}$	RpoD	<i>E. coli</i>	TTGACA	TATAAT
$\sigma^{32}$	RpoH	<i>E. coli</i>	TCTC-CCCCTTGAA	CCCCAT-TA
$\sigma^{38}$	RpoS	Varios	GTTAAGC*	TATACT*
$\sigma^{24}$	RpoE	<i>E. coli</i>	GAACTT	TCTGA
$\sigma^A$	sigA	<i>B. subtilis</i>	TTGACA	TATAAT

\* Dependiendo del promotor estudiado se le ha atribuido diferente relevancia a uno u otro elemento (Mol Microbiol 1995, 18:841; Nucl Acid Res 1995, 23:827; Ann Rev Microbiol 1994, 48:53)

Tabla 2 - Secuencias promotoras identificadas positivamente en arqueobacterias por medio del mapeo de inicio de la transcripción y de cambios específicos de secuencia.

Organismo	Gene	-25	+1
<i>M. thermolithotrophicus</i>	<i>nifH1</i>	TATA	TTGT
<i>M. thermoautotrophicum</i>	<i>purE</i>	ATTAAATA	ND
<i>S. shibatae</i>	TFB	TTTATAA	ND

ND- No determinado

Tabla 3 - Secuencias de factores de iniciación transcripcional incluidos en el análisis comparativo.

Organismo	Familia	Proteína	NCBI gi	Siglas
<i>Klebsiella pneumoniae</i>	$\sigma^{54}$	RpoN	133294	s54-kpn
<i>Alcaligenes eutrophus</i>	$\sigma^{54}$	RpoN	38785	s54-aeu
<i>Azotobacter vinelandii</i>	$\sigma^{54}$	RpoN	133291	s54-avi
<i>Thiobacillus ferrooxidans</i>	$\sigma^{54}$	RpoN	133300	s54-tfe
<i>R. leguminosarum</i> bv. phaseoli	$\sigma^{54}$	RpoN	1046231	s54-rph
<i>Escherichia coli</i>	$\sigma^{54}$	RpoN	414885	s54-eco
<i>Caulobacter crescentus</i>	$\sigma^{70}$	RpoD	1079588	s70-ccr
<i>Escherichia coli</i>	$\sigma^{70}$	RpoD	133309	s70-eco
<i>Lactococcus lactis</i>	$\sigma^{70}$	RpoD	216733	s70-lla
<i>Agrobacterium tumefaciens</i>	$\sigma^{70}$	SigA	464685	s70-atu
<i>Myxococcus xanthus</i>	$\sigma^{70}$	RpoD	133312	s70-mxa
<i>Escherichia coli</i>	$\sigma^{38}$	RpoS	401023	s38-eco
<i>Salmonella thyphimurium</i>	$\sigma^{38}$	RpoS	44116	s38-sth
<i>Yersinia enterocolitica</i>	$\sigma^{38}$	RpoS	687583	s38-yen
<i>Escherichia coli</i>	$\sigma^{32}$	RpoH	133287	s32-eco
<i>Pseudomonas aeruginosa</i>	$\sigma^{32}$	RpoH	493078	s32-pae
<i>Caulobacter crescentus</i>	$\sigma^{32}$	RpoH	1022898	s32-ccr
<i>Escherichia coli</i>	$\sigma^{24}$	RpoE	464681	s24-eco
<i>Salmonella thyphimurium</i>	$\sigma^{24}$	RpoE	585923	s24-sth
<i>Photobacterium</i> sp.	$\sigma^{24}$	RpoE	780317	s24-psp
<i>Pseudomonas aeruginosa</i>	$\sigma^{24}$	RpoE	550507	s24-pae
<i>Pyrococcus woesei</i>		aTBP	498649	atbp-pwo
<i>Thermococcus celer</i>		aTBP	498255	atbp-tce
<i>Halobacterium salinarum</i>		aTBP	1070345	atbp-hsa
<i>Sulfolobus shibatae</i>		aTBP	1097309	atbp-ssh
<i>Caenorhabditis elegans</i>		TBP	417896	tbp-cel
<i>Homo sapiens</i>		TBP	1070666	tbp-hum
<i>Emericella nidulans</i>		TBP	887880	tbp-eni
<i>Zea mays</i>		TBP	293906	tbp-zma
<i>Plasmodium falciparum</i>		TBP	160700	tbp-pfa

NCBI gi = número de identificación dentro de la colección de secuencias del NCBI.