

125  
2ej



**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

**FACULTAD DE CIENCIAS**

**C. I. F. N.**

**" CLASIFICACION Y POSIBLES RELACIONES  
FUNCIONALES DE UNA COLECCION DE  
REGULADORES TRANSCRIPCIONALES  
DE PROMOTORES SIGMA70 DE  
Escherichia coli "**

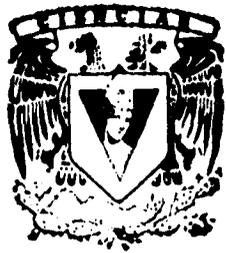
**T E S I S**

**PARA OBTENER EL TITULO DE:**

**B I O L O G O**

**P R E S E N T A**

**ERNESTO PEREZ RUEDA**



**MEXICO, D. F.**

**1995**

**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 esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

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



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

M. en C. Virginia Abrín Batule  
Jefe de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

" Clasificación y Posibles Relaciones Funcionales de una Colección  
de Reguladores Transcripcionales de Promotores Sigma70 de  
Escherichia coli ".  
realizado por

Ernesto Pérez Rueda  
con número de cuenta 8721205-3 , pasante de la carrera de Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis  
Propietario Dr. Pedro Julio Collado Vides  
Propietario Biólogo Carlos Alberto Castillo Pompeyo  
Propietario Biólogo Miguel Angel Meneses Pérez  
Suplente M. en C. Víctor Manuel Valdés López  
Suplente M. en B. José Adolfo Escalante Lozada

*En las Collado*  
*[Firma]*  
*[Firma]*  
*Carlos Alberto Castillo Pompeyo*  
*José Adolfo Escalante Lozada*

FACULTAD DE CIENCIAS

Coordinación General de Biología  
  
COORDINACIÓN GENERAL  
DE BIOLOGÍA

## *Escepticismo*

(I)- Dios de la controversia,  
concédeme el olvido  
de las mentiras en las que pude creer,

otórgame el perdón  
por las verdades que sostengo,  
a las que quizás un giro de la tierra  
vuelva falsas.

(II)- ¿Se sabrá algo para siempre?  
Nada se abraza como siempre,  
alma abrasada desde siempre.  
Si abras vacías habrás visto...

Ida Vitale

A mis padres: **Daniel y Carmen.**

Gracias:

Por sus noches de insomnio y preocupación

Por sus lágrimas que jamás he merecido.

Por toda su sabiduría y cariño que a veces no logro comprender; y

Sobre todo: Por haberme dado la vida.

A mis Hermanos y Amigos:

**Esperanza, Daniel, Ricardo,**

**Víctor:**

Gracias por todo.

Los quiero más que a mi vida.

Debo hacer un reconocimiento especial a todas aquellas personas que han hecho posible este trabajo, tanto directa como indirectamente; pero en especial debo dar las gracias:

A **Julio Collado Vides**, por su paciencia en la difícil labor de asesor y amigo a la vez.

A **David Link**, por su valiosa y desinteresada ayuda dentro del laboratorio.

A todos mis amigos de generación: **Rafael, Germán, Christian, Toriz, Fabian Romero, Gerardo y Alejandro**. Por los gratos momentos que pase con ellos y porque ni el tiempo ni la distancia los borrará de mi alma.

A **Carlos Castillo Pompeyo, José Luis Silencio, Bertha Espinoza**, por sus enseñanzas.

Al **CCH Sur** y a la **Universidad Nacional Autónoma de México**. Por dejarme y enseñarme a ser universitario.

## RESUMEN

Una colección de 44 reguladores transcripcionales de promotores sigma70 fue analizada en términos de secuencias de aminoácidos, propiedades regulatorias, sitios de pegado al DNA y posición del dominio que se pega al DNA (Amino (N-) o Carboxilo (C)-terminal). Se obtuvieron un total de 14 familias de proteínas reguladoras, a partir de la comparación de secuencias de aminoácidos, algunas de estas familias ya han sido establecidas previamente. Se propone un total de 4 nuevas familias con base en criterios de similitud de secuencias, así como de algunas propiedades asociadas para cada uno de los grupos. Las proteínas fueron reagrupadas en términos de sus actividades regulatorias, observándose familias muy estables funcionalmente hablando, tales como la familia GalR/LacI y LysR. Se proponen algunas hipótesis acerca del probable origen de la actividad regulatoria positiva y negativa de dichas proteínas. Se describen además dos posibles relaciones entre actividad regulatoria y posición del dominio de pegado al DNA dentro de la proteína, y actividad regulatoria en relación al tipo de simetría de la secuencia de DNA del operador.

## ABSTRACT

We have analyzed a collection of 44 regulatory proteins of sigma70 promoters in *E. coli* in terms of amino acid sequences, regulatory properties, DNA binding domains (Amino or carboxyl terminal) and DNA binding sites. We have obtained 14 families of regulatory proteins by amino acid sequence comparisons, of which some families already have been reported. We are proposing 4 new families. The regulatory proteins were grouped by regulatory activities. Very stable families such as GalR/LacI and LysR were observed. We propose some evolutive hypotheses about the probable origin of regulatory activity: repression and activation. We described relations between DNA binding sites *versus* regulatory activity and the position of DNA binding domain *versus* regulatory activity.

## INDICE

	PAG
I <b>INTRODUCCION</b> .....	11
I.1 Regulación de la expresión Génica.....	12
I.2 Operadores.....	18
I.3 Promotores Sigma70.....	21
I.4 Transcripción.....	23
I.5 Regulación del Inicio de la Transcripción.....	24
II <b>ANTECEDENTES</b> .....	28
III <b>OBJETIVOS</b> .....	31
IV <b>MATERIAL Y METODO</b> .....	33
V <b>RESULTADOS</b> .....	36
VI <b>DISCUSION Y CONCLUSIONES</b> .....	57
VII <b>PERSPECTIVAS</b> .....	65
VIII <b>REFERENCIAS</b> .....	66
<b>APENDICE 1</b> .....	73
<b>APENDICE 2</b> .....	83

**TABLA DE EQUIVALENCIAS DE  
AMINOACIDOS**

<b>A</b>	<b>Alanina</b>
<b>C</b>	<b>Cisteína</b>
<b>D</b>	<b>Ácido Aspártico</b>
<b>E</b>	<b>Ácido Glutámico</b>
<b>F</b>	<b>Fenilalanina</b>
<b>G</b>	<b>Glicina</b>
<b>H</b>	<b>Histidina</b>
<b>I</b>	<b>Isoleucina</b>
<b>K</b>	<b>Lisina</b>
<b>L</b>	<b>Leucina</b>
<b>M</b>	<b>Metionina</b>
<b>N</b>	<b>Asparagina</b>
<b>P</b>	<b>Prolina</b>
<b>Q</b>	<b>Glutamina</b>
<b>R</b>	<b>Arginina</b>
<b>S</b>	<b>Serina</b>
<b>T</b>	<b>Treonina</b>
<b>V</b>	<b>Valina</b>
<b>W</b>	<b>Triptófano</b>
<b>Y</b>	<b>Tirosina</b>

## I. INTRODUCCION

La necesidad de tener una visión global acerca de los fenómenos biológicos, ha originado que hoy en día se lleven a cabo grandes esfuerzos en biología molecular, principalmente en proyectos de secuenciación del genoma de *Escherichia coli*, previéndose que aproximadamente en un lapso de dos años pueda completarse dicha secuenciación. El sistema de *E.coli*, no es el único estudiado desde la perspectiva de secuenciación de genomas, ya que hay otros organismos en los que se trabaja bajo la misma pauta, tal es el caso de *Coenorhabditis elegans*, *Saccharomyces cerevisiae* y *Homo sapiens sapiens* (humano) entre otros.

Estos proyectos originan cada día una mayor cantidad de información biológica, que se acumula a la información disponible, producto de trabajos realizados por tratar de entender la fisiología de los organismos en diferentes condiciones de estudio.

Vemos que cada vez se hace más evidente la necesidad de englobar, clasificar y correlacionar los resultados que por décadas se han venido generando, sobre todo en un campo que puede proporcionar las pistas necesarias para poder conocer los principios generales de la organización molecular: La regulación de la expresión génica.

Bajo esta perspectiva hemos decidido estudiar las propiedades de una colección de proteínas reguladoras de promotores sigma<sup>70</sup> de *E. coli*, por lo cual es importante y a la vez necesario dar una introducción sobre el mecanismo de la regulación de la expresión génica y de sus componentes principales, entre los que debemos destacar las proteínas reguladoras.

## I.1 REGULACION DE LA EXPRESION GENICA

### Niveles de Organización y Regulación Génica.

Con la elucidación de las bases moleculares, las biomoléculas han incrementado su papel en estudios taxonómicos, p.e. el DNA, RNA y las proteínas proveen un amplio conjunto de datos acerca de las diferencias o semejanzas en la estructura genética de las poblaciones, estimándose muchas de las veces relaciones entre taxa. Estos estudios moleculares han generado la realización de grandes bases de datos que proveen una herramienta "comparativa", que puede proporcionarnos algunas pistas en la organización molecular.

Uno de los varios sistemas que han proporcionado gran información a nivel regulación génica es *Escherichia coli*, debido a su gran plasticidad de respuesta ante los diversos cambios ambientales por el uso "selectivo" de sus genes. *E.coli* presenta de 3000 a 4000 genes, usando aproximadamente la mitad de ellos en condiciones normales, el resto se utiliza en condiciones de stress (Neidhardt, F. *et al.* 1990). Toda esta flexibilidad en la expresión de los genes en diferentes circunstancias refleja la importancia de los mecanismos de regulación génica como uno de los temas centrales de investigación en biología molecular.

Así mismo, debemos mencionar los mecanismos de regulación integral en procariontes que son importantes para el metabolismo bacteriano, para su plasticidad de respuesta ante los estímulos ambientales y así lograr una mayor adaptación al medio. (Hoopes y MacClure, 1987):

**Regulación Global:** Incluye la regulación en respuesta a diversas fuentes nutritivas, tales como: N, C, O<sub>2</sub> o P, respuesta al choque térmico, daño y reparación al DNA. Dicha regulación se lleva a cabo por proteínas reguladoras, tales como: CRP, Fnr y LexA entre otras.

**Regulación específica:** Donde los metabolitos o catabolitos regulan la expresión de genes involucrados en su propia síntesis o degradación. Algunas proteínas involucradas en este tipo de regulación son AraC, MalT y LacI, entre otras.

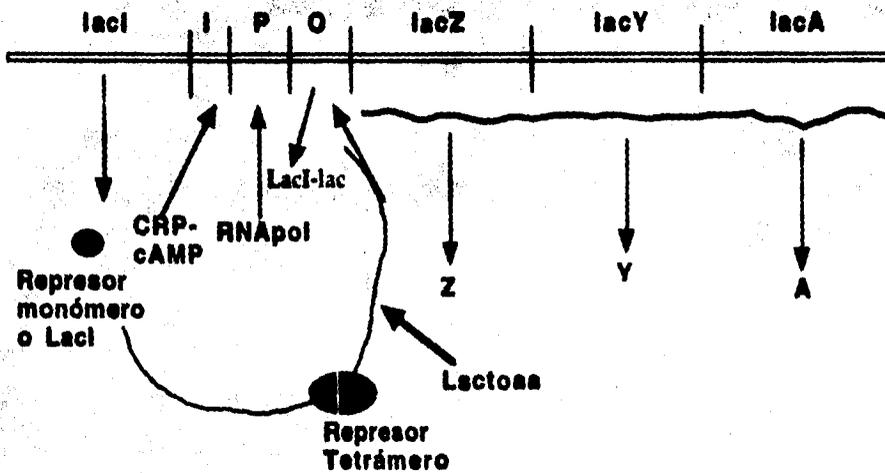
**Regulación temporal:** Regulación para replicación y/o ciclo celular .

Es necesario que los mecanismos de regulación involucren genes organizados en diferentes sistemas, de manera que puedan funcionar íntegra y/o adecuadamente ante los estímulos ambientales. De esta forma los sistemas de organización de los genes conocidos podemos clasificarlos con base en su asociación y/o respuesta, de la siguiente manera:

**Operón:** Conjunto de genes estructurales, promotor y operador que son regulados por una sola proteína reguladora, que se encuentran adyacentes unos con otros y que son transcritos en una cadena de RNAm policistrónico. Figura 1.

**Regulón:** Red de operones o genes individuales controlados por una sola proteína reguladora (que reconoce una secuencia particular común a las regiones de los operones o genes) y su ligando efector. Por ejemplo el regulón *pho* (para la toma de fósforo) (Neidhart, F. C. *et al.* 1987). Figura 2

## OPERON LACTOSA



**P: PROMOTOR**  
**O: OPERADOR**  
**I: Sitio de Pegado de la proteína activadora**  
**CRP-cAMP: PROTEINA REGULADORA**  
**RNAPol: HOLOENZIMA (RNAPol-s70)**  
**Genes del Operón: lacZ, lacY y lacA.**  
**Inductor del Operón: Lactosa**  
**LacI-lac: Complejo represor-metabolito que se separa del DNA**

Fig.1. El operón *lac* en *E. coli*, consiste de 3 genes adyacentes (*lacZYA*) y un gene regulador, *lacI*. El operón responde a la presencia de lactosa como inductor y es reprimido con el pegado de LacI en posiciones -80 y +1 aproximadamente, con respecto al inicio de la transcripción (+1). El sistema es dependiente además del sistema de cAMP-CRP, durante la represión catabólica. Modificado de Lehninger, A. L., 1975.



## Proteínas Regulatoras y Dominios de Pegado al DNA.

Las proteínas que regulan la transcripción reconocen sitios específicos dentro del DNA por medio de pequeños dominios. En algunos casos, estos dominios como por ejemplo el dominio de pegado al DNA, pueden ser intercambiados entre proteínas, mostrando que pueden comportarse como unidades de plegamiento independientes (Harrison, S. C. 1991).

Un dominio es la unidad funcional de la estructura terciaria de las proteínas, y se define como una cadena polipeptídica que puede plegarse independientemente en una estructura terciaria estable. Los dominios son unidades funcionales; así los diversos dominios de una proteína están asociados también a diferentes funciones. (Branden, C. y J. Tooze, 1991).

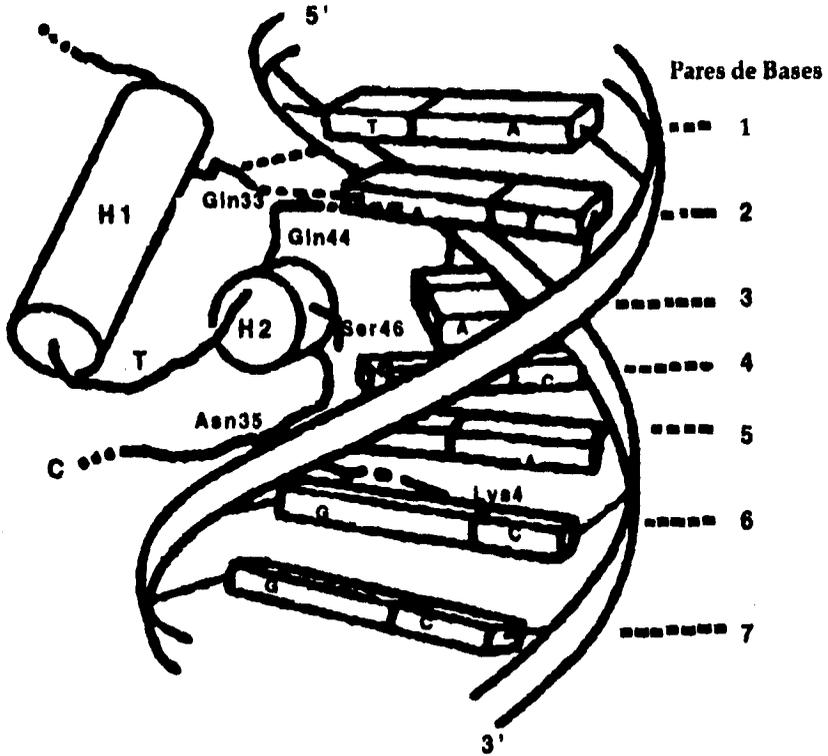
Las proteínas pueden comprender desde un sólo dominio hasta una amplia variedad de los mismos (Branden, C. y J. Tooze, 1991). Los dominios de pegado al DNA revelan una variedad de diseños: pueden plegarse y presentarse como protuberancias en la superficie molecular o bien como estructuras flexibles que se extienden a través de la proteína para contactar con las bases del DNA. Estos contactos ocurren en el surco mayor del DNA e incluyen puentes de hidrógeno, mediados directamente por moléculas de agua, así como por interacciones de Van der Waals. (Harrison, S. C., 1991)

La mayoría de las proteínas reguladoras que pertenecen a la colección analizada presentan una estructura denominada Hélice-Vuelta-Hélice o HTH, que está definida por un segmento de 20 residuos de aminoácidos, con dos  $\alpha$ -hélices que forman un ángulo de 120 grados.

La segunda hélice se denomina de reconocimiento, ya que contacta con bases muy específicas de la región del surco mayor del DNA, así mismo en experimentos mutacionales se ha observado la importancia de dicha hélice. (Harrison, S. C. y A. K. Aggarwal, 1990). Figura 3.

Se ha sugerido que las estructuras HTH de las diferentes proteínas reguladoras en las

### ESTRUCTURA HELICE-VUELTA-HELICE



**FIG. 3** Estructura de pegado al DNA, hélice-vuelta-hélice. La hélice N-terminal o hélice 1 contacta con los grupos fosfato del DNA, mientras que la hélice 2 contacta con bases específicas del surco mayor del DNA. Esquema tomado de Adhya, S. y S. Garges, 1990.

cuales está presente, han derivado de un precursor común y que la capacidad de pegarse al DNA en operadores específicos ha sido adquirida por reemplazamientos de aminoácidos dentro de la estructura del HTH. (Harrison S. C. y A. K. Aggarwal, 1990)

En otros casos se ha detectado la presencia de una estructura de hoja  $\beta$ -plegada antiparalela, que se constituye de 4  $\alpha$ -hélices, como en MetJ y ArcA o Dye.

#### Proteínas Activadoras.

Las proteínas reguladoras pueden ser activadoras del inicio de la transcripción, incrementando la afinidad de la holoenzima por el promotor a nivel de \*formación del complejo cerrado, incrementando la tasa de formación del complejo abierto del promotor o bien aumentando la velocidad con que la holoenzima abandona el promotor (antes o después de la liberación del factor sigma de la holoenzima, para la elongación de la cadena de RNAm), (Von Hippel, P. H. *et al.* 1992). En general, las proteínas activadoras responden a concentraciones de metabolitos, sustratos, cofactores y segundos mensajeros (Von Hippel, P. H. *et al.* 1992), por ejemplo CRP responde al nivel de cAMP intracelular.

#### Proteínas Represoras

Los represores están sujetos a controles similares a los de los reguladores positivos y pueden operar en varios niveles de la función del promotor con respecto a su asociación con la RNAPol; ya sea bloqueando totalmente un promotor, previniendo el pegado de la RNA polimerasa y evitando la formación del \*complejo cerrado o bien evitando la isomerización del complejo cerrado a complejo abierto (Von Hippel, P. H. *et al.* 1992).

\* Para mayores detalles de los complejos de isomerización DNA-Holoenzima ver la sección I.3, I.4 y la figura 5.

## I.2 OPERADORES

Un operador puede definirse con base en la función que desempeña, esto es, es una secuencia de DNA que es reconocida por una proteína reguladora a la cual se asocia para formar un complejo nucleoproteico y que puede interferir con el inicio de transcripción, ya sea para activar o bien para reprimir.

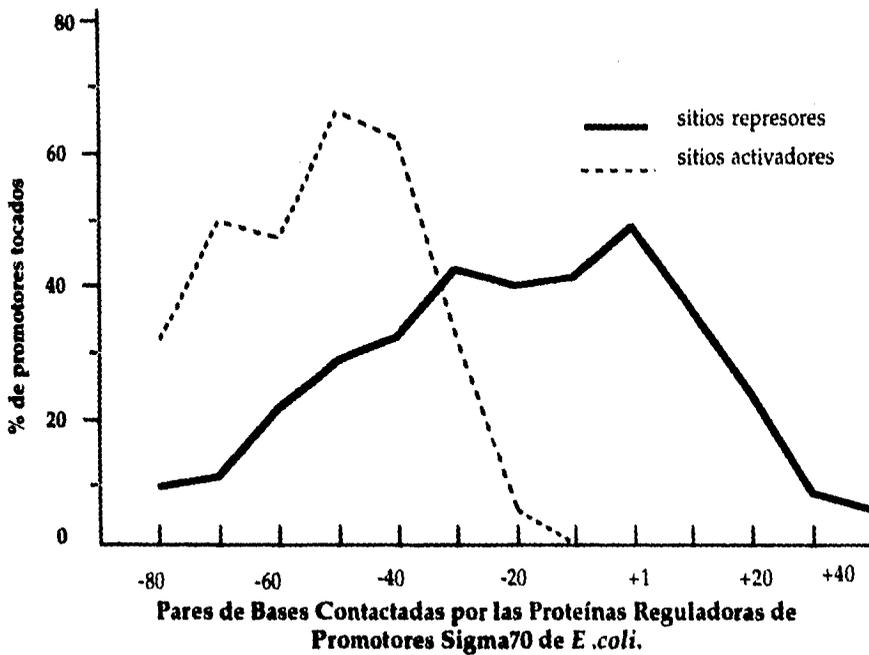
La estructura interna de la secuencia del operador es una característica muy importante, ya que puede presentarse en la forma de repetición directa, repetición invertida o con asimetría. Estas diferentes formas de secuencias de los operadores, son importantes para la actividad regulatoria asociada (como se verá en resultados y discusión), así como la posición en que se hallan ubicados dentro del DNA con relación al inicio de la transcripción.

En relación a la posición que ocupan los sitios regulatorios u operadores y que afectan el inicio de la transcripción, podemos distinguir dos tipos de localización, definidos con base al promotor *lac*. y la proteína reguladora CRP. Los sitios proximales están definidos como aquellos que están entre el sitio de pegado corriente arriba para CRP (-65) y el sitio de pegado corriente abajo de la RNAPol (+20). Esto significa que el rango va de -65 a +20, con respecto al inicio de la transcripción. Aquellos sitios que no tocan este rango de posiciones, se consideran como sitios remotos. La definición se basa en que aproximadamente dentro de este rango de posiciones CRP puede contactar la RNAPol para ayudar a la transcripción del operón *lac* (Collado-Vides, J. *et al.* 1991).

Sitios de Pegado de las proteínas reguladoras en el DNA.

En un estudio realizado por Collado-Vides, J. y J. D. Gralla (en prensa), se analizó una colección de promotores sigma70, en conjunto con proteínas reguladoras, observándose zonas de exclusión o permitidas para que se de la activación o la represión en base a los sitios de pegado de las proteínas en base al +1 o inicio de la transcripción. La figura 4 muestra que los activadores se pegan predominantemente a posiciones entre -80 y -30, aunque dentro de estas 50 pares de bases no hay una fuerte predominancia hacia posiciones determinadas.

Para el caso de las proteínas represoras, que necesitan bloquear el inicio de la transcripción, como se indicó anteriormente, no se observa una alta preferencia hacia una posición de pegado en particular (va del rango de -80 a +50). Lo interesante es que la zona de pegado de los represores coinciden con el sitio de pegado de la RNAPol. La flexibilidad en la localización de sitios de pegado para las proteínas represoras puede reflejar la capacidad de interferencia con la mayoría de los sitios de pegado para la RNAPol. Figura 4.



**Figura 4.** Distribución de los sitios de pegado al DNA de las proteínas reguladoras de promotores sigma70 de *Escherichia coli*. Gráfica tomada de Collado-Vides, J y Gralla, J., en prensa.

### I. 3. PROMOTORES Sigma70

Por definición, los promotores son secuencias de DNA que son reconocidos por la RNAPolimerasa (RNAPol)- $\sigma^{70}$ , para que se de el inicio de la transcripción con la respectiva síntesis de un RNAm policistrónico.

La secuencia de DNA de un promotor refleja la capacidad de actuar como un sitio de reconocimiento específico para una proteína y su subsecuente manipulación enzimática, resultando en la separación de la doble cadena de DNA. El reconocimiento diferencial de los promotores es determinado por el factor sigma (sigma70 en este estudio) siendo la base principal para la separación de las secuencias de los promotores en clases. Ver tabla 1. La clase más abundante de promotores son los reconocidos por la RNAPol-sigma70, existiendo además otras clases de promotores reconocidos por otros factores sigma como se muestra en la tabla1.

Sigma	Organismo	región -35	región -10
70	<i>E. coli</i>	TTGACA	TATAAT
32	<i>E. coli</i>	TCTC-CCCTTGAA	CCCCAT-TA
A	<i>B. subtilis</i>	TTGACA	TATAAT
B	<i>B. subtilis</i>	AGGTTTAA	GGGTAT
D	<i>B. subtilis</i>	CTAAA	CCGATAT
E	<i>B. subtilis</i>	ATATT	ATACA
G	<i>B. subtilis</i>	TGAATA	CATACTA
K	<i>B. subtilis</i>	AC	CATA--T
H	<i>B. subtilis</i>	CAGGA	GAATT--T
gp28	fago SPO1	AGGAGA	TTT-ITT
gp55	Fago T4	Ninguna	TATAAATA
Sigma	Organismo	región -24	región -12
54	<i>E. coli, K. pneumoniae</i> <i>Rhizobium sp., S. typhimurium</i>	CTGGA-A	TTGCA

Tabla 1. Secuencias consenso de promotores reconocidos por la RNA polimerasa conteniendo varios factores sigma (Gross, C. A., M. Lonetto y R. Losick. 1992).

Los promotores sigma70 tienen una estructura que consiste en una secuencia consenso hexanucleotídica o caja de Pribnow (TATAAT) centrada aproximadamente 10 pares de bases antes del inicio de la transcripción (caja -10). Una segunda secuencia hexanucleotídica (TTGACA), centrada aproximadamente 35 pares de bases (pb) corriente arriba (o upstream) con respecto al inicio de la transcripción. Además de 2 secuencias espaciadoras entre el -10 y el inicio de transcripción (+1) con una longitud relativamente variable y que va de 5 a 9 pb, y otra secuencia espaciadora entre el -10 y el -35 que varía un rango de 16 a 18 pb. La secuencia en la región intermedia entre cada caja puede no tener relevancia, pero la distancia puede ser muy importante para mantener los dos sitios con la separación apropiada para la disposición espacial de la RNAPol.

Se sugiere que la función de la secuencia -35 es suministrar la señal para el reconocimiento por la RNAPol, mientras que la secuencia -10 permite la conversión del complejo cerrado a complejo abierto. Figura 5.

Como ya se ha mencionado anteriormente, el inicio de la transcripción depende de la fuerza del promotor, la cual se basa en la capacidad de establecer interacciones estables con la RNAPol. Se han realizado mutaciones de promotores aumentando la homología con la secuencia consenso descrita arriba o bien reduciendo la distancia del espaciador a longitudes cercanas a 17 pb, produciendo un aumento en la transcripción; por otro lado, cuando se realizan mutaciones que disminuyen la semejanza de cualquiera de los sitios con la secuencia consenso o aumentan la distancia entre ellas a más de 17 pb, disminuye la frecuencia de transcripción de los genes adyacentes. Estos resultados indican claramente que la fuerza del promotor depende de cuán cercana se encuentra su secuencia de la secuencia consenso.

Los promotores pueden clasificarse en simples o controlados por un sólo sistema regulatorio o una sola proteína reguladora; promotores múltiples si su región regulatoria está cerca a otro promotor y ambos son controlados por lo menos por una proteína común o bien transcriben el mismo gene, por ejemplo *gal* con dos promotores sobrelapados o *malE*, *K* con sus dos promotores divergentes, y promotores complejos, que pueden ser múltiples como *gal* o bien formar parte de un sistema más complejo de regulación, tal como *lac* que es regulado por dos proteínas reguladoras, como CRP y LacI.

De un total de 107 promotores  $\sigma_{70}$  de *E. coli*, 37 % son promotores múltiples, 20 % son de regulación múltiple y 40% con regulación simple (Collado-Vides, J. *et al.* 1991).

Estos resultados sugieren que la mayoría de los promotores son regulados aisladamente; sin embargo, muchos de estos promotores simples son parte de grandes regulones. Si uno toma en cuenta la conexión vía reguladores, sólo 4 de 107 promotores parece que están regulados aisladamente de los otros promotores (Collado-Vides, J. *et al.* 1991).

En los apartados anteriores se han explicado los elementos necesarios para que se de la regulación génica, tales como promotores, operadores y proteínas reguladoras, también como sus principales papeles dentro del mecanismo de la regulación de la expresión génica; ahora bien, es necesario describir el mecanismo de la transcripción y su regulación, como un proceso global que involucra a todos los componentes antes mencionados.

#### I.4 TRANSCRIPCION

La respuesta de todos los genes necesita estar regulada, para lo cual se ha desarrollado un mecanismo a lo largo de la evolución capaz de modular la expresión de los genes ante un estímulo dado; tal como la utilización de carbono, nitrógeno o respuesta al choque térmico, entre otros. El mecanismo de control de la actividad de los genes bacterianos se ejerce principalmente en la transcripción, ya sea bloqueando el reconocimiento de la polimerasa al promotor, evitando o estimulando cualquier etapa intermedia para que se de la elongación del RNA mensajero (RNAm).

La transcripción es un evento preciso, sitio específico y con orientaciones específicas en el cromosoma. En *E.coli* la frecuencia del inicio de la transcripción varía en un rango de 10 000 veces como resultado de los mecanismos regulatorios y de la fuerza inherente de los promotores (Neidhardt, F. *et al.* 1990).

Para comprender mejor el mecanismo de regulación génica por activación y/o represión por pegado al DNA por proteínas reguladoras, es necesario dar un panorama de como se lleva a cabo el inicio de la transcripción en promotores sigma70 en procariontes:

El modelo del inicio de la transcripción sugiere los siguientes pasos: Figura 5.

A. Formación del complejo RNAPol-factor sigma70, generando la holoenzima. Figura 5A. El factor  $\sigma$ 70 confiere especificidad al complejo para el reconocimiento del promotor.

B. Reconocimiento del promotor por la holoenzima, dado por el factor sigma70, formando el complejo cerrado. Figura 5B.

C. Desnaturalización o apertura del DNA en la zona de contacto entre DNA-RNA pol, aproximadamente entre la posición -10 y +1, exponiendo el sitio de inicio de la transcripción. Este es el denominado complejo abierto. Figura 5C.

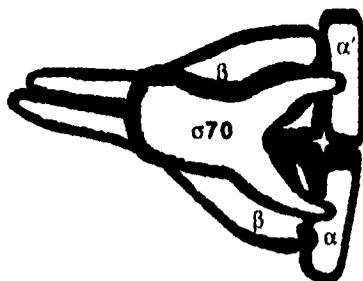
D. Al iniciar la transcripción el factor sigma se libera, volviendo a reciclarse. Se sintetizan cadenas cortas de RNAm "abortivas", al sintetizarse una cadena lo suficientemente estable de 8 a 9 nucleótidos, se produce la elongación de la cadena del RNAm. La transcripción se realiza corriente abajo del promotor. Figura 5D.

Como dato adicional, se han observado pausas en la elongación de la cadena, sin darse todavía una explicación clara de este proceso.

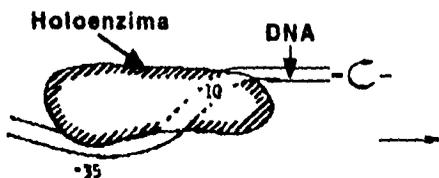
## **I. 5 REGULACION DEL INICIO DE LA TRANSCRIPCION**

En *E.coli*, la tasa del inicio de la transcripción en los promotores sigma70 es regulada de acuerdo a los requerimientos celulares. Esta regulación opera a dos niveles: primero, una directa modulación de la especificidad y afinidad de la RNAPol-sigma70 para el reconocimiento de su promotor, y segundo, el control por reguladores que interactúan con sitios de pegado adicionales al promotor cercanos generalmente al sitio de pegado de

## TRANSCRIPCIÓN EN PROMOTORES SIGMA70



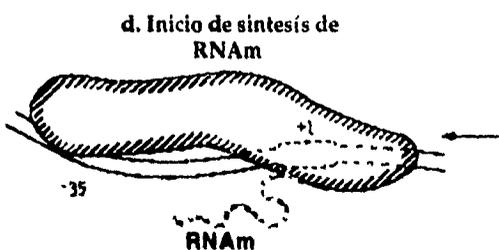
a. Complejo RNApol-sigma70= Holoenzima



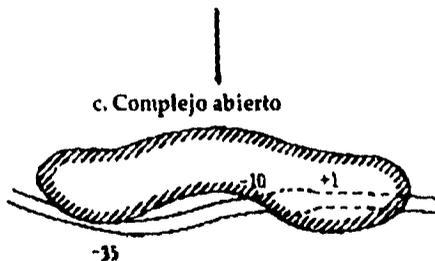
b. Reconocimiento del promotor por la holoenzima



c. Complejo cerrado



d. Inicio de síntesis de RNAm



c. Complejo abierto

FIG. 5. El proceso de la transcripción se da por un mecanismo de reconocimiento de un promotor específico por el complejo RNApol-sigma70. La transcripción se da por una serie de pasos que consisten en: Formación del complejo RNApol-sigma70, reconocimiento del promotor (con cajas -35 y -10), formación del complejo cerrado DNA-RNApol-sigma70, apertura del DNA en la región del -10 e inicio de polimerización del RNAm. (Para detalles ver el texto).

la RNAPol-sigma70.

Los sistemas de regulación que interactúan con sitios de pegado al DNA (operadores) adicionales al promotor pueden actuar positiva o negativamente, definida esta propiedad en base a la respuesta del operón cuando no se encuentra una proteína reguladora.

Los genes bajo control negativo se expresan, a menos que una proteína represora los inactive. Una proteína represora se pega al DNA para impedir que la RNAPol inicie la transcripción. Las proteínas reguladoras represoras se pegan al DNA en un sitio localizado corriente abajo del promotor o bien sobrelapado con el promotor, de manera que la RNAPol no puede reconocer el -10 o el -35 del promotor y por lo tanto no puede iniciar la transcripción.

El operón *lac* ha sido el modelo de estudio de la regulación negativa. Los tres genes *lac* (*lacZYA*) son expresados en altos niveles cuando se encuentra en el medio el sustrato adecuado (lactosa). La regulación ocurre como sigue: el gene *lacI* codifica una proteína represora LacI que se pega a un sitio corriente arriba del promotor de *lac*, inhibiendo el inicio de la transcripción, en ausencia de lactosa en el medio. En presencia de lactosa en el medio, el complejo represor-operador es desestabilizado por el pegado del inductor a LacI, formándose el complejo represor-inductor, produciendo la separación de LacI del DNA, conduciendo así, al inicio de la transcripción por el pegado de la RNAPol al promotor. Figura 1.

Para genes bajo control positivo, la expresión únicamente es posible cuando una proteína reguladora activadora está presente. El pegado de la proteína reguladora cercana al sitio de inicio de la transcripción incrementa la tasa de transcripción, incrementando la formación del complejo abierto.

Los promotores son, en este caso, débiles para la interacción con la holoenzima, puesto que sus secuencias normalmente difieren de la secuencia consenso. La proteína activadora interactúa con el DNA y con la RNAPol para ayudar al inicio de la transcripción. Para el caso que se ha propuesto en el operón *lac*, la interacción de CRP

con la RNAPol se da cuando la proteína represora del operón (LacI) se despega del operador; la subunidad alfa de la RNAPol hace contacto con un segmento de la proteína activadora CRP, que va del aminoácido 158 al 164, ayudando al inicio de la transcripción.

El papel de las proteínas reguladoras en el inicio de la transcripción es fundamental, puesto que contribuyen a la expresión de los genes bajo diversas condiciones ambientales, desde la presencia de un metabolito (como el caso de lactosa), hasta la ausencia total del mismo.

Es necesario pues, que el estudio de los componentes del mecanismo de la regulación génica, y en particular de las proteínas reguladoras de los promotores sigma70, se realice desde una perspectiva más global, que pueda ayudarnos a entender el comportamiento de la colección de proteínas reguladoras desde el ángulo de la clasificación de las mismas y de una relación con sus actividades regulatorias.

## II. ANTECEDENTES

### Familias de Proteínas Reguladoras.

Las clasificaciones en biología son muy importantes, ya que son las herramientas que nos ayudan a resumir y describir la diversidad orgánica y por las cuales tratamos de encontrar explicaciones a los fenómenos biológicos. Las clasificaciones son el almacén de atributos o características no sólo de organismos, sino también de moléculas.

Las proteínas son agrupadas en familias cuyos miembros han divergido de un ancestro común, tienen plegamientos similares y presentan por lo general secuencias y funciones similares (Chothia, C. 1992).

Uno de los temas de estudio más interesantes lo constituyen las proteínas involucradas en la regulación de la expresión génica, tales como las que regulan la expresión de los genes u operones asociados al factor sigma70. Su clasificación se basa generalmente en la comparación de secuencias de aminoácidos o estructuras primarias, por lo cual hemos conjuntado una colección de 44 proteínas reguladoras de promotores sigma70 de *Escherichia coli*, de las cuales 26 proteínas ya han sido identificadas dentro de alguna familia proteica.

Dentro de los estudios realizados para clasificar las proteínas reguladoras, encontramos los trabajos de Haydon D. J. y J. R. Guest, 1991, que describen la familia GntR en base a similaridad de secuencias con respecto al represor GntR del operón gluconato en *Bacillus subtilis*, además se incluyen los reguladores para la utilización de histidina en *Pseudomonas putida* (HutC<sub>pp</sub>) y HutC<sub>ka</sub> de *Klebsiella aerogenes*, el represor FadR para la utilización de ácidos grasos, entre otros.

Henikoff, S. *et al.* 1988, realiza el trabajo de clasificación de los miembros de la familia LysR con base en alineamientos de secuencias y además predice una estructura a manera de hélice-vuelta-hélice como dominio de reconocimiento del DNA. Con base en este trabajo se deduce un posible ancestro común para todas las proteínas de esta familia, que pueden presentar plegamientos similares.

Por lo que respecta a la familia AraC/XylS, la descripción de sus miembros se basa en la comparación de sus secuencias con respecto a AraC y XylS, dos proteínas reguladoras muy estudiadas y que exhiben gran similitud entre sus secuencias. (Ramos, J. L. *et al.* 1990). Ver tabla 2, pág. 39, para más detalles.

En el caso de la familia GalR/LacI, las proteínas se han agrupado comparando su secuencia de aminoácidos, observándose la presencia de proteínas que no se pegan al DNA y que por lo tanto no son reguladoras, sino transportadoras de azúcares asociadas a membranas celulares.

Sin embargo, no todas las proteínas reguladoras se han clasificado comparando las secuencias de aminoácidos, sino que algunas clasificaciones responden a las propiedades funcionales de las proteínas *per se*, caso concreto de esta aseveración lo constituye la familia de los componentes en la cual se han agrupado proteínas que responden a estímulos provenientes del medio y que comparten un dominio de pegado al DNA, además de un dominio que es fosforilado por una proteína cinasa. (Stock, J. *et al.* 1989).

En la base de datos "SwissProt" podemos encontrar todas las familias de proteínas reportadas, tales como la familia AraC/XylS, GalR/LacI, LysR, Crp, ArsR, entre otras; con sus respectivos miembros. Como se describe en dicha base de datos, las proteínas reguladoras se han agrupado comparando el dominio que se pega al DNA, como la estructura denominada hélice-vuelta-hélice.

Las clasificaciones en proteínas reguladoras de la expresión génica, se basan en la comparación de secuencias, tratando de encontrar posibles relaciones entre cada una de las proteínas, como ya se ha mencionado anteriormente, a excepción de la familia de los dos componentes en que el agrupamiento se basa en la funcionalidad de sus integrantes.

### III. OBJETIVOS

El presente trabajo es un primer intento en tratar de relacionar la información que a cada día surge de cientos de laboratorios en el mundo en el área de regulación de la expresión génica y con el cual pretendemos conocer desde una perspectiva global, la dinámica de la regulación génica, en promotores sigma70 de *Escherichia coli*.

Uno podría preguntarse por qué *E. coli* y por qué restringirse a los promotores sigma70, considerando que la información que se genera en regulación no sólo se restringe a estos dos campos de estudio. Pues bien, consideramos estos aspectos, debido a que tenemos por un lado, que la colección de promotores sigma70 representan una de las colecciones más conocidas y completas, además que representan parte del "housekeeping" bacteriano, de allí su importancia no sólo para la bacteria sino también en el conocimiento en biología; y por otro lado, la importancia de *E. coli* se ve reflejada, en que es uno de los principales modelos de estudio para poder entender sistemas biológicos mayores, como por ejemplo el hombre. Por lo cual, el genoma de *E. coli* está siendo secuenciado (y podría ser el primer genoma celular completamente secuenciado) como un proceso de conocimiento previo para la secuenciación de grandes genomas eucariontes.

Dentro de este proceso de conocimiento de los grandes genomas, debemos además estudiar o conocer los componentes accesorios que están involucrados en la regulación de la expresión génica; para lo cual disponemos de una colección lo suficientemente grande de 44 proteínas reguladoras de promotores sigma70, con sus respectivas propiedades regulatorias.

Las propiedades que consideramos como necesarias de conocer son: actividad regulatoria (proteínas represoras, activadoras o con ambas actividades), número y papel de los dominios que conforman cada una de las proteínas, enfatizando aquéllos dominios necesarios para la regulación (dónde se localiza el HTH u otra estructura de pegado al DNA), tipo de simetría de la secuencia de DNA del operador; como están

organizados los genes, ya sea en regulones u operones; entre otras.

Finalmente, con este trabajo pretendemos realizar una clasificación de la colección de las 44 proteínas reguladoras desde una perspectiva más dinámica, englobando todas sus propiedades disponibles en relación a la regulación génica y adicionando nuevas propiedades a cada familia proteica para reforzar la noción de familia. Además de realizar posibles relaciones funcionales entre dichas propiedades, para tratar de entender un poco más la regulación de la expresión génica en un sistema tal como *E. coli*, a la vez de estar tratando de realizar posibles hipótesis evolutivas en relación a la actividad regulatoria asociada a cada proteína.

## **IV. MATERIAL Y METODO**

Para la realización del presente trabajo se empleó una colección de 44 proteínas reguladoras de promotores sigma70 de *Escherichia coli*, de las cuales 33 se mencionan en la revisión de Collado-Vides, J. *et al.* 1991 y las restantes 11 proteínas se obtuvieron por revisión bibliográfica.

### **IV.1 CARACTERISTICAS FUNCIONALES**

En la revisión mencionada se describen algunas características importantes de las proteínas, tales como promotores asociados a cada proteína; la posición del operador con respecto al inicio de la transcripción (+1); la asociación de promotores, esto es, si son simples o múltiples; la actividad regulatoria, si son represoras, activadoras o duales y la presencia de autoregulación. Para el caso de las proteínas restantes, se hicieron revisiones bibliográficas para cada proteína, aproximadamente de 1985 a la fecha, obteniéndose de manera general las siguientes propiedades para cada proteína reguladora:

- Nombre y secuencia de aminoácidos de la proteína reguladora;
- Familia a la cual pertenece: en el caso de haber sido reportada previamente como perteneciente a alguna familia;
- Actividad regulatoria: si es represora, activadora o con actividad dual;
- Posición dentro de la proteína del dominio de pegado al DNA (N- o C- terminal), cuando este determinado;
- Tipo de simetría de la secuencia de DNA del operador (Asimétrica, de repetición invertida o repetición directa),
- Multimerización o número de subunidades que se asocian en solución,
- Número de sitios de pegado al DNA por promotor,
- Número de nucleótidos protegidos por la proteína en el DNA ("footprinting"),
- Papel de algunos aminoácidos dentro de cada proteína, entre otras propiedades.

## **IV. 2 BASE DE DATOS**

Con las características mencionadas anteriormente para cada proteína, obtenidas por revisión bibliográfica, se realizó una base de datos en el Software ClarisWorks 2.1 v3, 1994 para la posterior evaluación de las funciones encontradas de forma conjunta, en una Macintosh Classic II.

## **IV. 3 SECUENCIAS**

La secuencia de aminoácidos de cada una de las proteínas fue obtenida a partir del GenBank , en el software GeneWorks v.8 1994, en una Macintosh Quadra 805.

## **IV. 4 GCG-PILEUP: ALINEAMIENTO MULTIPLE.**

Para el análisis de las secuencias, se trabajó en el paquete Genetics Computer Group (GCG)-Pileup, versión 7.2 (Universidad de Wisconsin, Madison), que crea un alineamiento múltiple de secuencias, a partir de un grupo de secuencias relacionadas. También realiza un árbol o dendograma, mostrando las relaciones en grupos, basado en el alineamiento, en el cual la distancia del eje vertical es proporcional a la diferencia entre secuencias; la distancia del eje horizontal no tiene significado alguno.

El procedimiento de alineamiento múltiple inicia con la comparación de las dos secuencias más similares generando un grupo o "cluster". Este grupo puede entonces compararse con la siguiente secuencia más similar o bien a otro grupo de secuencias alineadas. El alineamiento final es realizado por una serie de comparaciones progresivas entre pares, que incluye un decremento en similaridad entre las secuencias y clusters, hasta que todas las secuencias han sido incluidas en el alineamiento final. Figura 8 y apéndice 1.

Antes del alineamiento, las secuencias son primero agrupadas por similaridad para producir un dendograma o representación gráfica del grado de similaridad entre las distintas secuencias alineadas. Figuras 6 y 9 .

Se realizaron dos alineamientos múltiples con GCG-Pileup, uno utilizando las secuencias totales de las proteínas y el otro con el uso únicamente de las secuencias de los dominios de pegado al DNA, ya fuera hélice-vuelta-hélice o una hoja  $\beta$ -plegada.

#### **IV. 5 GCG-MOTIFS (MOTIVOS)**

Este software busca motivos dentro de las secuencias de proteínas en base a la comparación con los patrones definidos en el "PROSITE Dictionary of Protein Sites and Patterns". Despliega además una serie de reportes bibliográficos para lectura adicional. Con este software se detectaron las familias en que se agrupan 26 proteínas reguladoras, del total de 44.

#### **IV. 6 GCG-BESTFIT**

Bestfit realiza el alineamiento más óptimo del mejor segmento de similitud entre dos secuencias. Los alineamientos óptimos son realizados por inserción de "gaps" para maximizar el número de pares. Con este software se localizaron dominios de pegado al DNA dentro de las proteínas que no han sido reportados previamente.

#### **IV. 7 ProDom (Protein Domains)**

Para enriquecer la descripción de las proteínas dentro de una familia, decidimos identificar la organización en dominios de las 44 proteínas reguladoras, auxiliándonos en la base de datos construída por D. Kahn: ProDom o Protein Domains. Esta base de datos es similar al programa de BLOCKS en NCBI.

Uno de los pasos iniciales en la construcción de estos dominios es un alineamiento múltiple sin gaps (del total de secuencias conocidas de proteínas no fragmentadas y que están reportadas dentro de la base de datos de Swiss-Prot 21.0); subsecuentemente, las secuencias de las proteínas son cortadas en diferentes dominios. Se generan así consensos para cada familia o dominio, que son utilizados para la búsqueda rápida de homologías con otras proteínas. La consulta de Prodom fue realizada vía correo electrónico en la dirección "prodom@toulouse.inra.fr".

## V. RESULTADOS

En resumen, los resultados considerados como parte central de este trabajo son los siguientes:

1. Clasificación de las 44 proteínas reguladoras, con base en la comparación de las secuencias totales de aminoácidos y con la realización de un dendrograma que muestra las relaciones de similitud. Obtención de 14 familias, de las cuales 10 son familias previamente establecidas y 4 son deducidas de este estudio. De las 10 familias ya establecidas a 4 se les agregan nuevos miembros que forman parte de la colección de proteínas reguladoras de nuestro estudio. Ver figura 6 y apéndice 1.

1.A Enriquecimiento de los criterios o propiedades para cada familia o grupo obtenido. Los criterios van desde porcentaje identidad entre los miembros de cada grupo hasta propiedades relacionadas con la regulación génica. Tabla 2.

1.B Detección de los posibles dominios en que se compone cada proteína por medio de la base de datos "ProDom" y organizados en familias que se han obtenido en este trabajo. Figura 7.

1.C Comparación de los dominios de pegado al DNA para las proteínas en que se ha detectado la estructura de HTH u hoja  $\beta$ -antiparalela. Tabla 3 y figuras 8 y 9.

2. Reagrupación de las familias reguladoras en respecto a su función reguladora (activador, represor o dual). Tabla 4.

3. Relación entre la actividad reguladora de las proteínas reguladoras y el tipo de simetría de la secuencia de DNA del operador. Tabla 5.

4. Relación entre la actividad y la posición del dominio de pegado al DNA dentro de la proteína (N- o C-terminal). Tabla 6.

A continuación, se describen de manera más extensa los resultados obtenidos, así mismo, en la discusión se realiza una posible interpretación de estos resultados.

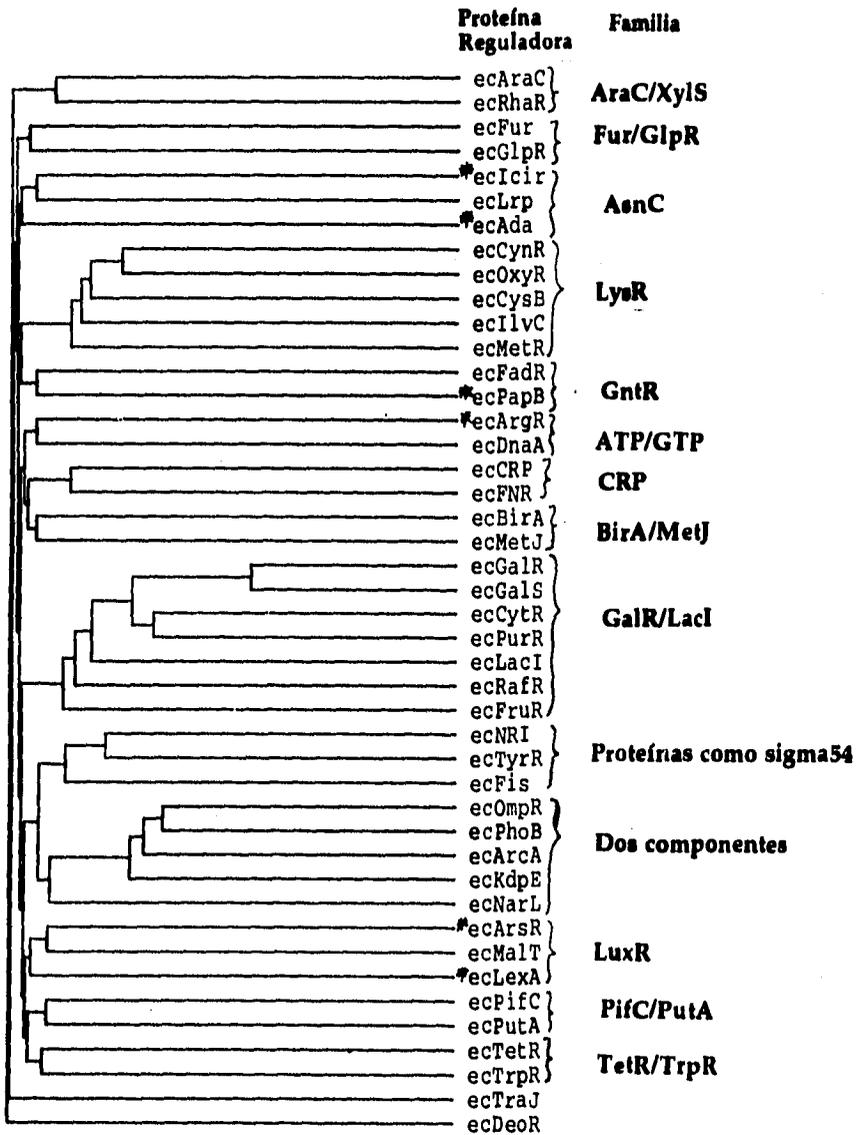
A partir de una colección de 44 proteínas reguladoras de promotores sigma<sup>70</sup>, obtenidas por medio de una revisión bibliográfica, así como de sus secuencias de aminoácidos obtenidas a partir del GenBank y alineadas (Apéndice 1) utilizando el programa GCG-Pileup se obtuvieron los agrupamientos de las proteínas reguladoras.

Los grupos proteicos obtenidos por medio de la comparación de sus secuencias de aminoácidos, se pueden ver en el dendograma de la figura 6, en el que se representan gráficamente las relaciones de similitud entre las estructuras primarias de cada proteína. Estos grupos nos muestran la estabilidad de familias previamente establecidas, tales como AraC/XylS, GalR/LacI, LysR y la familia de los dos componentes, y que además nos sirvieron como grupos control para la descripción de nuevas familias como: Fur/GlpR, PifC/PutA, TrpR/TetR y BirA/MetJ, tabla 2.

Del total de las 44 proteínas reguladoras de la colección se ha determinado que 26 proteínas se reportan como integrantes de familias previamente establecidas, mientras que 18 proteínas se ubican como integrantes de nuevas familias o bien se adicionan a familias ya reportadas.

Las 26 proteínas que ya se han reportado como integrantes de familias ya establecidas, se agrupan en 6 diferentes familias; mientras que de las 18 proteínas restantes, 7 proteínas se adicionan como miembros de familias ya existentes y 11 proteínas forman 4 nuevas familias propuestas en este trabajo, a excepción de 2 proteínas (DeoR y TraJ) que pueden considerarse como casos separados. Tabla 2.

En resumen, en la tabla 2, se muestran en negrillas las nuevas familias que se proponen como parte de este trabajo, así como los nuevos integrantes de familias previamente reportadas, y algunas de las propiedades importantes asociadas con cada familia proteica; también se muestran las familias que no sufren modificación alguna, por medio de nuestra clasificación.



**Fig 6.** Las proteínas fueron agrupadas con base en sus secuencias totales, por medio del programa Pileup-GCG. Los nombres de las familias son mencionados.

\* Nuevos miembros en familias ya establecidas,

\*\*Nuevas familias obtenidas en este trabajo.

**Tabla 2. Familias de Proteínas Reguladoras de Promotores sigma70 en Procariontes: *Escherichia coli* y *Salmonella typhimurium*.**

Se muestran las familias y sus miembros que forman parte de la colección de 44 proteínas reguladoras. Se indica además el número total de integrantes por familias y propiedades establecidas por trabajos previos. En negrillas mostramos las nuevas familias e integrantes, así como las propiedades asociadas a cada familia en algunas familias establecidas.

Las propiedades nuevas que se adicionan como parte de este trabajo se indican también en negrillas.

Nota: Se agrega en esta tabla a *Salmonella typhimurium*, debido a que muchas proteínas de la colección también se han secuenciado en esta bacteria, además que *S. typhimurium* es considerada como un organismo muy cercano a *E. coli*, evolutivamente hablando.

Familias de Proteínas Reguladoras de Promotores sigma70 en Procariontes:  
*Escherichia coli* y *Salmonella typhimurium*.

FAMILIA	MIEMBROS REPORTADOS	PROPIEDADES
AraC/XylS	AraC, RhaP y 13 proteínas adicionales. Total: 15 proteínas	Todas las proteínas presentan entre 77 y 85% de similaridad, en segmentos ali- neados; además de una fuerte homología en la región C-terminal; necesitan efec- tores para regular. La mayoría están involucrados en el metabolismo de azúcares. (Ramos, 1990).  <b>AraC y RhaP: regulan genes organizados en operones; son dímeros en solución y ambas reconocen sitios dentro del DNA con simetría directa. Los aminoácidos conservados son: A3, D4, Q54, Q56, G70, E107, L161, S208, I238, G246, F247, D249, Y252 respecto a AraC.</b>
GalR/LacI	GalR, GalS, LacI, CytR, PurR, FruR, RafR y 7 proteínas adicionales. Total: 14 proteínas.	Todas las proteínas tienen 60% o más de similaridad. La secuencia de DNA de reconocimiento está altamente conservada para todas las proteínas. Todas ne- cesitan azúcares como efectores o inductores; regulan además genes para el metabolismo de azúcares, excepto CytR y PurR que están involucrados en el metabolismo de nucleótidos. Presentan además una Arg197 para LacI conserva- da en el grupo. (Weickert y Adhya, 1992). La mayoría de las proteínas son represores. (Swiss-prot). El HTH se ubica en el N-term de las proteínas. <b>El sitio de pegado al DNA presenta simetría invertida; los amino- ácidos conservados en el grupo con respecto a LacI son: S16, T19, S21, P49, P144, R197.</b>
LysR	IlvY, CysB, MetR, CysR OxyR y 23 proteínas adicionales. Total: 28 proteínas.	La mayoría de las proteínas de este grupo regulan operones biosintéticos trans- critos por promotores divergentes. No hay un consenso para sus sitios de pega- do al DNA, aunque se presenta una estructura HTH (Henikoff, 1988). Las proteínas activadoras son autoreguladas negativamente. El sitio de activación transcrip- cional del gene(s) estructural requiere de una molécula inductora y de la proteína reguladora. Presentan un tamaño similar, de 300 aminoácidos en promedio Una fuerte homología en secuencia de aa se presenta en la región N-terminal de los activadores. (Francoise, A.1993).  <b>El aminoácido conservado con respecto a CysB es: Q36; se detecta 47% de similaridad.</b>

FAMILIA	MIEMBROS REPORTADOS	PROPIEDADES
CRP	CRP, FNR y 8 proteínas adicionales Total: 10 proteínas.	El HTH se localiza hacia el C-term. Con secuencias muy similares (Swiss-Prot). <b>Ambas proteínas son reguladores duales, dímeros en solución y se pegan a sitios en el DNA con simetría invertida. Los aminoácidos conservados con respecto a Fnr, para el grupo son: K50, T57, G74, G96, M137, I190, G218, I221. Además se observa un 45% de similitud en secuencias; regulan genes organizados en modulones.</b>
ASNC	Lrp y 2 proteínas adicionales. Total: 3 proteínas. *Ada e IciR.	Presentan 25% de identidad. el HTH se desplaza hacia el N-terminal (Willins, 1991). La mayoría son proteína activadoras (Swiss-Prot). <b>excepto IciR que es represor. Los aminoácidos conservados para el grupo con respecto a Lrp son: L12; además que se observa un 45% de similitud.</b>
GntR	FadR y 6 proteínas adicionales. Total: 7 proteínas. *PapB *FadA y FadB de E.coli	Presentan un posible HTH conservado hacia el N term; similares en peso molecular; además de ser proteínas con una amplia distribución funcional. (Haydon y Guest, 1991). <b>Son reguladores duales, con 46% de similitud, los aminoácidos conservados en el grupo con respecto a PapB son: P25, G26, T97</b>
LuxR	MaiT y 12 proteínas adicionales. Total: 14 proteínas. *LexA y AraR	El tamaño promedio de estas proteínas va de 190 a 230 aminoácidos. excepto MaiT que presenta 900 aminoácidos de longitud. El HTH se encuentra en la región C-term. (Swiss-Prot). <b>Se observa un 48% de similitud de secuencias.</b>
2 componentes	OmpR, PhoB, KdpE, ArcA, NarL 18 proteínas adicionales. Total: 23 proteínas.	Presentan repeticiones cortas y directas dentro de sus secuencias; con proteínas involucradas en repuestas adaptativas, controladas por proteínas transductoras de señales. La fosforilación de las proteínas reguladoras es una característica fundamental en el mecanismo de la transducción de señales en la región N-terminal. Se observa un dominio hacia el N-terminal conservado de aproximadamente 100 aminoácidos (Stock, 1989). <b>Los aminoácidos conservados dentro del grupo, con respecto a OmpR son: D12, R17 y K115; además se presenta un 56.4% de similitud en secuencias.</b>

41

**TABLA 2**  
**continuación...**

FAMILIA	MIEMBROS REPORTADOS	PROPIEDADES
Sigma 54	TyrR, NR(I) 12 proteínas adicionales. Total: 14 proteínas. *Fis	Proteínas con dominio conservado de 230 aminoácidos en la región central para interacción con el factor sigma54 y otro dominio menos conservado de 45 aa en el C-term. Todas las proteínas son reguladoras duales(Stock,1989). Se pegan a sitios en el DNA con simetría invertida; con 50.4% de similitud y sus aminoácidos conservados con respecto a Fis, son: Q4, L63.
ATP/GTP	DnaA. *ArgR *DnaA de Pseudomonas sp *DnaA de Micrococcus sp *DnaA de Protens sp *DnaA de Bacillus sp *DnaA de Streptomyces sp	Proteínas que pegan ATP o GTP; con una región conservada rica en Glicina (Swiss-Prot) El NTE se localiza hacia el C-term; los aminoácidos conservados del grupo con respecto a ArgR son: K12. Se detecta un 47% de similitud. Regulan genes organizados en regulones.
#Fur/GlpR	Fur y GlpR	Proteínas represoras, que se pegan a sitios dentro del DNA con simetría invertida; los aminoácidos conservados con respecto a GlpR son: L6,D75,D140; se presenta además un 48.1% de similitud.
#PifC/PutA	PifC y PutA *PfcI de E.coli.	PifC y PutA son proteínas represoras, que se pegan a sitios dentro del DNA con simetría invertida. Los aminoácidos conservados, con respecto a PutA son: R17; se detecta 47.3% similitud. Regulan genes organizados en operones.
#TrpR/ TetR	TrpR y TetR *Ter1, Ter2, Ter3 y Ter5 de E.coli	TrpR y TetR son proteínas represoras, dímeros en solución, sus sitios de pegado al DNA presentan simetría invertida. Los aminoácidos conservados para el grupo son(con respecto a TrpR): S8, Q17, T53, R56, P93; además de detectarse 51% de similitud.
#BirA/ MetJ	BirA y MetJ MetJ de Salmonella sp	BirA y MetJ son proteína represoras y dímeros en solución; los aa conservados con relación a MetJ son: P11, G16. Se detecta además un 47.5% de similitud en secuencia.

42

**TABLA 2**  
continuación...

Adicionalmente, para cada nueva familia propuesta, se presentan por lo menos dos proteínas reguladoras de promotores sigma70 que son parte de esta colección, además de otras proteínas candidatas que aunque no forman parte de la colección de 44 proteínas, se adicionan por el criterio de porcentaje de similaridad con respecto a los integrantes de cada nuevo grupo, así como por compartir un dominio común con los miembros de la familia; finalmente, se determina que por cada familia hay por lo menos de 4 a 6 miembros proteicos. Figura 7.

Además del criterio de similaridad entre secuencias de proteínas, se incluyen propiedades funcionales o estructurales para la mejor caracterización de cada grupo en particular, como por ejemplo:

a) Se identificaron de 2 a 3 aminoácidos conservados (a manera de una firma) para cada familia, para el caso de la familia de los dos componentes, dos de estos tres aminoácidos se reportan conservados dentro del gran número de proteínas miembros de la familia (Stock, J. B. *et al.* 1989);

b) Se detectó la presencia de dominios comunes para cada familia;

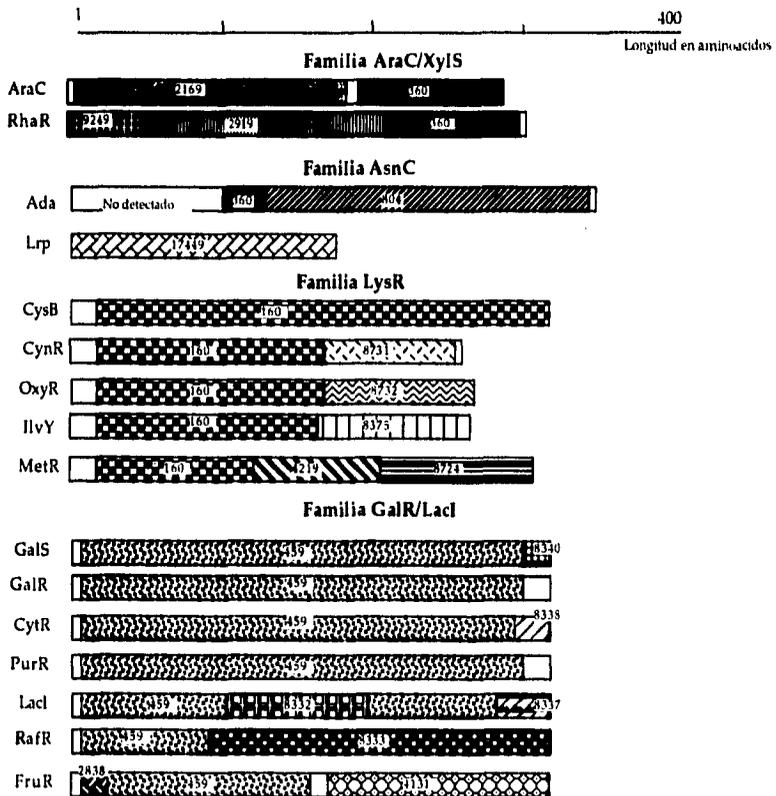
c) Presencia de un HTH en una región determinada dentro de cada miembro de la familia (en posición N- o C-terminal o bien presencia de una estructura diferente al HTH, que también se pega al DNA como la hoja  $\beta$ -plegada de MetJ y Dye); y

c) Papel que cumplen dentro de un proceso metabólico; entre otras propiedades, como puede observarse en la tabla 2.

Como ya se ha mencionado, una de las características o propiedades adjuntas a cada familia proteica lo constituyó la organización de los dominios de cada una de las 44 proteínas reguladoras y que fue considerada para reforzar los grupos obtenidos por alineamiento y relaciones de secuencias por medio de Pileup-GCG.

La utilización de una herramienta capaz de detectar o predecir posibles dominios o módulos dentro de las proteínas es importante desde la perspectiva de predicción, para lo cual la base de datos "Prodom" o "Protein domains" basada en "SwissProt" o base de datos de proteínas fue empleada para detectar los posibles dominios o módu-

## DOMINIOS PROTEICOS POR FAMILIAS



**FIGURA 7.** Las proteínas reguladoras de promotores sigma70 de *E.coli* fueron agrupadas en base a sus respectivas familias. Se detectaron los posibles dominios para cada proteína por Prodom, a excepción de tres proteínas no detectadas en esta base de datos y que son: PutA, IciR y TraJ. Los números indican asignación de cada familia dada por Prodom y de acuerdo a la nomenclatura dada por Sonnhammer y Kahn, 1994. Los espacios en blanco representan segmentos proteicos no detectados por Prodom.

DOMINIOS PROTEICOS POR FAMILIAS

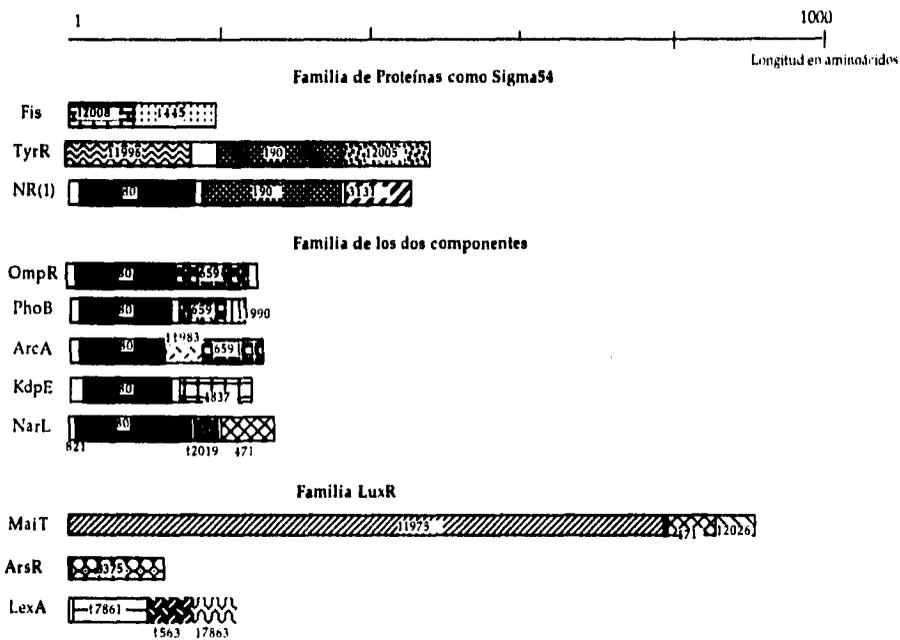


FIGURA 7 continuación...

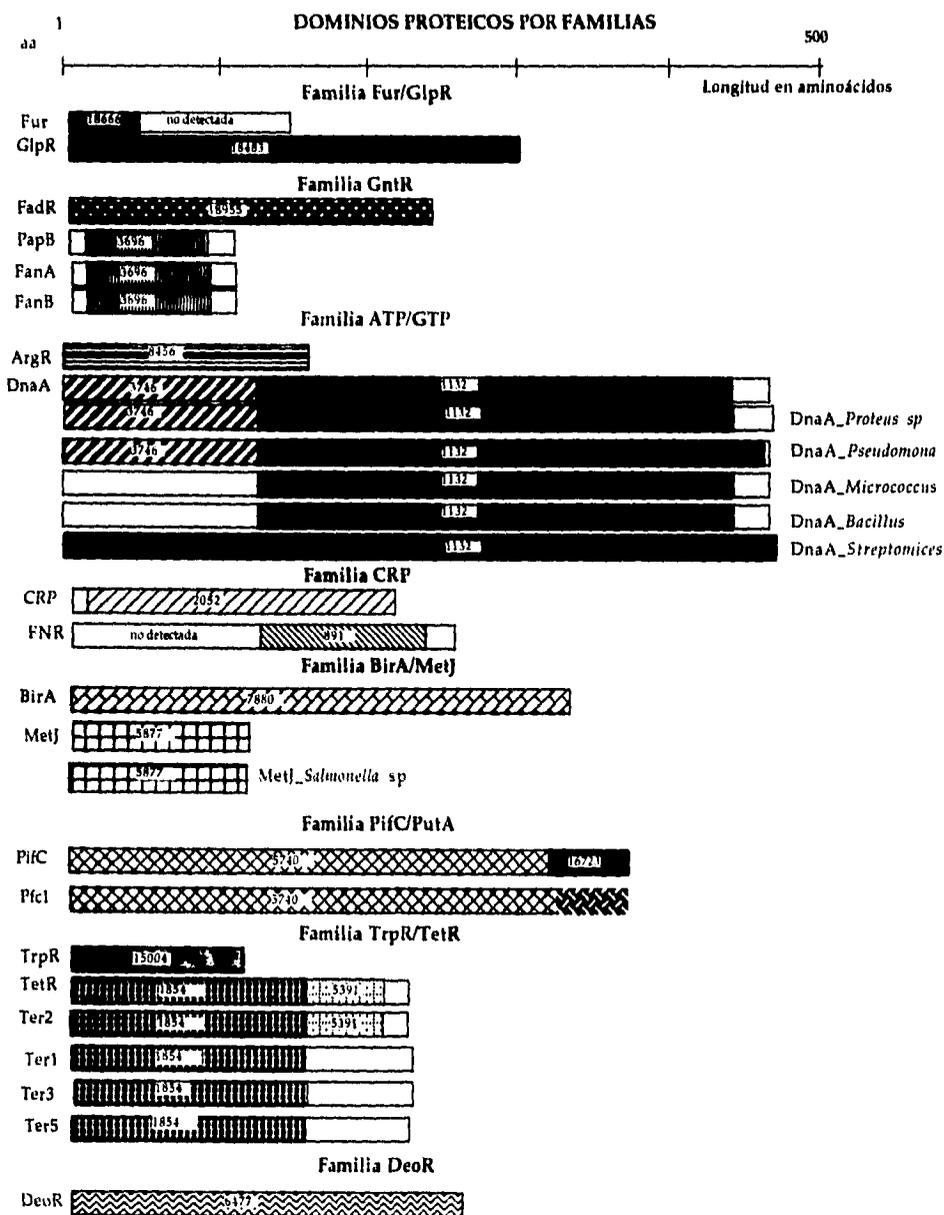


FIGURA 7. continuación...

los para cada proteína reguladora de la colección. Los dominios detectados fueron agrupados por su presencia en las respectivas familias proteicas formadas previamente, Figura 7.

Los dominios detectados al interior de cada familia pueden actuar como marcadores específicos para cada una de ellas, por ejemplo las proteínas de las familias GalR/LacI, LysR y de los dos componentes presentan dominios particulares que no son compartidos por proteínas de otras familias.

Los dominios "firma o indicadores" importantes con sus respectivas familias son los siguientes:

- El dominio 459, de la familia GalR/LacI que se ubica en la región N-terminal se conserva para los 7 miembros de la familia que forman parte de la colección de 44 proteínas reguladoras;

- El dominio 160, de la familia LysR, ubicado hacia el N-terminal y que es característico de los integrantes de esta familia y pertenecientes a nuestra colección; y

- El dominio 80, presente en los miembros de la familia de los dos componentes pertenecientes a nuestra colección.

Por otro lado, así como se presentan dominios característicos de cada familia proteica obtenidas en este trabajo que pueden servir como indicadores o firmas de un grupo en particular, también se presentan dominios que sólo se encuentran en proteínas particulares, es decir, son dominios exclusivos a ciertas proteínas, por ejemplo:

- El dominio número 8456 de la proteína ArgR (de la familia ATP/GTP) y el dominio número 7880 de BirA (de la familia BirA/MetJ) entre otros casos. Figura 7.

Así mismo, hay familias en que sus miembros muestran dominios compartidos, es decir, dominios "cruzados", por ejemplo:

- El dominio número 80 de NR(I) (de la familia de proteínas similares a sigma54), que es común a los miembros de la familia de los dos componentes; además del dominio 190 que comparte con TyrR, que es miembro de su misma familia (proteínas como sigma54).

- NarL presenta dos dominios compartidos por dos familias diferentes, un dominio compartido por los miembros de la familia de los dos componentes

(número 80) y otro dominio que es similar a los de la familia LuxR (471), Figura 7.

Continuando con la noción de dominios proteicos, se detectaron las estructuras que se pegan al DNA (tabla 3), denominadas HTH o hélice-vuelta-hélice para cada una de las proteínas de la colección de 44 proteínas reguladoras, obteniéndose que aproximadamente en 30 proteínas se detecta la estructura HTH, mientras que en el resto o no hay evidencia clara de que estructura de pegado al DNA se presenta o bien se presenta una estructura denominada hoja  $\beta$ -plegada bien determinada para el caso de MetJ y Dye.

Con base en la secuencia del dominio de pegado al DNA detectado en las proteínas de la colección se realizó un alineamiento que puede verse en la figura. 8, además de un agrupamiento en base al dendograma presentado en la figura 9.

Los grupos formados en este análisis fueron los siguientes:

- a) CynR, OxyR, IlvY, MetR, CysB y RhaR.
- b) LexA, MetJ y PapB.
- c) Fur y MalT.
- d) Lrp, TyrR, NarL y TetR.
- e) AraC y TrpR.
- f) Fis y NR(I).
- g) Ada, AraC2 e IciR.
- h) GalR, GalS, PurR, LacI, FruR, CytR, RafR y DeoR.
- i) CRP, FNR, FadR y BirA.

Las diferencias en cuanto a los grupos obtenidos por medio de este agrupamiento en relación a los grupos obtenidos por alineamiento total de secuencias, se discutirá más adelante.

Regresando a la formación de familias o grupos por medio de la comparación de secuencias totales, se adicionaron propiedades importantes para las proteínas de las diferentes familias obtenidas, que refuerzan la noción de grupos (Tabla 2), por ejemplo:

**Estructura de pegado al DNA para proteínas reguladoras de promotores sigma70 de *Escherichia coli*.**

PROTEINA	ESTRUCTURA	EVIDENCIA	REFERENCIAS
Ada	HTH		Samson, 1992.
AraC	HTH	B y D	Schleif, R. 1992.
ArcA	hoja $\beta$ -plegada	B	Breg, J. N. et al. 1990.
BirA	HTH	B y C	Brennan, R. G. et al. 1989.
CRP	HTH	A y C	Spiro, et al. 1990. Harrison, 1990
CynR	HTH		
CysB	HTH	B	Ostrowski, J. y Kredich, 1991.
CytR	HTH		Brennan y Matthews, 1989.
DeoR	HTH	B	Brennan y Matthews, 1989.
FadR	HTH		DiRusso, et al. 1991. Brennan, 1989.
Fis	HTH	A	Harrison, C.S. 1991.
FNR	HTH	B y C	Shaw, et al. 1983. Spiro, et al. 1990.
FruR	HTH		Weickert y Adhya. 1992.
Fur	HTH		
GalR	HTH	B	Shaw, et al. 1983. Weickert y Adhya. 1992.
GalS	HTH		Weickert y Adhya. 1992.
IciR	HTH		
IlvY	HTH		
KdpE	Prob. hoja $\beta$ -plegada		Bestfit-GCC
LacI	HTH	NMR	Harrison, C. S. 1990.
LexA	HTH		Gicquel-Sanzev y P.Cossart, 1992.
Lrp	HTH		
MalT	HTH	B y E	Cole, T. S. y O. Raibaud. 1986
MetJ	hoja $\beta$ -plegada A		Rafferty, et al. 1989. Somers y Philips, 1992.
MetR	HTH		
NarL	HTH		
NR(I)	HTH		Pittard y Davidson, 1991.
OmpR	Prob. hoja $\beta$ -plegada		Bestfit-GCC
OxyR	HTH		
PapB	HTH		
PhoB	Prob. hoja $\beta$ -plegada		Bestfit-GCC
PurR	HTH		Weickert y Adhya, 1992.
RafR	HTH		Aslanidis, 1990.
RhaR	HTH		
TetR	HTH	B	Brennan y Mathews, 1989.
TrpR	HTH	A	Harrison, C. S. 1990
TyrR	HTH	C	Pittard y Davidson, 1991.

**Tabla 3.** Las proteínas reguladoras de promotores sigma70 de *E. coli* exhiben una estructura de pegado al DNA, que se muestra en la tabla, así como la evidencia por la que fue detectada y las referencias correspondientes. Las proteínas que no se presentan, no hay evidencias claras de la estructura de pegado al DNA, entre ellas: ArgR, ArsR, DnaA, GlpR, PifC, PutA y TraJ.

NMR: Resonancia Magnética Nuclear. A: Difracción por rayos X. B: Comparación con secuencias de otras proteínas reguladoras. C: Mutagénesis. D: Clonación y sobreexpresión del dominio de pegado al DNA. E: Predicción de estructura secundaria. HTH: Hélice-Vuelta-Hélice.

	*	$\alpha$ Hélice 1 - Vuelta - $\alpha$ Hélice 2	*	No de aa
CynR	18	...SFTRAASA LHS QPALSQQIR.....	37	300
OxyR	18	...HFRRRAADS CHS QPTLSGQIR.....	37	305
IlyY	18	...HFGRSARA MHS PSTLSRQIQ.....	37	298
MetR	19	...SLAAAAAT LHT QSALSHQFS.....	38	318
CysB	21	...STAEGLYS QPG QPGISKQVR.....	39	325
RhaR	225	..LDKFCDEASC .S ERVLRQQFR.....	246	312
LexA	141	...NGQVVVA RID DEVTVKRLKK.....	161	203
MetJ	21	.....QV KKITVSIPLK VLKILTDERTRRQVNNLRHATNSELLCEAFLHAFT	67	106
PapB	9	...RSGNAFLL NIR ESVLLPGSM.....	38	105
Fur	114	.....IAAKH GIR LTNHSLYL.....	129	148
Malt	741	...ALKLANRT GFI SHFVIEGEA.....	760	902
Lrp	28	RISNVELSKRV GLS PTPCLERVRRLERQG.....	56	165
TyrR	481	..PSTRKLAARL GVS HTAIANKLR.....	502	514
NarL	209	...NKMIARRL DIT ESTVKVHVK.....	228	255
TetR	26	..TTRKLAQKL GIE QPTLYWHVK.....	47	218
AraC	236	...AKLLSTT RMP IATVGRNVG.....	255	294
TrpR	68	...QRELKNEL GAG IATITRGSN.....	87	109
Fis	74	...QTRAALMM GIN RGTLRKKLK.....	93	99
NR(I)	445	...QEAAALL GWG RNTLTRKLEL.....	465	468
Ada	100	..VTLEALADQV AMS PFHLHRLFK.....	121	355
AraC2	197	..DIASVAQHV CLS PSRLSHLFR.....	217	294
IciR	47	...LTELAAQQA GLP NSTTHRLLT.....	66	275
GalR	4	...IKDVARLA GVS VATVSRVIN.....	24	344
GalS	4	...IRDVARQA GVS VATVSRVLN.....	24	347
PurR	4	...IKDVAKRA NVS TTTVSHVIN.....	24	342
LacI	6	...LYDVAEYA GVS YQTVSRVVN.....	25	361
FruR	3	...LDEIARLA GVS RTTASYVIN.....	23	335
CytR	12	...MRDVALKV KVS TATVSRALM.....	31	341
RafR	3	...LKAIAATL GIS VTTVSRALG.....	23	337
DeoR	24	...LKDAAALL GVS EMTIRRDLN.....	44	352
CRP	169	...RQEIGQIV GCS RETVGRILK.....	188	211
FNR	197	...RGDIGNYL GLT VETISRLLG.....	216	251
FadR	35	...ERELESELI GVT RTTLREVLQ.....	53	240
BirA	22	...GEQLGETL GMS RAAINKHIQ.....	41	322

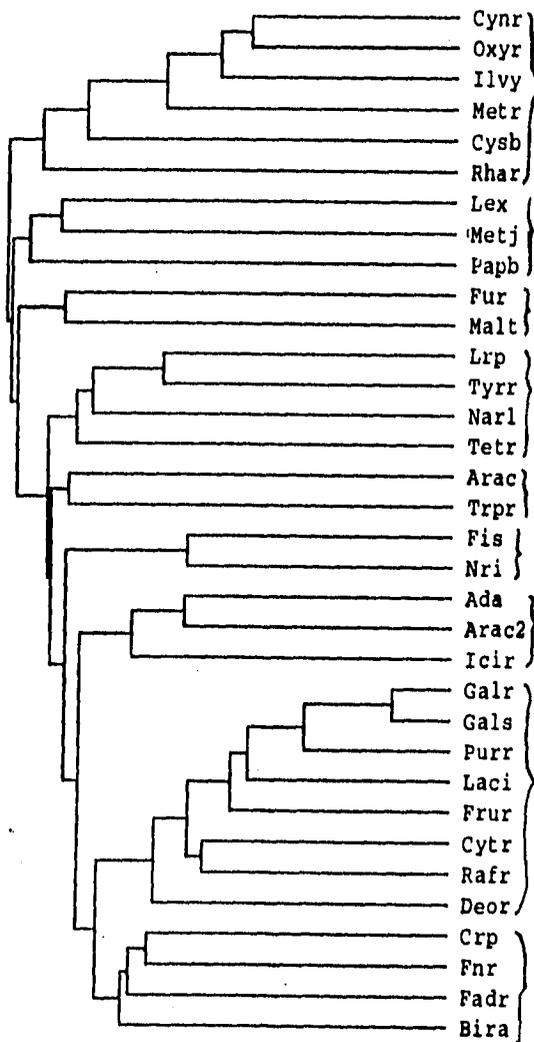
FIG. 8. Alineamiento de los dominios de pegado al DNA: hélice-vuelta-hélice y hoja- $\beta$ -plegada, para las proteínas reguladoras de promotores sigma70 de *E.coli*. Se observa la presencia de aminoácidos muy conservados, tales como Glicina, Valina o Alanina.

Se propone que la glicina en conjunto con tres aminoácidos hidrofóbicos ayudan a estabilizar el arreglo de las dos hélices en el HTH y a empaquetar el dominio contra el resto de la proteína.

\*: Posición de la estructura dentro de la proteína total,

No. de aa: Total de aminoácidos de cada proteína.

Sólo se presentan las estructuras ya detectadas en las proteínas reguladoras, ver tabla 3.



**Fig. 9.** Agrupamiento de proteínas con base en dominios de pegado al DNA. Las proteínas fueron agrupadas con base a sus dominios de pegado al DNA, por medio de Pileup-GCG. En esta figura se muestran 34 dominios de pegado al DNA, para el resto de los dominios de las proteínas que no se muestran en este análisis no se conoce la secuencia de la estructura de pegado al DNA. Todas las proteínas presentan una estructura Hélice-Vuelta-Hélice, excepto Metj que presenta una estructura hoja  $\beta$ -plegada antiparalela.

- Los integrantes de la familia LysR (de nuestra colección), se presentan como reguladores duales, es decir, son represores y activadores, entre otras propiedades.

Así como podemos encontrar familias muy estables en cuanto a secuencias y propiedades asociadas, también encontramos familias con una alta variabilidad entre sus integrantes, desde la perspectiva de propiedades regulatorias. Por ejemplo, la que agrupa a FadR y PapB.

Las proteínas reguladoras se regrouparon de acuerdo a sus respectivas familias y en relación a sus actividades regulatorias, como puede observarse en la tabla 4, con el objeto de obtener posibles pistas acerca del origen de los miembros de cada una de las familias, obteniéndose que:

- Aproximadamente la mitad de las 14 familias contienen reguladores con dos actividades, ya sea activadores y represores, activadores y duales o bien represores y duales,
- De las 20 proteínas represoras, 35% aproximadamente pertenecen a un grupo que contiene exclusivamente proteínas represoras (familia GalR/LacI),
- En el caso opuesto, sólo 28% de las 16 proteínas duales pertenecen a un grupo como LysR, que contiene exclusivamente proteínas con actividad dual, y
- Finalmente las 4 proteínas activadoras restantes pertenecen a familias con miembros represores o con actividad dual, adicionales.

Por otra parte, pensamos en la posible relación que puede existir entre las actividades regulatorias de las proteínas de la colección con respecto a la orientación del sitio de pegado al DNA u operador, para lo cual se agruparon todas las proteínas reguladoras con respecto a la orientación del operador ( es decir, repetición directa, invertida o asimetría en el sitio), Tabla 5. Se observa que:

- Aproximadamente 90% de los represores se asocian con sitios dentro del DNA que exhiben una orientación de simetría invertida,
- Aproximadamente 40% de proteínas activadoras se asocian a sitios con orientación de simetría directa, y

Proteínas agrupadas por familias y actividades regulatorias			
Familia	Actividad en promotores sigma70 de <i>E.coli</i>		
	Represor (-)	Activador (+)	Represor y Activador (±)
Arac / XylS		RhaR(2)	AraC(1)
LysR			OxyR(1), CysB(1), IlvY(1), MetR(1), CynR
CRP			CRP, FNR(1)
GalR / LacI	GalS(1), GalR(1), CytR(1), PurR(1), LacI, FruR, RafR(1)		
Dos componentes		PhoB(2), KdpE	OmpR, NarL, ArcA o Dye
Similares a Sigma54			TyR(1), Fis(1) NR(1)*
Fur / GlpR	Fur, GlpR		
AsnC	IciR		Lrp, Ada(2)
GntR			FadR, PapB(2)
ATP / GTP	ArgR(1)		DnaA(1)
BirA / MetJ	BirA(1), MetJ(1)		
LuxR	LexA(1), ArsR	MalT**	
TrpR / TetR	TetR(1), TrpR(1)		
PifC / PutA	PifC(1), PutA(1)		
TraJ		TraJ	
DeoR	DeoR(1)		

Tabla 4. Las proteínas reguladoras de promotores sigma70 son reagrupadas por familia y actividad regulatoria. Las indicaciones hechas dentro de la tabla indican lo siguiente: (1) Proteínas autoreguladas negativamente y (2) autoreguladas positivamente. (\*) Activa promotores sigma54 y (\*\*) el gene para MalT esta regulado positivamente por CRP.

- 20% de las mismas proteínas activadoras que se asocian a sitios con orientación asimétrica.

- Para el caso de las proteínas con actividad dual no hay una tendencia clara de asociación con la disposición de un sitio de pegado al DNA.

Siguiendo por esta línea de relaciones entre las propiedades necesarias para la regulación de la expresión génica, se relaciona la actividad regulatoria con respecto a la posición del dominio que se pega al DNA (Tabla 6), ya sea que se ubique en el extremo C- , N-terminal o a la mitad de la proteína (posición definida en base a la posición del dominio de pegado al DNA con respecto al resto de la proteína, ver figura 8 . Lo más relevante de la tabla donde se describe esta relación es que ninguna de las proteínas que presentan el dominio de pegado al DNA desplazado hacia el N-terminal son activadores estrictos, sino que pueden ser represores o con actividad dual. En el caso opuesto, las proteínas activadoras estrictas no presentan el dominio de pegado al DNA desplazado hacia el extremo N-terminal de la proteína.

Proteínas agrupadas por actividades regulatorias y orientación de los sitios de pegado al DNA				
Actividad	Orientación del sitio de pegado al DNA			
	Simetría Directa	Simetría Invertida	Asimetría	No determinada
Represores		Fur, GlpR, IciR, ArgR, BirA, MetJ, GalS, GalR, PurR, FruR, LacI, RafR, LexA, PifC, PutA, TrpR, TetR, DeoR		CytR, ArcA o Dye, ArsR
Activadores	RhaR, PhoB		MalT	TraJ, KdpE
Duales	AraC, DnaA, OmpR	IlvY, MetR, PapB, CRP, FNR, TyrR, Fis, NR(I)	Ada	Lrp, CynR, OxyR, FadR, NarL

Tabla 5. Las proteínas reguladoras de promotores sigma70 son agrupadas en base a la orientación del sitio de pegado al DNA y su actividad regulatoria. Las proteínas reguladoras represoras muestran una tendencia clara a la asociación con sitios de pegado al DNA con orientación simetría invertida, pero con sitios con simetría directa. No se observa la asociación de activadores con sitios de simetría invertida.

Proteínas agrupadas por posición del dominio de pegado al DNA y actividades regulatorias				
Actividad	Posición del dominio de pegado al DNA			
	C-terminal	N-terminal	Toda	No determinada
Represores	IciR, Fur, TrpR(1), ArgR, LexA(1), ArsR	GalS(1), GalR(1), CytR(1), PurR(1), LacI, FruR, RafR(1), TetR(1), DeoR(1), BirA(1)	ArcA o Dye, MetJ(1)	GlpR, PifC(1), PutA(1)
Activadores	RhaR(2), Malt(**)		PhoB(2), KdpE	TraJ
Duales	AraC(1), DnaA(1), NarL, Fis(1), NR(1)(*), TyrR(1)	CRP, FNR(1), FadR, PapB(2), OxyR(1), CysB(1), IlvY(1), MetR(1), CynR, Lrp, Ada(2)	OmpR	

Tabla 6. Las proteínas reguladoras de promotores sigma70 son agrupadas en relación a la posición del dominio de pegado al DNA y su actividad regulatoria. Aquéllas proteínas reguladoras cuyo dominio de pegado al DNA se desplaza hacia el extremo N-terminal, son represoras o duales, pero no activadoras estrictas.

El dominio de pegado al DNA se define de acuerdo a su ubicación dentro de la proteína total como puede observarse en la figura 8.

Las indicaciones dentro de la tabla indican: (1) Proteínas autoreguladas negativamente, (2) Autoreguladas positivamente, (\*) Activa promotores sigma54 y (\*\*) el gene para Malt es regulado positivamente por CRP.

## VI. DISCUSION Y CONCLUSIONES

Las proteínas son uno de los componentes más importantes en la organización de los procesos metabólicos celulares, pero la organización de esta red metabólica es sólo una de las muchas funciones que son realizadas por las proteínas, ya que desempeñan otras funciones, también muy importantes, como por ejemplo: forman parte estructural de la célula (componentes del citoesqueleto o formando parte de membranas celulares), proteínas receptoras de estímulos ambientales, anticuerpos en organismos superiores y por supuesto las proteínas reguladoras de la expresión génica (presentes no sólo en células procariontes, sino también en eucariontes y virus, entre otros sistemas biológicos). (En el apéndice 2 se pueden ver las funciones reguladas por las proteínas de la colección analizada).

La misma diversidad funcional que se exhibe en las proteínas es producto de la gran diversidad de secuencias de aminoácidos, lo que ha provocado que las clasificaciones proteicas cobren gran importancia desde el punto de vista biológico, ya que nos ayudan a comprender tanta diversidad funcional y estructural, por medio de una organización en grupos con propiedades comunes. Esta comprensión muchas veces se relaciona con pistas acerca de los cambios a través de la historia evolutiva de los organismos y de las moléculas en general. Para el caso de las proteínas reguladoras, se ha observado que una clasificación de las mismas, nos vendría a dar mayores pistas en la evolución de mecanismos tales como el de la regulación génica.

Las clasificaciones de las proteínas se basan en la comparación de la estructura primaria o secuencia de aminoácidos. Para el caso de las 44 proteínas reguladoras de la colección, la comparación de las secuencias de aminoácidos dio por resultado la formación de varios grupos que se pueden ver en las figuras 6 y 7 de la sección de resultados. Un resultado interesante del alineamiento y por ende de la formación de grupos proteicos, es la exhibición de heterogeneidad entre las proteínas reguladoras, refiriendonos por heterogeneidad a una amplia diversidad de secuencias dentro de la estructura primaria

de cada una de las proteínas de la colección, así como también podemos observar la presencia de segmentos o dominios comunes como la estructura que se pega al DNA denominado Hélice-Vuelta-Hélice, y que no es exclusiva de las proteínas reguladoras de ésta colección.

La heterogeneidad entre las secuencias de las proteínas reguladoras de la colección mencionada anteriormente podemos comprenderla en base al mecanismo de evolución por módulos o dominios que se plantea para las proteínas en general. El mecanismo se plantea como la posibilidad de una evolución en segmentos proteicos o módulos, por lo cual lo que realmente vemos hoy día es el resultado de una diversidad de historias evolutivas asociadas en una unidad proteica.

En resumen, los alineamientos de las secuencias completas de las proteínas reguladoras de la colección, condujo a la formación de 14 grupos, que se pueden observar en conjunto con algunas propiedades asociadas en la tabla 2 de la sección de resultados. La misma heterogeneidad de secuencias condujo a la formación de este alto número de grupos. En promedio podemos encontrar 3 proteínas reguladoras de esta colección por grupo formado.

Aunado a la diversidad de secuencias de aminoácidos para las proteínas reguladoras de la colección es necesario precisar los criterios mínimos para que se pueda realizar una clasificación que pueda proporcionarnos la mayor cantidad de información posible acerca de la regulación de la expresión génica. Entre los criterios de clasificación que consideramos como importantes, son:

- porcentaje de similitud entre secuencias de proteínas reguladoras,
- detección del número y ubicación de los diferentes dominios para cada proteína (para lo cual "ProDom o Protein Domains database" será una buena herramienta). Figura 7.
- organización de los genes regulados por las proteínas reguladoras de la colección, esto es, si se encuentran arreglados en operones, regulones o modulones.
- Actividad regulatoria: Activadora, Represora o con ambas actividades.
- Ubicación del dominio de pegado al DNA dentro de la proteína.

La característica más generalizada dentro de las proteínas de nuestra colección, es la

presencia de una secuencia corta de aproximadamente 20 aminoácidos, con excepción de 6 proteínas (5 son miembros de la familia de los dos componentes y uno miembro de la familia BirA/MetJ) que parece que exhiben una estructura de pegado al DNA diferente y de mayor tamaño al HTH.

Las proteínas reguladoras presentan diferentes dominios involucrados en la regulación de la expresión génica, que podemos agruparlos como sigue:

- a) Dominio de pegado al DNA, ya sea una estructura Hélice-Vuelta-Hélice (o HTH) o bien una estructura a manera de hoja  $\beta$ -plegada antiparalela.
- b) Dominio de multimerización.
- c) Dominio de reconocimiento de pegado a metabolitos, tales como grupos fosfato o bien carbohidratos, entre otros, es decir, dónde se detecta una señal fisiológica.

La detección de dominios probables para cada proteína reguladora es de suma importancia puesto que podemos reconstruir como se ha dado la evolución de las proteínas no sólo reguladoras, sino de todas en general, así de sus propiedades reguladoras como asociación a los diferentes módulos o dominios proteicos.

Por otra parte, la tabla 4 muestra la asociación de familias de proteínas reguladoras con respecto a su función, los resultados podrían implicar un origen evolutivo común si realmente las familias proteicas muestran un origen común como lo menciona Chothia, C. 1992 (ver la sección de Antecedentes).

La mayoría de los activadores y reguladores duales pertenecen a familias que además presentan miembros con una función regulatoria diferente (proteínas represoras o duales), esto es, en los miembros de estas familias se han presentado cambios de actividad reguladora:

De positivos a duales, de duales a positivos o negativos, etc.

Estos cambios no parecen ser una regla común dentro del conjunto de proteínas represoras, debido a que se muestran más estables como clase, evolutivamente hablando, que las otras clases; como ejemplo tenemos la familia GalR/LacI, que presenta a todos sus miembros como represores, pero ningún miembro es activador o presenta actividad dual.

Podemos pensar que los reguladores duales derivaron de ancestros activadores (o quizás el proceso se dió a la inversa), más que de ancestros represores. Esta idea la podemos constatar con base a varias evidencias, tales como:

- Las familias con miembros represores se mantienen muy estables, sin cambios a través de la evolución, tal como se muestra en la familia GalR/LacI,

- En el conjunto total de las proteínas reguladoras hay más proteínas represoras que activadoras. De 44 proteínas, sólo 5 son estrictamente activadoras, mientras que 20 son represoras estrictas.

La posibilidad de cambios dentro de la distribución de actividades antes mencionada para la colección de las proteínas reguladoras se ha contemplado, sin embargo existe la certeza de que aun con descubrimientos de nuevas funciones para las proteínas, la distribución no cambiará radicalmente, sobre todo por las siguientes razones:

- Un gran número de represores como: LacI, LexA, ArgR y GalR, y de activadores como MalT, se han estudiado cuidadosamente desde hacia ya varios años, por lo cual es difícil la descripción de nuevas funciones regulatorias asociadas a estas proteínas,

- La autoregulación negativa podría ser una fuerza de selección para cambiar un activador en proteína con actividad dual (debido a que para la mayor parte de las proteínas con actividad dual de ésta colección, la regulación de su propio gene es de tipo negativa ,ver tabla 4), y para varios de estos reguladores, la regulación de su propio gene ya se ha caracterizado,

- Las proteínas con ambas actividades no se modificarán en cuanto a su actividad se refiere por actividades adicionales ya que presentan ambas actividades, es decir, presentan la propiedad de ser represores y activadores.

La probable selección evolutiva para cambiar un activador en proteína dual se debe a la necesidad de autoregulación negativa:

De las 29 proteínas reguladoras en dónde la regulación de su propio gene está conocida:

- 24 son autoreguladas negativamente,

- 4 son autoreguladas positivamente, y

- MalT que es regulado positivamente por CRP.

Lo anterior muestra una clara tendencia a la regulación negativa de un gene regulador.

Las proteínas reguladoras son generalmente producidas en pequeñas cantidades en la célula. La teoría de la demanda de la expresión génica predice proteínas que son requeridas en bajas cantidades, para estar bajo control de regulación negativa, mientras que en el caso inverso, proteínas o genes requeridos en grandes cantidades están bajo control positivo (Savageau, M., 1983). En relación a esta teoría, nueve de las 16 proteínas duales son autoreguladas negativamente, y 3 lo son positivamente (incluida NR(I)), mientras que dos activadores estrictos son autoregulados positivamente (incluida MalT) tabla 4. Una comparación más cuantitativa de la cantidad de producción de estas proteínas se requerirá para evaluar si la teoría de la demanda es suficiente para explicar estas diferencias en la regulación de los genes reguladores o si hay restricciones adicionales involucradas en la conservación de estas propiedades regulatorias.

Las proteínas pudieron haber evolucionado como resultado de la separación de los diferentes dominios de las proteínas ancestrales -idea relacionada a la evolución proteica en base a dominios o módulos-, debido a que los grupos proteicos obtenidos previamente por comparación de secuencias completas (Figura 6) se modifican cuando se comparan los dominios de pegado al DNA para las proteínas reguladoras (Figura 9), por ejemplo: TetR y TrpR, dos reguladores de la misma familia están separados en términos de su HTH, así como Lrp y Ada ( que pertenecen a la misma familia y que son dos reguladores con actividad dual), AraC, una proteína con actividad dual y RhaR un activador de la misma familia (AraC/XylS) quedan separadas en términos de su HTH; o bien proteínas en familias más estables que se mantienen sin variaciones o con algún nuevo integrante con propiedades similares, tal es el caso de la familia GalR/LacI que se conserva como tal, con la adición de DeoR, un represor; sugiriendo un posible origen común para las proteínas reguladoras de esta familia y que corroborado por el análisis realizado por Vartak, et. al., 1991. Otro caso es el de los reguladores de la familia LysR, con actividad dual en su totalidad, se retienen estables como un grupo basado en su dominio de pegado al DNA, con adición de RhaR, un activador estricto adicionado a este grupo. En este caso, aunque los miembros de la familia pudieron haber sido activadores que adquirieron más tarde la función de represores, emergieron posiblemente por duplicación génica de un ancestro común que presentaba la capacidad de activar la

transcripción.

Así, la historia evolutiva de las proteínas reguladoras podría no ser homogénea, lo que hace sentido con la hipótesis evolutiva por dominios para las proteínas en general. Desafortunadamente, no se dispone de información suficiente para proteínas de otros grupos.

Por otro lado, uno de los pocos artículos que se aventuran a hacer una hipótesis acerca del origen de las proteínas reguladoras, Raibaud (1989) sugiere que CRP fue inicialmente un represor que se convirtió después en un activador. Raibaud pensó en esta posibilidad, a partir de la observación de que CRP se pega a sitios dentro del DNA con simetría invertida, mientras que diversos activadores se pegan a sitios con simetría directa. Las proteínas se pegan frecuentemente a sitios de aproximadamente 20 pares de bases, con aproximadamente 10 pares de bases de pegado por un monómero y la otra mitad por un segundo monómero, de una proteína dimérica en solución. La posición cara a cara de los monómeros correlacionará con la orientación del sitio de pegado al DNA, siendo de simetría directa o invertida (ver Raibaud, O. 1989). La observación de que los activadores se pegan a sitios con simetría directa se pretende explicar suponiendo una interacción débil entre sitios heterólogos que la simetría implica, comparada a una fuerte interacción entre sitios isólogos para proteínas que se pegan a sitios con simetría invertida. Los activadores que se pegan a sitios con simetría directa dimerizarán con menos frecuencia que las proteínas que se pegan a sitios con simetría invertida. Raibaud agrega que es más común encontrar proteínas como monómeros en solución y que sean activadoras o reguladoras duales, esto incluye a MalT, Ada y DnaA. Nosotros catalogamos a Ada y DnaA como reguladores duales, mientras que de los reguladores positivos estrictos sólo se dispone de información del estado multimérico para MalT, un monómero, y RhaR, un dímero. (Tobin y Schleif, 1990).

En nuestra colección, 18 de 22 represores se pegan a sitios con simetría invertida en el DNA y sólo MetJ, se pega a repeticiones que no se han determinado como directas o invertidas (Rafferty et. al., 1989). De 16 proteínas duales, 7 se asocian a sitios con simetría invertida. 2 de los activadores se pegan a sitios con simetría directa, mientras que para las otras 2 la simetría no está clara todavía. Tabla 5.

A manera de resumen, podemos decir que tenemos evidencias para suponer que las proteínas con actividad dual provinieron de proteínas activadoras, tales evidencias son: el bajo número de proteínas activadoras estrictas en la colección, parece ser que una fuerza de selección es la autoregulación, predominando la del tipo negativa en la mayoría de las proteínas de nuestra colección.

Por lo que respecta a la relación existente entre los sitios de pegado al DNA y la multimerización de las proteínas, podemos decir que las proteínas represoras presentan simetría estructural (es decir asociados en pares, ya sea que se asocien en dímeros, tetrameros, etc), así como simetría en el sitio de pegado en el DNA, predominando la asociación con la de orientación de repetición inversa. Podríamos pensar de acuerdo con esto, que los sitios de pegado para las proteínas en el DNA, determinan el tipo de arreglo espacial de los multímeros, ya que de alguna manera podrían determinar como va a ser la orientación de las subunidades proteicas con respecto al DNA y con respecto al resto de la proteína con la cual interaccionan.

En otro orden de ideas, es interesante ver la posición del dominio de pegado al DNA dentro de las proteínas reguladoras, con relación a la actividad asociada (represor, activador o dual). Los resultados de la tabla 6 muestran una mayor tendencia a la asociación de proteínas represoras a presentar el dominio de pegado al DNA en posición N-terminal, mientras que ninguna proteína activadora estricta presenta el dominio de pegado al DNA en ese extremo.

Dentro de las proteínas represoras, tenemos que se presenta la familia LacI/GalR, una familia muy estable evolutivamente hablando lo que trae consigo la importancia de tal relación funcional.

Posiblemente la posición del dominio este involucrada con la posición de otro dominio que se da a la tarea de reconocer metabolitos necesarios para la actividad de la proteína, sin embargo, esta línea de investigación todavía no esta clara.

La actividad de reguladores de la expresión génica de las proteínas de la colección es una propiedad inherente de las mismas, sin que esto quiera decir que es la única función que desempeñan.

Uno de los aspectos tomados en cuenta al estudiar la colección de proteínas reguladoras como fuente de información lo constituye la presencia de ciertos aminoácidos, tales como los aminoácidos hidrofóbicos que se presentan en alta cantidad y que determinan la función y localización de las proteínas dentro de la célula, es decir, en base al contenido de aminoácidos podemos tener proteínas asociadas a membrana como el caso de PutA o bien pueden ser proteínas asociadas a DNA como lo son todas las proteínas reguladoras de la colección con la cual se realizó el trabajo, o bien proteínas que ayudan a la transferencia del plásmido F de una bacteria a otra, tal como TraJ.

## VII. PERSPECTIVAS

El presente trabajo forma parte del proyecto de maestría que se está realizando en el Centro de Investigación Sobre Fijación del Nitrógeno y que tiene por objetivo el tratar de encontrar posibles relaciones funcionales entre todos los elementos involucrados en la regulación de la expresión génica, tales como operadores, promotores asociados al factor sigma70, genes y proteínas reguladoras en *Escherichia coli*.

Los resultados nos muestran que podemos seguir varias líneas de investigación, que en algunos casos ya se empezaron a abordar, casos concretos son los siguientes:

- Análisis evolutivo de las secuencias de las 44 proteínas reguladoras de la colección por medio del software Phylip;
- Búsqueda de secuencias de proteínas reguladoras de promotores sigma70 de otros organismos no *E.coli*, tales como *Salmonella typhimurium*, *Klebsiella aerogenes*, entre otros;
- Reunir evidencias para poder explicar la relación existente entre la posición del dominio de pegado al DNA y la actividad asociada para las proteínas reguladoras, siendo las siguientes evidencias las siguientes: Posición del HTH para proteínas reguladoras (en general y no sólo de *E.coli*), de acuerdo a los extremos N- o C-terminal, así como las actividad regulatoria asociada (si es represor o activador o dual);

Por otro lado, y de orden más técnico debemos de hacer mención que se esta realizando una base de datos con información de regulación génica, con el objeto de ordenar las propiedades de interés para un análisis más detallado de los mecanismos y componentes de la regulación de la expresión génica.

Finalmente, tenemos que mencionar que el trabajo se continuará como una línea importante del proyecto de maestría.

## VI. REFERENCIAS

- Allard, J. D., M. L. Gibson, L. H. Vu, T. T. Nguyen y K. P. Bertrand. 1993. Nucleotide sequence of class D tetracycline resistance genes from *Salmonella ordonez*. *Mol. Gen. Genet.* 237: 301-305.
- Amouyal, M., L. Mortensen, H. Buc y K. Hammer. 1989. Single and double loop formation when *deoR* repressor binds to its natural operator sites. *Cell.* 58: 545-551.
- Andrews, A. E., B. Dickson, B. Lawley, C. Cobbett y A. J. Pittard. 1991. Importance of the position of TYR R boxes for repression and activation of the *tyrP* and *aroF* genes in *Escherichia coli*. *J. Bacter.* 278(16): 5079-5085.
- Aslanidis, Ch., I. Muizniekis y R. Schmitt. 1990. Successive binding of *raf* repressor to adjacent *raf* operator sites in vitro. *Mol. Gen. Genet.* 223:297-304.
- Branden, C. y J. Tooze. 1991. Introduction to Protein Structure. Garland Publishing, Inc. New York and London. p. 85-110.
- Brennan, R. G. y B. W. Matthews. 1989. The Helix-Turn-Helix DNA Binding Motif. *J. Biol. Chem.* 264(4): 1903-1906.
- Brennan, R. G., S. Vasu, B. W. Matthews y A. J. Otsuka. 1989. Crystallization of the bifunctional biotin operon repressor. *J. Biol. Chem.* 264(1): 5.
- Breg, J. A., J. H. J. Opheusden, M. J. M. Burgering, R. Boelens y R. Kaptein. 1990. Structure of Arc repressor in solution: evidence for a family of  $\beta$ -sheet DNA-binding proteins. *Nature.* 346: 586-589.
- Brissette, R. E. K. Tsung y M. Inouye. 1991. Supression of a Mutation in *OmpR* at the putative phosphorylation center by a mutant *EnvZ* protein in *Escherichia coli*. *J. Bacter.* 173(2): 601-608.
- Charlier, D. 1992. Arginine regulon of *Escherichia coli* K-12. *J. Mol. Biol.* 226: 367-386.
- Chothia, C. 1992. One thousand families for the molecular biologist. *Nature.* 357: 543-544.

- Cole, T. S. y O. Raibaud. 1986. The nucleotide sequence of the *malT* gene encoding the positive regulator of the *Escherichia coli* maltose regulon. *Gene*. 42:201-208.
- Collado-Vides, J., Magasanik, B. y J. D. Gralla. 1991. Control site location and transcriptional regulation in *Escherichia coli*. *Microb. Rev.* 55(3): 371-394.
- Cronnan, J. E. 1989. The *E. coli bio* operon: transcriptional repression by an essential protein modification enzyme. *Cell*. 58: 427-429.
- Drury, L. S. y R. S. Buxton. 1985. DNA sequence analysis of the *dye* gene of *Escherichia coli* reveals amino acid homology between the Dye and OmpR proteins. *J. Biol. Chem.* 260(7): 4236-4242.
- Ernsting, B., J. W. Denninger, R. M. Blumenthal y R. G. Matthews. 1993. Regulation of the *glbBDF* operon of *Escherichia coli*: How is a leucine-insensitive operon regulated by the leucine-responsive regulatory protein?. *J. Bacter.* 175(22): 7160-7169.
- Feng, S. y J. A. DeMoss. 1987. Promoter region of the *nar* operon of *Escherichia coli*: Nucleotide sequence and transcription initiation signals. *J. Bacter.* 169(10): 4614-4620.
- Fuller, R., B. E. Funnell y A. Kornberg. 1984. The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell*. 38: 889-900.
- Göransson, M., K. Forsman, P. Nilsson y B. E. Uhlin. 1989. Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in *Escherichia coli*. *Mol. Microb.* 3(11): 1557-1565.
- Gralla, J. D. 1990. Promoter Recognition and mRNA Initiation by *Escherichia coli*  $\sigma^{70}$ . *Methods in Enzymology*. 185:37-54.
- Gralla, J. D. y J. Collado-Vides. Organization and function of transcription regulatory elements. en prensa.
- Gross, C. A., M. Lonetto y R. Losick. 1992. Bacterial sigma factors. *Transcriptional Regulation*. Vol. I. CSHL. p. 129-176.
- Gunsalus, R. P. y C. Yanofsky. 1980. Nucleotide sequence and expression of *Escherichia coli trpR*, the structural gene for the *trp* aporepressor. *PNAS*. 77(12): 7117-7121.

- Hansen, D. y W. Hillen. 1987. Tryptophan in  $\alpha$ -helix 3 of Tet repressor forms a sequence-specific contact with *tet* operator in solution. *J. Biol. Chem.* 262(25): 12269-12274.
- Harrison, S. C. 1991. A structural taxonomy of DNA-binding domains. *Nature.* 353: 715-718.
- Harrison, S. C. y A. K. Aggarwal. 1990. DNA Recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* 59: 933-69.
- Haydon, D. y J. Guest. 1991. A new family of bacterial regulatory proteins. *FEMS Microb. Letters.* 79: 291-296.
- Henikoff, S., G. W. Haughn, J. M. Calvo y J. Wallace. 1988. A large family of bacterial activator proteins. *PNAS.* 85: 6602-6606.
- Henry, M. F. y J. E. Cronnan Jr. 1992. A new mechanism of transcriptional regulation: release of an activator triggered by small molecule binding. *Cell.* 70: 671-679.
- Hoopes y MacClure. 1987. Strategies in regulation of transcription initiation. in *Escherichia coli* and *Salmonella typhimurium*. Vol. II
- Kennedy, M., M. Chandler y D. Lane. 1990. Mapping and regulation of the *pifC* promoter of the F plasmid. *Bioch. et Biophys. Acta.* 950: 75-80.
- Lamblin, J. A-F. y J. A. Fuchs. 1993. Expression and purification of the *cynR* regulatory gene product: CynR Is a DNA-binding protein. 175(24): 7990-7999.
- Lee, N., Ch. Francklyn y E. P. Hamilton. 1987. Arabinose-induced binding of AraC protein to *araI2* activates the *araBAD* operon promoter. *PNAS.* 98: 8814-8818.
- Lehninger, A. L. 1975. Biochemistry. 2nd. edition. Worth Publishers, Inc. New York, U. S. A.
- Lobell, R. B. y R. F. Schleif. 1990. DNA looping and unlooping by AraC protein. *Science.* 250:528-532.
- Maeda, S., K. Takayanagi, Y. Nishimura, T. Naruyama, K. Sato y T. Mizuno. 1991. Activation of the osmoregulated *ompC* gene by the OmpR protein in *E.coli*: A study involving synthetic OmpR-binding sequences. *J. Biochem.* 110: 324-327.
- Maeda, S., y T. Mizuno. 1990. Evidence for multiple OmpR-binding sites in the upstream activation sequence of the *ompC* promoter in *E. coli*: a single OmpR-binding site is capable of activating the promoter. *J. Bacter.* 172(1): 501-503.

- Maloy, S. 1991. The proline utilization operon. *Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology. Tomo II
- Menzel, R. y J. Roth. 1981. Enzymatic properties of the purified *putA* protein from *Salmonella typhimurium*. J. Biol. Chem. 256(18): 9762-9766.
- Miller, S. 1989. The structure of interfaces between subunits of dimeric and tetrameric proteins. Protein Engineering. 3(2): 77-83.
- Nath, S. K. 1988. Attenuation of transcription of biotin genes in *E. coli*. Can. J. Microb. 34: 1288-1296.
- Neidhart F. C., J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter y H. E. Umbarger. 1987. *Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology. Vol. II. p. 1313-1317, 1231-1240, 1439-1526.
- Neidhart F. C., J. L. Ingraham y M. Schaechter. 1990. Physiology of the bacterial cell. A Molecular Approach. SINAUER. p. 351-388.
- Ogden, S., D. Haggerty, C. M. Stoner, D. Kolodrubetz y R. Schleif. 1980. The *E. coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. PNAS. 77(6): 3346-3350.
- Ostrowski, J. y N. M. Kredich. 1991. Negative autoregulation of *cysB* in *Salmonella typhimurium*: In vitro interactions of CysB protein with the *cysB* promoter. J. Bacter. 173(7): 2212-2218.
- Otsuka, A. y J. Abelson. 1978. The regulatory region of the biotin operon in *E. coli*. Nature. 276(14): 689-694.
- Pedersen, H., L. Sogaard-Andersen, B. Holst, P. Gerlach. E. Bremer y P. Valentin-Hansen. 1992. cAMP-CRP activator complex and the CytR repressor protein bind co-operatively to the *cytRP* promoter in *E. coli* and CytR antagonizes the cAMP-CRP-induced DNA bend. J. Mol. Biol. 227: 396-406.
- Pittard, A. J. y B. E. Davidson. 1991. TyrR protein of *Escherichia coli* and its role as repressor and activator. Mol. Microb. 5(7):1585-1592.
- Rafferty, J. B., W. S. Somers, I. Saint-Girons y S. E. V. Phillips. 1989. Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor. Nature. 341: 705-710.
- Raibaud, O. y M. Schwartz. 1984. Positive control of transcription initiation in bacteria. Ann. Rev. Genet. 18: 173-206.

- Raibaud, O. 1989. Nucleoprotein structures at positively regulated bacterial promoters: homology with replication origins and some hypotheses on the quaternary structure of the activator proteins in these complexes. *Mol. Microb.* 3(3): 455-458.
- Ramos, J. L., F. Rojo, L. Zhou y K. N. Timmis. 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and *Escherichia coli* AraC activators. *NAR.* 18(8): 2149-2152.
- Ramseier, T. M. D. Nègre, J-C. Cortay, M. Scarabel, A.J. Cozzone y M. H. Saier Jr. 1993. In vitro binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru*, *pps*, *ace*, *pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*. *J. Mol. Biol.* 234: 28-44.
- Reznikoff, W. S., D. A. Siegele, D. W. Cowing y C. A. Gross. 1985. The Regulation of Transcription Initiation in Bacteria. *Ann. Rev. Genet.* 19:355-87.
- Richet, E. y O. Raibaud. 1989. MalT, the regulatory protein of the *Escherichia coli* maltose system, is an ATP-dependent transcriptional activator. *EMBO Journal.* 8(3): 981-987.
- Rolfes, R. J. y H. Zalkin. 1990. Autoregulation of *Escherichia coli purR* requires two control sites downstream of the promoter. *J. Bacter.* 172(10): 5758-5766.
- Ross, W., J. F. Thompson, J. T. Newlands y R. L. Gourse. 1990. *E. coli* Fis protein activates ribosomal RNA transcription in vitro and in vivo. *EMBO Journal.* 9(11): 3733-3742.
- Sankar, A. y S. Garges. 1990. Positive control. *J. Biol. Chem.* 265(19): 10797-10800.
- Savageau, M. 1983. Regulation of differentiated cell-specific functions. *PNAS.* 80: 1411-1415.
- Schleif, R. 1992. Regulation of the L-arabinose catabolic operon *araBAD*. in *Transcriptional regulation.* p. 643-665.
- Shaw, D. J., D. W. Rice y J. R. Guest. 1983. Homology between CAP and FNR, a regulator of anaerobic respiration in *Escherichia coli*. *J. Mol. Biol.* 166:241-247.
- Shevell, D. E. y G. Walker. 1991. A region of the Ada DNA-repair protein required for the activation of *ada* transcription is not necessary for activation of *alkA*. *PNAS.* 88: 9001-9005.

- Søgaard-Andersen, L. y P. Valentin-Hansen. 1993. Protein-Protein interactions in gene regulation: The cAMP-CRP complex sets the specificity of a second DNA-binding protein, the CytR repressor. *Cell*. 75: 557-566.
- Somers, W. S. y S. E. V. Phillips. 1992. Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by  $\beta$ -strands. *Nature*. 359: 387-393.
- Sonnhammer, E. L. L. y D. Kahn. 1994. The modular arrangement of proteins as inferred from analysis of homology. *Prot. Science*. 3: 482-492.
- Spiro, S. y Gaston, K. L. 1990. Interconversion of DNA-binding specificities. *Mol. Microb.* 4(11): 1831-1838.
- Stock, J. B, A. J. Ninfa y A. M. Stock. 1989. Protein phosphorylation of adaptive responses in bacteria. *Microb. Rev.* 53(4): 450-490.
- Storz, G., L. A. Tartaglia y B. N. Ames. 1990. Transcriptional regulator of oxidative stress-inducible genes: direct activation of oxidation. *Science*. 248: 189-194.
- Vartak, N. B., J. Reizer, A. Reizer, J. T. Gripp, E.A. Groisman, L.-F. Wu, J. M. Tomich y M. H. Saier, Jr. 1991. Sequence and evolution of the FruR protein a pleiotropic transcriptional regulatory protein possessing both activator and repressor functions which is homologous to the periplasmic ribose-binding protein. *Res. Microbiol.* 142: 951-963.
- Vidal-Ingigliardi, D. y O. Raibaud. 1991. Three adjacent binding sites for cAMP receptor protein are involved in the activation of the divergent *malEp-malKp* promoters. *PNAS*. 88: 229-233.
- Voelkner, P., W. Pupper y K. Alhendorf. 1993. Characterization of the KdpD protein, the sensor kinase of the K<sup>+</sup>-translocating *Kdp* system of *Escherichia coli*. *Eur. J. Biochem.* 217: 1019-1026.
- Von Hippel, P. H., T. D. Yager y S. C. Gill. 1992. Quantitative aspects of the transcription cycle in *Escherichia coli*. *Transcriptional Regulation Vol. I. CSHL*. p. 179-201
- Walker, M. S. y J. A. DeMoss. 1992. Role of alternative promoter elements in transcription from the *nar* promoter of *Escherichia coli*. *J. Bacter.* 174(4): 1119-1123.

- Wang, Q. y J. M. Kanugi. 1987. Transcriptional repression of the *dnaA* gene of *E. coli* by *dnaA* protein. *Mol. Gen. Genet.* 209: 518-525.
- Wang, Q. y J. M. Kanugi. 1989. *dnaA* protein regulates transcription of the *rpoH* gene of *E. coli*. *J. Biol. Chem.* 264(13): 7338-7344.
- Wanner, B. L. 1987. Phosphate regulation of gene expression in *Escherichia coli*. *Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology. Vol. II
- Weickert, M. J. y S. Adhya. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* 267(22): 15869-15874.



continuación...

	201				250
Ecoara	.....	.....	.....	.....	.....
Ecrhac	.....	.....	.....	.....	.....
Ecfur	.....	.....	.....	.....	.....
Ecoglpr	.....	.....	.....	.....	.....
Ecoicir	.....	.....	.....	.....	..MVAPIP
Ecolrp	.....	.....	.....	.....	.....
Ecada	LTTDQRWQSV	LARDPNADGE	FVFAVRTTGI	FCRPSCRARH	ALRENVSYFA
Ecocynr	.....	.....	.....	.....	.....
Ecokyr	.....	.....	.....	.....	.....
Ecocysba	.....	.....	.....	.....	.....
Ecoilvc	.....	.....	.....	.....	.....
Ecometr	.....	.....	.....	.....	.....
Ecfadr	.....	.....	.....	.....	.....
Ecpapak	.....	.....	.....	.....	.....
Ecoargr	.....	.....	.....	.....	.....
Ecodnaa	KPVTQTPQAA	VTSNVAAPAQ	VAQTQPQRAA	PSTRSGWDNV	PAPAEPYRS
Ecocrp	.....	.....	.....	.....	.....
Econirr	.....	.....	.....	.....	.....
Ecobira	.....MKD	NTVPLKLIAL	LANGEFHSGE	QLGNTLGMSR	AAINKHIQTL
Ecometja	.....	.....	.....	.....	.....
Ecgal5	.NITIRDVAR	QAGVSVATVS	RVLNNSL..	.VSADTREAV	MKAVSELDYR
Ecogal	.MATIKDVAR	LAGVSVATVS	RVINNSPK..	.ASEASRLAV	HSAMESLSYH
Ecocytr	TAATMKDVAL	KAKVSTATVS	RALMNPDK..	.VSQATRNRV	EKAAREVGYL
Ecopurr	.MATIKDVAR	RANVSTTVS	HVINKTRF..	.VAEETRNAV	WAAIKELHYS
Eclac	KPVTLYDVAE	YAGVSYQTVS	RVVNQASH..	.VSAKTRKRV	EAAAMELNYI
Ecorafro	.MSLKALAT	TLGISVTVS	RALGGFSD..	.VAASTRETV	EAEARRRGYR
Ecfrrr	.VKLDEIAR	LAGVSRRTAS	YVINGKAKQY	RVSDKTVÉKV	MVVREHNYH
Ecgln	NGPTTDIIAK	PAMQDVFRII	GRLSRSSISV	LINGE <u>SGTGK</u>	ELVAHALHRH
Ecotyrr	SASFQIVAVS	PKMKHVVEQA	QKLAMLSAPL	LITGDTGTGK	DLFAYACHQA
Ecofis	.....	.....	.....	.....	.....
Ecoompr	.....	.....	.....	.....	.....
Ecphob	.....	.....	.....	.....	.....
Ecodye	.....	.....	.....	.....	.....
Ecokdpe	.....	.....	.....	.....	.....
Ecnarl	.....	.....	.....	.....	.....
Ecarsr	.....	.....	.....	.....	.....
Ecomalt	.....	.....	.....	.....MLIP	SKLSRPVRLD
Ecolexa	.....	.....	.....	.....	.....
Ecpife	.....	.....	.....	.....	.....
Ecpup	.....	.....	.....	.....	.....
Ecotrpr	.....	.....	.....	.....	.....
Ectetra	.....	.....	.....	.....	.....
Ecetraaj	.....	.....	.....	.....	.....
Ecdeor	.....	.....	.....	.....	.....

La estructura de pegado al DNA o Hélice-Vuelta-Hélice (HTH) del residuo 205 al 214, para el grupo de GalS, GalR, CytR, PurR, lacI, RafR y FruR; mientras que para BirA, se presenta del 217 al 248. Además se presenta una región rica en Valina para las proteínas antes mencionadas, excepto en BirA, que va del residuo 208-240.

En RafR hay tres aminoácidos involucrados en el reconocimiento del operador, que son: V216, T217, R221; localizados en la hélice de reconocimiento del HTH. (Aslanidis, 1990).

Para el caso de Gln (NR(I)) y TyrR, se presenta un dominio de pegado de ATP que va de los residuos 214 a 221. Así mismo se presenta un dominio de interacción con el factor sigma 54 en Gln(NR(I)), que va de los residuos 225 al 247. (Austin, 1991)

continuación...

	251	300
Ecoara	.....	SM AE....AQND
Echrac	.....	MAFCN NANLLNVFVR HIANNQLRSL AEVATVAHQ
Ecfur	.....	.....MTDMN
Ecogfpr	.....	.....
Ecoicir	AKRGRKPAVA TAPATQQVQS LTRGLKLEW IA.....E	SNGSVALTEL
Ecolrp	.....	MVDS KRRPGKDLDR IDRNILNELQ KDGRISNVEL
Ecada	NASEALAAGF RPKRCQPEK ANAQHRLDK ITHAC RLLE	QETPVTLEAL
Ecocynr	.....	.....MLSRHI
Ecocyx	.....	.....MMIRDL
Ecocysba	.....	.....MKLQQL
Ecoilvc	.....	.....VDLRDL
Ecometr	.....	.....MIEVKHL
Ecfadr	.....	.....
Ecpapak	.....	.....
Ecoargr	.....	.....
Ecodnaa	NVNVKHTFDN FVEGKSNQLA RAAARQVADN PGGAYNPLFL	<u>YGGTGLGKTH</u>
Ecocrp	.....	.....M
Econirr	.....	MIPEKRRI RRIQSGGCAI <u>HQDCSISQL</u>
Ecobira	RDWGVDFVTV PGKGYSLEPE IQLLNKQIL GQLDGGSVAV	LPVIDSTNGY
Ecometja	.....	.....
Ecgal5	PNANAQALAT QVSDTIGVVV MDVSDAFFGA LVKAV....	DLVAQQHQKY
Ecogal	PNANARALAQ QTETVGLVV GDVSDFFFGA MVKAV....	EQVAYHTGNF
Ecocytr	PQPMGRNVKR NESRTILVIV PDICDPFFSE IIRGI....	EVTAAHNGYL
Ecopurr	PSAVARSILKV NHTKSIGLLA TSSEAAYFAE IIEAV....	EKNCFQKGYT
Eclac	PNRVAQQLAG KQSLLIGVAT SSLALHAPSQ IVAAI....	KSRADQLGAS
Ecorafro	ENTQARRLKT GRTDAIGLVY PENDVPFNSG VFMDMVS CIS	RELAYHDIDL
Ecfrr	ENVAAGLRA GRTRSIGLVI PDLENTSYTR IANYL....	ERQARQRYQ
Ecqln	SPRAKAPFIA LNMAAIKDL IESELEFGHEK GAFTGANTIR	<u>QGRFEQADGG</u>
Ecotyrr	SPRAGKPYLA LNCASIPEDA VESELEFGHAP EG.....	K RGFPEQANGG
Ecofis	.....	.....
Ecoompr	.....	.....MQENY
Ecphob	.....	.....MAR
Ecodye	.....	.....MQTP
Ecokdpe	.....	.....VT
Ecnarl	.....VRKVY EAIAASAVVN QVAPKWSPL FPKKLSQTSK	EIPMSHQEPA
Ecarsr	.....	.....
Ecomalt	HTVVRERLLA KLSGANNFRL ALITSPAGYG <u>KTTLISQWAA</u>	<u>GKNDIGWYSL</u>
Ecolexa	.....	.....
Ecpihc	.....	.....
Ecpup	.....	.....
Ecotrpr	.....	.....
Ectetra	.....	.....
Ecetraaj	.....	.....
Ecdeor	.....	.....

Un dominio de pegado a ATP para NR(I) y TyrR, se ubica entre los residuos: 297 a 305, y que se relaciona funcionalmente con el dominio de 234 a 241 de pegado a ATP. (Pittard, 1991). Se observa otro dominio de pegado a ATP para MalT, que va de 278 a 301. (Richet, 1989), mientras que para DnaA, va de los residuos 291 al 305.

Tres Cisteínas en FKN en posiciones 288, 292 y 295, definen el dominio involucrado en la sensibilidad al oxígeno del medio. (Dispensa, 1981).

La citosina en posición 263 de Ada, sirve como aceptor de grupos metilo provenientes del DNA, para la restauración del DNA y activación de la proteína. Además presenta el HTH entre los residuos 295 a 316.



continuación...

	351		400
Ecoara	VVKNQGREFV	CRPGDILLFP	PGEITHHYGRH PEAREWYHQW VYFRPRAYWH
Ecrhac	LHVLENDERPYR	ITRGDLFYIH	ADDKHSYA...SVNDELVLQN IICYCP....
Ecfur	MGEEI.....	GLATVYRVL	NQFD..DAGI VTRHNFEQCK SVFELT....
Ecoglpr	AGGELSRHG	RSLANGRSIL	SPSSAVDFGI LGISGIDSDG SLELFDYHEV
Ecoicir	GSSFQSRNL	.....LAI	VHPILRNLE ESGETVNVAV LDQSDHEATI
Ecolrpr	NPHYLDASLL	.....V..	.....FVE ITLNRGAPDV FEQNTAVQK
Ecada	NAGFPDSSSY	YRKADETLGM	TAKQFRHGGE NLAVRYALAD CELGRCLVAE
Ecocynr	.RLTDAGEVW	QYASRALQE	<u>LGAGKRAIHD</u> VADLTRGSLR IAVTPTFTSY
Ecoxyr	.LFTQAGMLL	VDQARTVLR	VKVLKEMASQ QGETMSGPLH IGLIPTVGPY
Ecocysba	TQVTPAGQEI	IRIAREVLSK	VDAIKSVAGE HTWPKGSLY IATHTQARY
Ecolivc	T.LTEAGEEL	RVFAQQTLQ	YQQLRHTIDQ QGFSLESELH IFCSVTAAYS
Ecometr	.RFTPQGEIL	LQLANQVLPQ	ISQALQACNE ...PQOTRLR IAIECHSCIQ
Ecfadr	...TTLEVL	QRLARDGWL	IQHCKPTKVN NFWETSGLNI LETLARLDHE
Ecpapak	...IHSRVI	..LAMKDYLV	GGHSRKEVCE KYQMNNGY.F STTLRGLR.L.
Ecoarpr	ALLKEE....	..KFSSQGEI	VAALQEQGFD NINQSKVSRM LTKGAVRTR
Ecodnaa	ALLIDDIQFF	ANKERSQEEF	FHTFNALLEG NQIILTSDR YFKEINQVED
Ecocrp	LIKDEEGKEM	ILSYLNQGDF	IGELGLFEEG QERSAWVRAK TACVAEISY
Econirr	YTITEQGDEQ	ITGFHLAGDL	VG.FDAIGSG HHP.S.FAQL ETSMVQEIPF
Ecobira	.LEQGPAAAI	GSLVIGIVM	AEVLRKLGAD KVRVKWPNLD YLQDRKL..A
Ecometja	..EHGKKEQ	VKKITVSIPL	..KVLKILTDE <u>RTRRQVNNLR</u> <u>HATNSGELLCE</u>
Ecqals5	PGMVLINRVV	PGYAHRCVCL	DNLSGARMAT RMLLNNGHQF IGYLSSSHGI
Ecogal	PGMVLINRIL	PGFENRCIAL	DDRYGAWLAT RHLIQGHTR IGYLCSNHSI
Ecocytr	PPMVMANEFA	PELELPTVHI	DNLTAAFDVAV NYLYEQGHKR IGCIAQPEEM
Ecopurf	PMVVMWGEA	KADFTDAVID	NAFEGGYMAG RYLIERGHRE IGVIPGFLER
Eclac	PALFL..DVS	DQTPINSIIP	SHEDGTRLGV EHLVALGRHQ IALLAGPLSS
Ecoafro	PFIALG.RVP	QQLPCAWFDF	DNHAGTWQAT QKLIALGHKS IALLSENTSH
Ecfzur	PIVALDRALD	REHFTSVVGA	DQDDAEMLAE E.LRKFAET VLYLGALPEL
Ecgln	EQRVQEGKFR	EDLPHRLNVI	RVHLPPLRER REDIPRLARH FLQVAARELG
Ecotyr	VELVQKGMFR	EDLYYRLNVL	TLNLPLLRDC PQOIMPLTEL FVARFVADEQ
Ecofis	.....	.....	.....
Ecoompr	<u>LDLMLPGEDG</u>	<u>LSICRRL</u> ...	.....R SQS..NPMPI IMVTAKGEEV
Ecphob	<u>LDWMLPGGSG</u>	<u>IQFIKHL</u> ...	.....K RESMTRDIPV VHLTARGEVE
Ecodye	<u>MDINLPGKNG</u>	<u>LLLAREL</u> ...	.....EQA...NVAL MFLTGRDNEV
Ecokdpe	<u>LDLGLPDGDG</u>	<u>IEFIRDL</u> ...	.....R QNS...AVPV IVLGARGEES
Ecnarl	<u>LDLNNPQMGNG</u>	<u>LETLDKL</u> ...	.....R EKSL...SGRI VVFSVSNHEE
Ecarsr	.....M	<u>LQLTPLQLFK</u>	NLS.....DETRLG IVDLLREMGE
Ecomalt	ELAEWHSPLY	LVIDDYHLIT	NPVIESMRF FIRHQENLT LVVLSRNLQ
Ecolexa	RSPNAABEHL	KALARKGVIE	IVSGASRGIR LLQEBEEGLP LV.....VAP
Ecpicf	...MLSQLN	LRFHKKLIEA	LKTRAGRENT SVNALAEFL DDGLKTVAPG
Ecpupr	....MGTTT	MGV..KLDDA	TRERIXFAAT RIDRTPHWLI KQAIYSYART
Ecotrpr	.....	.....	.....MA
Ectetra	AALELLNETG	IDGLTRKLA	QKLGIEQFTL <u>YHVVKNKRAL</u> LDALAVETLA
Ecetraaj	.....	.....MTV	VDPARFMYER NHFPSLTDKE FETLVLYCQM
Ecedeor	LNNHSAPVVL	LGQYIVLEPR	SASHYLLSDQ KSRLVEEKRR AAKLAATLVE

En CynR se encuentra un Probable sitio de pegado a ATP, entre los residuos 373 -376. El sitio de separación de Ada, por proteolisis, se localiza entre los aminoácidos Lisina(373) y glutamina(374).

En MetJ, se detecta la estructura hoja  $\beta$ -plegada antiparalela de pegado al DNA entre los residuos 360 -402. Los aminoácidos de reconocimiento al DNA para MetJ son: Glicina(355), lisina(357), Serina(358), arginina (381), y asparagina-serina en posición 395-396 respectivamente; mientras que los aminoácidos de dimerización son: Treonina(365, 382) (Somers y Phillips, 1991). Por otro lado, TetR presenta un HTH en posición 366 a 375.

continuación...

	401	450
Ecoara	EWLMWPSIFA	NTGPFPRPDEA HQPHF.....SDLFG QIINAGQQEG
Ecrhac	ERLKLNLDMQ	GAIPGFNASEA GQPHWRLGSM GMAQARQVIG QLEHSSQHV
Ecfur	.....QQHH	HDHLICLDGC K.....VIEFS DDSIEARQRE
Ecoqlr	RTKRAIENS	RHVMLVVDHS KFGRNAMVNM GSI SMVDVAVY TDAAAA SKRD
Ecoicir	IDVQCTHLM	RMS.....AP IGGKLPMSAS GAGKAFLAQL SEEQVTKLH
Ecolrp	LEEIQECHLV	SGD.....PD YLLKTRVPM SAYRKLLEGE.....TLR
Ecada	SERGICAILL	GDDDATLISE LQQMFPAAEN APADLMFQQH VREVIASLNQ
Ecoxynr	FIGPLMADFY	ARYPSITLQL QEMSQEKIED MLCRDELVDG IAFAPVH.SP
Ecoyir	LLPHIIPMLH	QTFPKLEMYL HEAQTHQLLA QLDSGKLDVCV ILALVKE.SE
Ecocysba	ALPNVIKGF	ERYPRVSLHM HQGSPTQIAD AVSKGNADFA IATEALHLYE
Ecoilvc	HLPFILDRFR	AENPSVEIKL TTQDAADAME KVVTGEADLA IAGKPETLPG
Ecometr	WLTPALENFH	KNWPQVEMDF KSGVTFDPQP ALQQGELDLV MTSIDLPRSG
Ecfadr	SVPQLIDNLL	SVRTHNISTIF IRTAFRQHPD KAQE.....VLATANEVAD
Ecpapak	.LNALAAARLA	PYYTDESSAF DX.....
Ecoargr	NAKHEMVYCL	PAELGVPTTS SPLKNLVLDI DYNDAV'V'V'VH TSPGAAQLIA
Ecodnaa	RLKSRFGWGL	TVAIEPPELE TRVAILMKKA DEND...IR LPGAFAFFIA
Ecoocrp	KKFRQL....	.IQVNPDLIM RLSAQMARRL QVTSEKVGNL AFLDVTGRIA
Econlrr	ETLDDL....	.SGKMPNLRQ QMMRLMSGEI KGDQDMILLL SKKNAEERLA
Ecoblra	GILVEL....	.TGKTDAAQ IV...IGAGI NMAMRRVEES VVNQWITLQ
Ecometja	<u>APLHAF</u> ....	.TGQP.....LPDDA DLRKERSDE. IPEAAKEIMR
Ecgals5	EDDAMRKAGW	MSALKKEQDII PPESWGAGT PDMPPGGEAAM VELLGRNQL
Ecoqal	SDAEDRLQGY	YDALAESGIA ANDRLVTFGE PDESQGEQAM TELLGRGRNF
Ecoeytr	PLCHYRLQGY	VQALRRCGIM VDPQYIARGD FTFEAGSKAN QQLDLDPQP
Ecopurr	<u>NTGAGRLAGF</u>	MKAMEAMIK VPESWIVQGD FEPESGYRAM QQILSQPHRP
Eclac	VSARLRLAGW	KNYLTRNLIQ PIAER...EGD WSAMGGFQQT MQMLNEGIVP
Ecorafro	SYVIARRQGW	LDALHEHGLK .DPLRLVLS PTRRAGYLAV MELMSLPAPP
Ecfrrur	SVSFLREQGF	RTAWKDDPRE V..HFLYAMS YEREAQAQLF EKWLETHPMP
Ecgln	VEAKLHPET	EALTRLAMP GNVRLQENT. CRWLTVMAA...GQEVLIQ
Ecotyrr	VERPKLAADL	HTVLTRYAWP GNVRLQKNAI YRALTLQLDGY ELRPQDILLP
Ecofis	.....	.....MFEQRVNSD
Ecoompr	DRIVGLEIGA	DDYIYKPFNP RELLARIRAV LRRQANELPG APSQEE.AVI
Ecpob	DRVRGLETGA	DDYITKPFSP KELVARIKAV MRR.....I SPMAVE.EVI
Ecodeye	DKILGLEIGA	DDYITKPFNP RELTIRARNL LSRTMNLGTV SEERRSVESY
Ecokdpe	DKIAALDAGA	DDYLSKPFGI GELQARLRVA LRRH.....SATTAPDPLV
Ecnarl	DVVTALKRGA	DGYLLKDMEP EDL...LKAL HQAAAGEMVL SEALTPLVLA
Ecarsr	LCVCDLCMAL	DQSQPKISRH LAMLR.ESGI LLDKQKQKMW HY...RLSP
Ecomalt	LGIANL.RVR	DQLEIGSQQ LAFTHQEANE FFDCLRSSPI EAESSRICD
Ecolexa	VAAAGEPLLAQ	QHIEGHYQVD PSLFKPNADF LL.....RVSG
Ecpifc	DGYFQLIADF	EATVRQLYRH ILLGQTFGTS ALSRDELRFV LVHVREAFLR
Ecpuip	TGKQRYASG	.ATCAAFWR.....
Ecotrpr	QQSPYSAAMA	EQRHQEWLRF VDL.LKHAYQ.NDLHLPLNL MLTPDER...E
Ectetra	RHHYSLPAA	GESWQSFLRN NAMSFRALL RYRDGAKVHL GTRPDEKQYD
Ecetraaj	MNVQMVADYQ	NRKPDVIKH LKSCRQKIGV ESDPFELYFIV INKFNVERV
Ecodeor	PDQTLFFDCG	TTTTWIIIEAI DNEIPFFAVC YSLNTPFLALK EKPHCRAFLC

Una alanina en posición 401 está involucrada en la dimerización de MetJ. (Somers y Phillips, 1991). Dos triptófanos en posición 414 y 382, están involucrados en el reconocimiento del sitio de pegado al DNA. (Hansen, 1987)

El dominio de inducción para el grupo de GalS a FruR se localiza en posición de 401 a 410.

En CRP, los aminoácidos TS en posición 433-434 son necesarios para el pegado a cAMP.

Para el grupo de OmpR a KdpE, se presentan tres aminoácidos conservados: ac. glutámico(407), arginina (426 y 433) necesarios para la dimerización de las proteínas.

Una alanina en posición 401 es necesaria para la dimerización de MetJ.

	451		500
Ecoara	RYSELLAINL	LEQL..LLRR	MEAINESLHP
Ecrhac	PFANEMAEEL	FGQLVHLLNR	HRYTSDSLPP
Ecfur	IAAKHGIRLT	NHSLYLYGHC	ABGDCCRED..
Ecoglpr	AGADGPPYST	GAVLILHGFP	RQKTRQVFA
Ecoicir	RKGLHAYTHA	TLVSPVHLKE	DLAQTRKR..
Ecolrp	LPGVNDTRY	VVMEEVKQSN	RLVIKTRX..
Ecada	RDTRLTLPLD	IRGTAFQOQV	WQALRTIPCG
Ecocynr	ELEATPLLTE	SLALVVAQHH	PLAVHEQVAL
Ecooxyr	AFIEVPLFDE	PMLLAIYEDH	PWANRECVPM
Ecocysba	DLVMLPCYHW	NRAIVVTPDH	FLAGKAITI
Ecoilvc	A.VAFSMLEN	LAVVLIAPAL	PCPVRNQSV
Ecometr	.LHYSMPFDY	EVRVLAPDH	PLAAKTRITP
Ecfadr	HADAFALDY	TIFRGLAFAS	GNPIYGLILN
Ecpapak			
Ecoargr	RLLDLSLGA.	EGILGTIAGD	DTIFTTPANG
Ecodnaa	KRLRSNVREL	EGALNRVIAN	ANF...TGRA
Ecocip	QTLNLLAKQP	DAMTH.PDGM	QIKITRQEIG
Econirr	APIYNLSRRF	AQRGFSREF	RLTMTTRGDIG
Ecobira	EAGINLDRN.	.....TL	AAMLIRELRA
Ecometja	EMGINPETW.	.....EY	X.....
EcgalS	TAVFAYNDM	AAGALTALD	NG.IAIPLHL
Ecogal	TAVACYNDSM	AAGAMGLND	NG.IDVPEI
Ecocytr	TAVFCHSDVM	ALGALSQAKR	QG.LKVPEDL
Ecopurr	TAVFCGGDIM	AMGALCADE	MG.LRVQDV
Eclac	TAMLVANDQM	ALGAMRAITE	SG.LRVGADI
Ecorafro	TAIITDNDLS	GDDAAMALQL	RGRLSGEAV
Ecfur	QALFTTSPAL	LQGVMDVTLR	RDG.KLPSDL
Ecgln	DLPGELFEST	V.....AES	TSQMOPDSWA
Ecotyrr	DYDA...AT	V.....AVG	EDAMEGS...
Ecofia	VLTVSTVNSQ	D.....QVT	QKPLRDSVKQ
Ecoompr	AFGKFKLNLG	TREMF.REDE	EMPLTSGEFA
Ecphob	EMQGLSLDPT	SHRVM.AGEE	PLEMGPTFEK
Ecodye	KFNWELDIN	SRSLIGDGE	QYKLPREFR
Ecokdpe	KFSDVTVOLA	AR.VIHRGEE	EVHLTPIEFR
Ecnarl	SLRANRAT..	.....TERD	VNQLTPREPD
Ecarsr	HIFSWAAQII	EQAWLSQQD.	.DVQVIARKL
Ecomalt	DVSGWATALQ	LTALSARQNT	HSAHKSARRL
Ecolexa	.MSMKDIGIM	EGDLLAVHKT	QDVRRNGQVVV
Ecpifc	GHRNLATLPA	LDTLDDITGN	LLAQVEHDR
Ecpup			
Ecotrpr	ALGTRVRIVE	E...LLRGM	SQRELKNELG
Ectetra	TVETQLRFMT	ENGFSLRDGL	YAISAVSHFT
Ecetraaj	FPFLTSEQIN	ILAAFSFYPK	RSTIARRFDI
Ecdedor	GGEFHASNAI	FKPIDFQQT.	MNFCPDIAFY

Para el caso del grupo de ArgR a BirA, se detectan dos aminoácidos involucrados en el reconocimiento del operador, un primer aminoácido más variable, en posición 487(Y, R, V o Q) y un E conservado en posición 488.

En FNR la presencia de D en posición 478 es necesario para estimular la transcripción.

Se detectan los HTH para CRP y FNR de 476-496; de 451-466 para Fur; 498-524 para NarL; de 472-496 para TrpR y de 474-499 para LexA.

Alanina en posición 481 en TrpR es necesario para el pegado del L- triptofano (inductor).

El dominio involucrado en la dimerización para el grupo de GalS-FruR, se ubica en la posición 485 a 493; mientras que el de inducción se ubica entre los residuos 456-459 y 483-486. Sitio de pegado a la RNApol para CRP se localiza entre los residuos 463 a 470, mientras que para FNR se ubica entre los residuos 474-478.

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA

continuación...

	501	550
Ecoara	.DSNFDIASV AQHVCLSPSR LSHLFRQQLG ISVLSWREDQ RISQAKLLLS	
Ecrhac	LKSPFALDKF CDEASCSESV LRQQFRQQTG MTINQYLQV RVCHAQYLLQ	
Ecfur	.....	
Ecoglr	SGCGKRLCIA ADSDNVQLH TVEIDVFCL CAHERRKRGG AALLPVIAQT	
Ecoicir	IFDEHREPFA AISISGPISR ITDDRVTTEFG AMVIKAAKEV TLAYGGMRX.	
Ecolrp	.....	
Ecada	VASACAANKL AIIIPCHRVV RGDGTLGQYR WGVSRKAQLL RREAENEERX	
Ecocynr	EQIDHYCEKA GLHPQVVEIA NSISAVLELI RRTSLSTLLP AAIATQHDGL	
Ecocyx	DQAMGFCFEA GAEDETHFRA TSLETLRNMV AAGSGITLLP ALAVPPERKR	
Ecocysba	SELDTAFNRA GLTPRIVFTA TDADVITYV RGLGQGVIA SMAVDGVPADP	
Ecolilvc	RIELWFRNK ISNPMIYATV GGHEAMVSMV ALGCGVALLP EVVL...ENSP	
Ecometr	DVWRHFLQPA GVSPLKSV D NTL.LLIQMV AARMGIAALP HWVVESEFQ	
Ecfadr	SLALGFYHKL SALCSEGAND QVYETVRRYG HESGEIWHRM QXNLPGDLAI	
Ecpapak	.....	
Ecoargr	.....	
Ecodnaa	VTIDNIQKTV AEYKIKVAD LLSKRRSRSV ARPRQAMAL AKELTNHSLP	
Ecocrp	NLISAHGKTI VVYGTRX...	
Econlrr	GMLAVKGYI TIENNDALAQ LAGHTRNVAX .....	
Ecobira	DNFINRPVKL IIGDKIEFGI SRGIDKQGL LLEQDGIKPK WMGGEISLRS	
Ecometja	.....	
Ecgals5	VRYPIASM.A KLATELALQG AAGNIDPRAS HCFMPTLVRR HSVATRQNAA	
Ecoqal	VRYPIVTH.A TQAELALAL ADNRLPEIT NVFSPTLVRR HSVSTPSLEA	
Ecocyx	IAQPRVEI.G REAMLLLLDQ MGGQHVSGS RLMDCELIIR GSTRALP...	
Ecopurr	IHQPKDSL.G ETAFNMLLDR IVNKREEPQS IEVHPRLIER RSVADGPFPRD	
Eelac	IKQDFRLL.G QTSVDRLLQL SQGQAV.KGN QLLPVSLVKR KTTLAPNTQT	
Ecorafro	VIQSTRSLVG RQISDMVYQI INGASPELQ ITWTFIFYPG STVHSPSFQ.	
Ecfur	VAQRHRDV.A ERVLEIVLAS LDEPRKPKPG LTRIKRNLIR RQVLSRSX..	
Ecgln	BAQPELERTL LTTALRHQTG HKQBAARLLG WGRNTLTRKL KELGME....	
Ecotyr	EITSRFERSV LTQLYRNYP S TRKLAKR.LG VSHTAIANKL REYGLSQKKN	
Ecofis	LVLAEVEQPL LDMVMQVTRG NQTRALMMG INRGTLRKKL KKYGMNK...	
Ecoompr	LARGREYSAM ERSIDVQISR LRRMVEEDPA HPRYIQTWVG LGYVFPDGS	
Ecphob	HVWGTNVYVE DRTVDVHIRR LRKALEPG.G HDRMVQTVRG TGYRFSTRF.	
Ecodye	KMTGRELKP DRTVDVTIRR IRKHFEPTD TPEIATHIG EGYRFGDLE	
Ecokdpe	QVWGPNAVEH SHYLRIYMGH LRQKLEQDPA RPRHPITETG IGYRFMLX..	
Ecnarl	KMIARRLDIT ESTVKVHVKH MLKMKL... KSRVEAAVW HGERIFX...	
Ecarsr	.....	
Ecomalt	DLATRHFLK SAILRSMNDA LITRVTGEEN GQMRLEEIER QGLFLQRMDD	
Ecolexa	GNKVLLPEN SEFKPIVVDL RQQSFTIEGL AVGVIRNGDW LX.....	
Ecplfc	EFEAFRAALR PVVDQMYAEH LLRPLESDCF GLAEVFDVAVL AEIFTLPRLK	
Ecputp	.....	
Ecotrpr	PVE.LRQWLE EVL..LKSDX .....	
Ectetra	PDENLPPLLR EALQIMSDD GEQAFHLGLE SLIRGFVQL TALLQIVGGD	
Ecetra	LESRLMFFM KITVFLX...	
Ecdeor	VKHWAMSMAQ KHVLVVDHDK FGKVRPARMG DLKRFDIYVS DCCPEDEYVK	

Estructura HTH detectada para: AraC(507-526), TyrR(519-541), Fis(522-541), NarL (498-524) y Gln(NR(I)) se ubica de 523 a 543.

En Ada se presenta un dominio de activación para la reparación del DNA, con una cisteína en posición 516 del alineamiento, que funciona como sitio activo de la proteína, ya que es el aminoácido donde se transfiere el grupo alil proveniente de la cisteína del DNA.

continuación...

	551		600
Ecoara	TTRMPIATVG	RNVGFDDQLY	FSRVFKKCTG ASPSEFFAGC EEKVNDDVAVK
Ecrhac	HSRLDISDIS	TECGFEDSNY	FSVVFTRETG MTPSQWRHLN SQKD.....
Ecfur	.....	.....	.....
Ecoqlpr	VYLIFNHVQS	DASASGKGGH	RVTLAESSRK LPRHVAVGEV AKANQAAALR
Ecoicir	.....	.....	.....
Ecolrp	.....	.....	.....
Ecada	.....	.....	.....
Ecocynr	KAISLAP...	..... PLL	ERTAVLLRR. KNSWQTAAAK AFLHMALDKC
Ecocyrr	DGVVYLPCK	PEPRRTIGLV	YRPGSPRS. RVEQLAEAIR ARMDGHFDKV
Ecocysba	D. LVRVDAHD	IFSHSTKIG	FRRSTFLRSY MYDFIQRFAP HLTRDQVDAE
Ecoilvc	EPVRRVMIL	ERSDEKTPFE	LGVCAQKKRL HEPLIEAFWK ILPNHXK...
Ecometr	GLVVTKTIGE	GLWSR.....	LYAAVRDGEQ RQEVTEAFIR SARHACDHL
Ecfadr	CGRX.....	.....	.....
Ecpapak	.....	.....	.....
Ecoagr	.....	.....	.....
Ecodnaa	EIGDAFGGRD	HTTVLHACRK	IEQLREESHDIKEDFSNLIR TLSSX.....
Ecocrp	.....	.....	.....
Econirr	.....	.....	.....
Ecobira	AEX.....	.....	.....
Ecometja	.....	.....	.....
Ecgals5	AITNSTNQAM	X.....	.....
Ecogal	SHHATSDX..	.....	.....
Ecocytr	.....	.....	.....
Ecopurr	YRX.....	.....	.....
Eclac	ASPRALADSL	MQLARQVSR	ESGQX.....
Ecorafro	.....	.....	.....
Ecfur	.....	.....	.....
Ecgln	.....	.....	.....
Ecotyrr	EEX.....	.....	.....
Ecofis	.....	.....	.....
Ecoompr	KAX.....	.....	.....
Ecpob	.....	.....	.....
Ecodye	DX.....	.....	.....
Ecokdpe	.....	.....	.....
Ecnarl	.....	.....	.....
Ecarsr	.....	.....	.....
Ecomalt	TGEWFCYHPL	FGNLRQRCQ	WELAAELPEI HRAAAESWMA QGFPEAIHH
Ecolexa	.....	.....	.....
Ecpifc	AVFPLMLRGL	DWNTQARTL	AQELRPVISA VTETIEAGTL RLEIRVGGQH
Ecpup	.....	.....	.....
Ecotrpr	.....	.....	.....
Ectetra	KLIIFC...	.....	.....
Ecetraaj	.....	.....	.....
Edeor	YAQTQRIKLM	Y.....	.....

Se localizan dos estructuras HTH: RhaR de 556 a 583 y para AraC de 545 a 564.  
 El dominio para la tetramerización de LacI se ubica en posiciones 556-570.

continuación...

	601	650
Ecoara	LSX.....	
Ecfur	.....	
Ecoglpr	C.....	
Ecada	.....	
Ecocynr	AVVGGNESRX.....	
Ecoxyr	LKQAV.....	
Ecocysba	VALRSNEEIE VMFKDIKLE KX.....	
Ecoilvc	.....	
Ecometr	PFVKSARPT YDAPTVRPGS PARLX.....	
Ecomalt	ALAAODALML RDILLNHAWS LFNHSELSL EESLKALPWD SLEENPQLVL	
Ecolexa	.....	
Ecpifc	PGERPGAWYT TPRLLHLLITG QDFVVPYGWE ALSELLGLFT LYARHPALT	
Ecpntp	.....	
	651	700
Ecarsr	.....	
Ecomalt	LQAWLMQSQH RYGEVNTLLA RAEHEIKDIR EDTMHAEPNA LRAQVAINDG	
Ecolexa	.....	
Ecpifc	HGHQGERVMF SPPGNVTPEG FFGIDGLRIP MPAAEFETLV RELATRCQEG	
Ecpntp	.....	
	701	750
Ecomalt	NPDEAERLAK LALEELPPGW FYSRIVATSV LGEVLHCKGE LTRSLALMQQ	
Ecolexa	.....	
Ecpifc	PLAEALTGLR CLYGDL.....	
	751	800
Ecomalt	TEQHARQHDV WHYALMSLIQ QSEILFAQGF LQTAWETQEK AFQLINEQHL	
	801	850
Ecomalt	EQLPMHEFLV RIRAQLLWAW ARLDEAEASA RSGIEVLSSY QPQQQLQCLA	
	851	900
Ecomalt	MLIQCSLARG DLDNARSQLN RLENLLGNGK YHSDWISNAN KVRVIYQMT	
	901	950
Ecomalt	GDKAAAANWL RHTAKPEFAN NHFLQGQWRN IARAQILLGE FEPAEIVLEE	
	951	1000
Ecomalt	LNENARSLRL MSDLNRNLLL LNQLYWQAGR KSDAQRVLLD ALKLANRTGF	
	1001	1050
Ecomalt	ISHFVIEGEA MAQQLRQLIQ LNTLPELEQH RAQRILREIN QHHRHKFAHF	
	1051	1100
Ecomalt	DENFVERLLN HPEVPELIRT SPLTQREWQV LGLIYSGYSN EQIAGELEVA	
	1101	1139
Ecomalt	ATTIKTHIRN LYQKLGVAHR QDAVQHAQQL LKMMGVGVX	

Dominio de pegado al DNA (HTH) en posición 991-1010.

**APENDICE 2. PROTEINAS REGULADORAS DE PROMOTORES SIGMA70 Y GENES REGULADOS EN PROCARIONTES.**

PROT ORGANISMO(S)	GENES REGULADOS	FUNCION METABOLICA	REFERENCIAS
<b>Ada</b> <i>E.coli</i>		Genes de respuesta adaptativa a agentes alquilantes	Shavell, D. y G. Walker, 1991.
<b>AraC</b> <i>E.coli</i>	Operón araC y araBAD	Utilización de L-arabinosa.	Lee, 1987; Ogden, 1980.
<b>AraC</b> <i>E.coli</i>		Expresión de alcalin-fosfatasa, transcripción de los genes del factor sexual F.	Drury, L. S. y R. S. Buxton, 1985.
<b>ArgR</b> <i>E.coli</i>	Regulón arg. organizados en 9 locis.	Síntesis de arginina.	Charlier, D. 1992; Smith, 1989.
	<i>Bacillus subtilis</i>		
<b>ArsR</b> <i>E.coli</i>	Operon ars	Operón para respuesta a arsénico. Detoxificar.	Swiss Prot
<b>BirA</b> <i>E.coli</i>	Operon bio BFCD	Síntesis de biotina.	Nath, S. K.,1988; Cronan, J. E, 1989
<b>CRP</b> <i>E.coli</i>	Operon lac, gal, mal, ara, cat, deo, tnaA, etc	Asimilación de fuentes de carbono y síntesis de pilis bacterianos.	Otsuka, A y J. Abelson, 1978.
	<i>S. typhimurium</i>		Raibaud, O y M. Schwartz, 1984; Kolb, A. et al. 1993; Göransson, M. et al, 1989.
<b>CynR</b> <i>E.coli</i>	Operón cynT, S, X y CynR.	Detoxificación de cianato(compuesto tóxico celular).	Francoise, 1993.
<b>CysB</b> <i>E.coli</i>	Regulón cysA, CD, JIH	Asimilación de sulfato y síntesis de cisteína.	Raibaud, O y M. Shwartz, 1984;
<b>CytR</b> <i>E.coli</i>	Operón cytRP	Catabolismo de nucleótidos.	Ostrowski, J y N. Kredich,1991.
<b>DeoR</b> <i>E.coli</i>	Operón deoCABD	Catabolismo de deoxiribonucleósidos y ribonucleósidos.	Pedersen, H. et al. 1992.
<b>DnaA</b> <i>E.coli</i>	dnaA, rpoH, oriC	inicio de la replicación del DNA.	Amouyal, M. et al, 1989.
<b>FadR</b> <i>E.coli</i>	Regulón de ac. grasos	Síntesis y degradación de ácidos grasos.	Wang, Q. y J. M. Kanugi, 1987; idem, 1989; Fuller, R. et al, 1984.
<b>Fis</b> <i>E.coli</i>	Siete operones de rRNA	Síntesis de RNA ribosomal	Henry, M y J.E. Cronan, 1992.
<b>FNR</b> <i>E.coli</i>	chiCI, trd, nar, etc.	Transporte anaeróbico de electrones	Ross, W. et al, 1990.
<b>FruR</b> <i>E.coli</i>	Operón fru	Utilización de sustratos gluconeogénicos(fructosa).	Raibaud, O. y M. Schwartz, 1984.
	<i>S. typhimurium</i>	Independiente de CRP-cAMP.	Ramseier, T. M. et al, 1993.
<b>Fur</b> <i>E.coli</i>	Regulón fur	Regulón de 20 genes para la toma de Hierro.	Stajiljkovic, 1994.
<b>GalR</b> <i>E.coli</i>	Operón galK,T,E	Metabolismo de D-galactosa	Adhya, 1991
<b>GalS</b> <i>E.coli</i>	Operón galK,T,E	Metabolismo de D-galactosa	Adhya, 1991
<b>GlpR</b> <i>E.coli</i>	Regulón glpACB, glpTQ, glpD, glpF, glpK, glpR.	Utilización de sn-Glicerol-3-Fosfato	Ye., y T. J. Larson, 1988.
<b>IciR</b> <i>E.coli</i>	Operón aceBAK	Catálisis de acetatos.	Cortay, J-C. et al, 1991.
<b>IlyY</b> <i>E.coli</i>	Regulón Ily	Biosíntesis de Isoleucina, Valina.	
<b>KdpE</b> <i>E.coli</i>	Operón kdp	Toma y utilización de Potasio	Voelner, P. et al, 1993.
<b>LacI</b> <i>E.coli</i>	Operón lacZYA.	Catálisis de lactosa.	Gralla, J. 1992.
<b>LexA</b> <i>E.coli</i>		Genes de respuesta a daños por alquilación al DNA.	Little, J. et al, 1981; ibidem, 1980;

PROT ORGANISMO(S)	GENES REGULADOS	FUNCION METABOLICA	REFERENCIAS
<b>Lrp</b> <i>E.coli</i>	Regulón de por lo menos 40 genes	Genes de respuesta a Leucina.	Ernsting, B. et al, 1993.
<b>Malt</b> <i>E.coli</i> <i>Klebsiella sp.</i>	Regulón mal, con 3 operones malPQ,EFG,malKlamBmalM	Toma y catabolismo de maltodextrinas y maltosa como fuentes de carbono.	Vidal-Ingigliardi,1991; Richet, 1989.
<b>MetJ</b> <i>E.coli</i>	Regulón met	Biosíntesis de metionina	Somers, W. S. y G. Walker, 1991.
<b>MetR</b> <i>E.coli</i>	Regulón met	Biosíntesis de metionina	Somers, W. S. y G. Walker, 1991.
<b>NR(I)</b> <i>E.coli</i> Enterobacterias	Regulón para la toma nitrógeno.	Asimilación de nitrógeno.	
<b>NarL</b> <i>E.coli</i>	Operón narG, H, JI	Inducción para la producción de nitrato reductasa por nitrato en anaerobiosis.	Feng, 1987; Walker, 1992.
<b>OmpR</b> <i>E.coli</i>	Genes ompC y ompF	Genes de repuesta al estress oxidativo.	Maeda, S. et al, 1991; ibidem, 1990; Brissette, R. et al, 1991.
<b>OxyR</b> <i>E.coli</i>		Genes de respuesta a estress osmótico y oxidativo	Storz, G. et al, 1990.
<b>PapB</b> <i>E.coli</i>	Operón pap.	Síntesis de Pílus-adhesina (pílis).	Göransson, M. et al, 1989.
<b>PhoB</b> <i>E.coli</i> <i>Pseudomona sp.</i>	Regulón phoA,E,S, etc.	Síntesis de alcalin-fosfatasa y toma de Fosfatos.	Wanner, B. L., 1991.
<b>PilC</b> <i>E.coli</i>	Operon pilCAB	Infección abortiva del plásmido F en presencia del fagoT7	Kennedy, M. et al, 1990.
<b>PurR</b> <i>E.coli</i>	Regulón pur arreglados en 9 locis.	Síntesis de novo de purinas y síntesis y vía de salvamiento de pirimidinas.	Rolles, 1990.
<b>PutA</b> <i>E.coli</i> <i>S. typhimurium</i>	Operon putA y putP	Oxidación de prolina a ac. glutámico.	Menzel, 1981; Ostrowski, 1991; Matoy,1991.
<b>RafR</b> <i>E.coli</i>	Operón raf	Toma y catálisis de rafinosa.	Aslanidis, C. et al, 1990.
<b>RhaR</b> <i>E.coli</i>	Operón rhaDBA	Catálisis de L-Ramnosa.	
<b>TetR</b> <i>E.coli</i>	operón tet	Genes de resistencia a antibióticos (tetraciclina).	Allard, J. D. et al, 1993.
<b>TrpR</b> <i>E.coli</i>	Operón trp	Síntesis de L-triptofano.	Gunsalus, R. P. y C. Yanofsky, 1980.
<b>TyrR</b> <i>E.coli</i>	Regulón tyr, con 8 unidades transcripcionales.	Síntesis de L-tirosina y aminoácidos aromáticos.	Pittard, 1991; Andrews, 1991.
<b>TraJ</b> <i>E.coli</i>	Operon traY-Z, traM	Control del plásmido F en la conjugación.	

**Apéndice 2. Se muestran los operones y/o genes regulados por las proteínas reguladoras de promotores sigma70, así como sus funciones y referencias.**