



03062 7
24.
UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

CENTRO DE INVESTIGACION SOBRE FIJACION DE
NITROGENO

LA POSICION DEL MOTIVO DE UNION AL DNA
COMO INDICADOR FUNCIONAL Y EVOLUTIVO EN
REGULADORES TRANSCRIPCIONALES

T E S I S
QUE PARA OBTENER EL GRADO DE:
MAESTRO EN INVESTIGACION
BIOMEDICA BASICA
P R E S E N T A :
ERNESTO } PEREZ RUEDA

CUERNAVACA, MOR.

1987

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

***La ciencia es todavía un candil que parpadea
en una gran caverna en tinieblas.
M. Vargas Llosa***

Debo hacer un reconocimiento especial a:

Dr. Julio Collado Vides

Dr. Jaime Mora

Dr. Enrique Morett

Dr. Lorenzo Segovia

Dr. Miguel Angel Cevallos

Dr. David Romero

Dr. Mario Soberón

Por su valiosa cooperación y confianza.

En este lugar he vivido cosas inesperadas y gratas. He conocido a gente muy valerosa a quienes admiro y aprecio. De hecho, la realización de este trabajo es el resultado del apoyo que me han brindado cada uno de los amigos y compañeros que me rodean. Realmente agradezco a todos que me han tendido la mano y han hecho más divertida mi estancia aquí, pero en especial:

A mis amigos de aventuras...

Humberto, Cano, Juan, Josué, Agustino, Adriana, Brenda, Ramón...

A los del laboratorio que me soportan todos los días...

Denis, Jack, Hely, Araceli, Víctor, Alma, Victoria, Arturo y Rosa María.

A mis amigos del CIFN...

Sergio, Carmen, Mere, Paulina, Pedro, Gaby, Mario, Rocto, Elizabeth, Nora, Ismael, Javier, Angeles, Sandra, Sara, A. Corvera, Edgar, Luis Ernesto, Alberto y los que me faltaron.

A Mis padres y Hermanos...

A quienes amo cada día más y de quienes aprendí el valor de la vida.

A los ausentes...

A la Universidad Nacional Autónoma de México y al Centro de Investigación Sobre Fijación de Nitrógeno por permitirme continuar con mi formación académica.

Este trabajo se realizó bajo la asesoría del Dr. Julio Collado Vides, en el Laboratorio de Biología Computacional del Centro de Investigación Sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México.

Durante mis estudios de Maestría fui apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT).

INDICE

RESUMEN	1
INTRODUCCION	2
Elementos de la Transcripción	2
Promotores $\sigma 70$	3
Promotores $\sigma 54$	4
Transcripción en Procariontes	4
Regulación del Inicio de la Transcripción	6
Proteínas Reguladoras de Procariontes	7
Transcripción en Eucariontes	9
Proteínas Reguladoras de Eucariontes	10
ANTECEDENTES	12
OBJETIVO	15
MATERIAL Y METODO	16
RESULTADOS	19
DISCUSION	24
CONCLUSIONES	31
PERSPECTIVAS	33
REFERENCIAS	35
FIGURAS Y TABLAS	41
APENDICE	66

RESUMEN

El mecanismo a nivel de inicio de la transcripción se regula por proteínas que se unen al DNA en regiones cercanas e inclusive sobre la secuencia que reconoce la RNA polimerasa (RNAPol). Estas proteínas reconocen secuencias específicas del DNA a través de estructuras conservadas, de las cuales nosotros hemos analizado cuatro: hélice-vuelta-hélice (HTH) de procariontes; homeodominios, hélice-loop-hélice (HLH) y dedos de zinc de eucariontes. Las secuencias proteicas se obtuvieron a partir de la base de datos Swiss-Prot. El análisis de los reguladores procariontes, dio por resultado una distribución uniforme de grupos funcionales asociados a la posición de su HTH. La mayoría de los represores tienen el HTH en posición amino terminal (NH₂-term), los activadores en posición carboxilo terminal (COOH-term); mientras que el grupo de proteínas duales que es constituido principalmente por miembros de la familia LysR lo presentan en el NH₂-term. La ausencia de una explicación mecanística aparente de dicha relación junto con otras evidencias adicionales, nos hace pensar que esta distribución puede ser explicada en términos evolutivos.

Algunos grupos de proteínas reguladoras eucariontes se analizaron de manera similar a las proteínas de procariontes. La distribución del motivo de unión al DNA se conserva fuertemente: las homeoproteínas presentan el motivo de unión al DNA en posición COOH-term, los dedos de zinc en posición NH₂-term y los HLH en posición central y/o COOH-term.

Para el caso de las proteínas reguladoras de procariontes, la distribución observada, en conjunto con otras evidencias, apoyan la hipótesis de un HTH ancestral al que se le fueron adicionando fragmentos proteicos que determinaron su función.

INTRODUCCION

Los genes se organizan en sistemas que responden a estímulos ambientales externos e internos y podemos clasificarlos con base en su asociación y/o respuesta, en:

Operón: Conjunto de genes estructurales se transcriben en un RNA mensajero (RNAm) policistrónico, a partir -en general- de un mismo promotor y que se regulan por proteínas que se unen a sitios reguladores en el DNA.

Regulón: Conjunto de operones o genes individuales controlados por una sola proteína reguladora (que reconoce una secuencia común a las regiones de los operones o genes) y su ligando efector. Por ejemplo, el regulón *pho* (para la toma de fósforo) (40), está constituido aproximadamente por 20 promotores, más de 14 genes y dos sistemas de transporte. Estos componentes se ubican en diferentes posiciones del cromosoma bacteriano (63). Cuando la concentración de fosfato es baja, la proteína moduladora PhoR se fosforila a sí misma y al regulador PhoB. La fosforilación de PhoB permite su pegado a una secuencia conservada 5'-CTGTGATA(A/T)A(T/A)CTGTCA(C/T)-3', activando la transcripción de los genes *pho*. Cuando la concentración normal de fosfato se restaura, PhoR cataliza la reacción de defosforilación de PhoB-fosforilado y la transcripción se suprime (34).

El control de la actividad de los genes bacterianos asociados en operones o regulones, se ejerce usualmente en el inicio de la transcripción (41), aunque también podemos encontrarla a otros niveles.

ELEMENTOS DE LA TRANSCRIPCION

RNA Polimerasa

La RNA polimerasa (RNAPol) de *Escherichia coli* (EC 2.7.7.6) (32) es la enzima responsable de la síntesis del RNAm, RNAr y RNAt en esta bacteria. La enzima básica de 480 kD es la responsable de la elongación de la cadena de RNA y consiste en un complejo multimérico con 2 subunidades α (35kD cada subunidad), una subunidad β (155kD) y β' (165kD), que son codificadas por los genes *rpoA*, *rpoB* y *rpoC* respectivamente (49). Las subunidades $\beta'\beta\alpha\alpha$ forman una estructura a manera de triángulo, con las 2 subunidades α en la base y las dos subunidades β y β' a los

lados (16). En condiciones normales de crecimiento, el factor $\sigma 70$ se une en el centro de este triángulo y hace contacto con cada una de las otras subunidades formando la holoenzima. Este factor, que es codificado por el gene *rpoD*, puede pegarse al DNA únicamente cuando está en contacto con la enzima básica, requiriéndose para el reconocimiento específico del promotor en la etapa de inicio de la transcripción (16, 49) Figura 1. La holoenzima cataliza la transferencia de un ribonucleósido monofosfato al 3'-OH terminal de la cadena de RNA en crecimiento, usando ribonucleósidos trifosfos como sustrato (61).

En el modelo anterior se describe únicamente al factor $\sigma 70$ como parte integral del complejo de reconocimiento específico del promotor. Sin embargo, la mayoría de las bacterias expresan una gran cantidad de factores sigma adicionales (23), que por medio de la comparación de secuencias se han agrupado en dos grandes familias. La primer familia está relacionada al factor $\sigma 70$ primario de *E. coli* (31); mientras el segundo grupo es similar al factor $\sigma 54$ alternativo de la misma bacteria e involucrado en transcribir genes regulados por nitrógeno (23). Los promotores son reconocidos por holoenzimas conteniendo algún factor σ y presentan secuencias conservadas y particulares para cada factor asociado (Tabla 1). Las holoenzimas que contienen factores del tipo- $\sigma 70$ son competentes para formar complejos abiertos en ausencia de factores auxiliares, mientras los que contienen factores del tipo- $\sigma 54$ requieren proteínas auxiliares para llevar a cabo la formación del complejo abierto (23).

A continuación se describirán los dos principales tipos de promotores descritos en bacterias, aquéllos que son reconocidos por los factores de la familia $\sigma 70$ y los reconocidos por factores de la familia $\sigma 54$.

PROMOTORES Sigma70

Un promotor se define como un segmento de DNA que contiene las señales que dirigen al pegado apropiado de la holoenzima y su subsecuente activación a una forma capaz de iniciar la transcripción. La comparación de secuencias en regiones reconocidas por el factor $\sigma 70$ muestran la presencia de dos hexanucleótidos conservados. El hexamero "TATAAT" está localizado aproximadamente 10 pares de bases (pb) "upstream" o corriente arriba con respecto al inicio de la transcripción (+1), por lo que se le ha denominado caja de Pribnow o -10. La otra caja "TTGACA" se localiza 35 pb "upstream" respecto al +1 (36). Ambas cajas están separadas por una secuencia espaciadora que varía entre 16 y 18 pb de longitud; mientras que una

segunda secuencia espaciadora se presenta entre el -10 y el +1 con una longitud de entre 5-9 pb (30). En algunos promotores de genes ribosomales, se ha descrito un elemento adicional denominado "UP-like element" rico en AT y localizado aproximadamente 20 pb arriba de la región -35 (9), su posible función es formar contactos adicionales con la holoenzima y aumentar la tasa del inicio de la transcripción.

El inicio de la transcripción depende de la capacidad o fuerza del promotor para establecer interacciones estables con la RNAPol. La fuerza del promotor presenta dos componentes, el de equilibrio y que relaciona el grado de ocupación del promotor (complejo cerrado) con una concentración dada de la E σ 70 y el componente cinético, que relaciona la velocidad a la cual se isomeriza el complejo cerrado a complejo abierto (61).

Promotores Sigma54

Los promotores que reconoce el factor σ 54 se describieron inicialmente en genes regulados por nitrógeno y subsecuentemente en algunos genes cuyos productos no son esenciales para todas las condiciones de crecimiento (33). Inclusive, se propone que la transcripción mediada por σ 54 se asemeja al mecanismo descrito para la RNAPol II de eucariontes, ya que este factor contiene dos "zipers" de leucina que son dominios característicos de factores transcripcionales de eucariontes (51).

La holoenzima (E σ 54) reconoce promotores con la secuencia conservada 5'-TGGCAC-NS-TTGCA-3', localizada en posición -12 y -24 con respecto al inicio de la transcripción. Los dinucleótidos GG en posición -24 y GC en posición -12 son invariables (Tabla 1). Aunque la E σ 54 se une al promotor y forma el complejo cerrado, la transición hacia el complejo abierto requiere la presencia de un activador, que se une a sitios localizados 100 a 150 pb "upstream" con respecto al +1 (21, 33). Los activadores transcripcionales en este tipo de promotores, ayudan a la isomerización del complejo cerrado al abierto involucrando la hidrólisis de ATP.

TRANSCRIPCIÓN EN PROCARIONTES

La transcripción es el proceso central en el desarrollo y crecimiento celular, en el cual los genes codificados en el DNA doble cadena del cromosoma, son selectivamente localizados, reconocidos y transcritos por la RNAPol para producir los RNA mensajeros

y los RNA estructurales (61). En *E. coli* la frecuencia del inicio de la transcripción varía en un intervalo de diez mil veces como resultado de los mecanismos regulatorios y de la fuerza inherente de los promotores (41).

El modelo del inicio de la transcripción en promotores de la familia $\sigma 70$ (el factor sigma más abundante en células en fase exponencial) sugiere los siguientes pasos (19). Figura 1:

A. Formación del complejo RNAPol-factor sigma70 (proceso reversible), generando la holoenzima ($E\sigma 70$). El factor $\sigma 70$ confiere especificidad al complejo para el reconocimiento del promotor.

B. Formación del complejo cerrado. La $E\sigma 70$ reconoce secuencias específicas en el DNA (promotor). El factor sigma aumenta la afinidad de la enzima básica por los promotores. El reconocimiento específico del promotor está en función del factor σ , formándose el complejo cerrado, que tiene un potencial definido para isomerizarse espontáneamente a un complejo abierto, caracterizado por la burbuja de transcripción.

C. Complejo abierto. Desnaturalización o apertura del DNA (reacción irreversible) en la zona de contacto entre el DNA y la $E\sigma 70$, aproximadamente entre la posición -10 a +1, permitiendo que la cadena templado se encuentre disponible para la complementación con los ribonucleótidos. La burbuja de transcripción se produce por la desestabilización de los puentes de hidrógeno que mantienen unida la doble hélice produciendo un desenrollamiento local que inicia en el sitio de unión de la $E\sigma 70$. La $E\sigma 70$ protege alrededor de 75 pb (aproximadamente 250 Å) al formar el complejo abierto (61).

D. Al iniciar la transcripción el factor sigma se libera, reciclándose. La RNAPol se mantiene unida al DNA (promotor) mientras se unen los primeros nucleótidos (9 aproximadamente). Inicialmente, se sintetizan cadenas cortas de RNAm "abortivas", pero al sintetizarse una cadena lo suficientemente estable de aproximadamente 8 a 9 nucleótidos, se produce la elongación de la cadena del RNAm. La transcripción se realiza corriente abajo del promotor en dirección $5' \rightarrow 3'$, esto es, la RNAPol se mueve a lo largo de la cadena templado en dirección $3' \rightarrow 5'$.

REGULACION DEL INICIO DE LA TRANSCRIPCION

En *E. coli*, la tasa del inicio de la transcripción en los promotores sigma70 se regula de acuerdo a los requerimientos celulares. La regulación opera a dos niveles: una modulación directa de la especificidad y afinidad de la RNAPol-sigma70 para el reconocimiento de su promotor, y el control por los reguladores que interactúan con los sitios de pegado generalmente cercanos o sobrelapados al sitio de pegado de la Eσ70.

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

Los genes bajo control negativo se expresan, a menos que una proteína represora los inactive (30). Una proteína represora se pega al DNA para impedir que la RNAPol se pegue al promotor e inicie la transcripción. Las proteínas reguladoras represoras se pegan al DNA en un sitio localizado corriente abajo del promotor o bien sobrelapado con este, por lo que la RNAPol no puede reconocer el promotor e iniciar la transcripción. En el operón *lactosa*, los tres genes *lacZYA* se expresan en altos niveles cuando se encuentra presente la alolactosa. En ausencia de alolactosa, el represor LacI se une a un sitio sobrelapando al promotor *lac*, e inhibe el inicio de la transcripción. En presencia de alolactosa, el complejo represor-operador se desestabiliza por el pegado del inductor a LacI, formándose el complejo represor-inductor, que se separa del DNA. Así, la Eσ70 puede reconocer el promotor e iniciar la transcripción (5, 20).

Para genes bajo control positivo, la expresión únicamente es posible cuando una proteína reguladora activadora está presente (30). El pegado de la proteína reguladora está cercano al promotor e incrementa la tasa de transcripción. Los promotores son, en este caso, débiles para la interacción con la Eσ70. La proteína activadora contacta la Eσ70 incrementando su afinidad por el promotor. Por ejemplo, el operón *lactosa* se regula positivamente por CRP (Proteína Receptora de Catabolito). La proteína se une a un sitio adyacente al promotor *lac*, centrado en posición -61.5 y extendiéndose desde -72 hasta -52 respecto al +1. Dos posibles mecanismos se proponen por los cuales CRP podría activar: es posible que CRP contacte con la subunidad alfa de la Eσ70, o bien cambie la estructura del DNA permitiendo que la Eσ70 se pegue al promotor con

mayor afinidad, ayudando al inicio de la transcripción (30).

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). (62). En general, las proteínas activadoras responden a concentraciones de metabolitos, sustratos, cofactores y segundos mensajeros (62). por ejemplo, CRP responde a concentraciones de cAMP intracelular. Los activadores se pegan predominantemente a posiciones entre -80 y -30 con respecto al inicio de la transcripción (21).

Proteínas Represoras

Los represores están sujetos a controles similares al de los reguladores positivos y pueden operar en varios niveles de la función del promotor con respecto a su asociación con la holoenzima; ya sea bloqueando totalmente un promotor (eliminandolo del grupo de promotores que compiten por la $E\sigma 70$) previniendo el pegado de la RNAPol y evitando la formación del complejo cerrado o bien evitando la isomerización del complejo cerrado a complejo abierto (62). La región preferencial de pegado de los represores coincide o sobrelapa con el sitio de pegado de la RNAPol (-35, -10) interfiriendo con su pegado al DNA (9).

PROTEINAS REGULADORAS DE PROCARIONTES

La unidad funcional de la estructura terciaria de las proteínas o dominio, se define como una cadena polipeptídica que puede plegarse independientemente y espontáneamente en una estructura terciaria estable (14). Las proteínas pueden comprender desde un sólo dominio hasta una amplia variedad de los mismos (6).

Las proteínas reguladoras de la transcripción presentan una organización modular con distintos dominios para varias funciones: dimerización, pegado a DNA y unión a inductores, entre otros (28). Estos dominios pueden conservarse entre los diferentes miembros de cada familia de reguladores transcripcionales. En algunos casos, como

el dominio de pegado al DNA, pueden intercambiarse entre proteínas, evidenciando que pueden comportarse como unidades de plegamiento independientes (25). Estos dominios además, 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, algunos mediados directamente por moléculas de agua (represor del operón triptófano), o bien por interacciones de Van der Waals (25).

Una de las estructuras de unión al DNA que es importante para clasificar las proteínas en grupos es la denominada hélice-vuelta-hélice (HTH). El HTH se presenta en reguladores transcripcionales, y también se le ha descrito en factores sigma de las familias del tipo $\sigma 54$ y $\sigma 70$ (31), además de 4 ó 5 proteínas que no se unen a DNA y que funcionan como enzimas (68).

El HTH consta de un segmento de 20 residuos de aminoácidos, dividida en dos α -hélices que forman aproximadamente un ángulo de 120° . La primer α -hélice contacta los grupos fosfato del DNA y la segunda α -hélice o de reconocimiento contacta los nucleótidos expuestos en el surco mayor del DNA, determinando la especificidad de la interacción (66). Figura 2.

La estructura del tipo HTH presente en la mayoría de las proteínas reguladoras de procariontes se ha propuesto como una estructura originada a partir de un ancestro común. La capacidad de pegarse al DNA específicamente es el producto de reemplazamientos de aminoácidos dentro de la estructura (24). Muchas proteínas tienen al menos tres α -hélices, que pueden formar probablemente la estructura globular más pequeña y estabilizar la hélice de reconocimiento: dos α -hélices pueden colocarse en un plano, pero 3 α -hélices pueden crear una verdadera estructura tridimensional, colocando los residuos hidrofóbicos dentro de la estructura y los hidrofílicos fuera (56).

El motivo de unión al DNA o HTH descrito inicialmente en la proteína CRO del fago lambda y posteriormente en CRP de *E. coli*, está ampliamente diversificado en reguladores transcripcionales de procariontes. Inclusive, sólo un segundo motivo estructural diferente al HTH se ha descrito en bacterias como *E. coli*. Esta estructura, esta constituida por 3 α -hélices y una β -plegada que ocupan aproximadamente el 50% de las proteínas MetJ y ArcA. Al asociarse las proteínas en dímeros, las hojas β -

plegadas forman la estructura funcional "β-plegada antiparalela", que se expone en la superficie de la proteína a manera de una protuberancia (6), reconociendo los nucleótidos del surco mayor del DNA (57). Figura 3.

Antes de describir los principales tipos de estructuras descritas en proteínas reguladoras de eucariontes, es necesario presentar un panorama general de como se lleva a cabo el mecanismo de transcripción en estos organismos.

TRANSCRIPCIÓN EN EUKARIOTES

Un gran número de factores de transcripción en eucariontes se han identificado. El mecanismo de transcripción es precedido por el ensamblaje de un complejo multiproteico, que consiste de la RNAPol II y seis factores generales de transcripción (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIIF y TFIIH. El primer paso en el ensamblaje del complejo de preiniciación es el pegado del factor TFIID a una secuencia específica de DNA -la caja TATA- localizada corriente arriba del inicio de la transcripción. TFIID es un factor que consiste de un polipéptido involucrado en el reconocimiento de la caja TATA, denominado TBP o "TATA Binding Protein" (30 kDa), y al menos ocho cadenas polipeptídicas adicionales, denominados TAF's o factores asociados a TBP. TBP puede unirse a la secuencia "TATAAA" del surco menor del DNA, localizada entre el -37 al -25. TFIIA ayuda al direccionamiento de la transcripción. Posteriormente TFIIB se une al complejo TBP-DNA y ayuda al reclutamiento de la RNAPol. Finalmente, TFIIIF incorpora la RNAPol al complejo e inicia la transcripción, mientras TFIIE podría estar involucrado en el posterior reclutamiento de TFIIH y TFIIJ. TFIIH con actividad de quinasa, fosforila al dominio COOH-term de la RNAPol (30). Figura 4.

Una vez que se asocian todos los factores antes mencionados a la RNAPol, se genera el complejo cerrado, que se isomeriza a complejo abierto e inicia la transcripción.

PROTEÍNAS REGULADORAS DE EUKARIOTES

Si bien es cierto que los reguladores procariontes presentan prácticamente un solo tipo de estructura de unión al DNA (HTH), en organismos eucariontes, aproximadamente el 80% de los factores de transcripción presentan uno o más de las tres estructuras de unión a DNA o "motivos estructurales" conocidos: hélice-loop-hélice, dedos de zinc, zipper de leucina y hélice-vuelta-hélice (homeodomínios) (6).

mismos que se describen a continuación:

Homeodominios

Los homeodominios (Figura 3) pertenecen a una gran familia de activadores transcripcionales y se han descrito en un amplio rango de organismos eucariotes, tales como levaduras, insectos y vertebrados (13). Estos motivos son estructuras de pegado al DNA de aproximadamente 60 aa que se pliegan en tres α -hélices (descritos originalmente en genes homeóticos de *Drosophila*). La parte que contiene las hélices 2 y 3 constituyen un HTH típico de procariontes. La hélice 3 o de reconocimiento interacciona con los grupos fosfato del surco mayor del DNA: es más larga que la descrita en procariontes y contiene más residuos básicos en su región COOH-term (26). El NH₂-term del homeodominio forma una estructura a manera de brazo (extendiéndose desde la hélice 1) que se pega en posición adyacente al surco menor. Aunque muchos homeodominios se pegan con alta afinidad al DNA como monómeros, en los sistemas vivos se pegan frecuentemente en dímeros a través de interacciones cooperativas (58).

Dedos de Zinc

Los dedos de zinc (Figura 4) son motivos estructurales que están repetidos en tandem para reconocer secuencias de DNA de diferentes longitudes. Cada dedo es similar en su composición e interactúa con un pequeño número de bases. El modelo clásico de dedos de zinc se describe como dos residuos de Cisteína (Cys) y dos residuos de histidina (His) unidos a un átomo de zinc (6). La secuencia consenso de un sólo dedo consiste en: Cys-X₂-4-Cys-X₃-Phe-X₅-Leu-X₂-His-X₃-His (27), donde X representa cualquier aminoácido.

Hélice-loop-hélice (HLH)

Otro grupo de proteínas que presentan similitudes al grupo de los zipper de leucina son las que presentan el motivo estructural denominado hélice-loop-hélice o HLH (Figura 5). Esta estructura consiste de dos α -hélices separadas por un "loop" y se describió inicialmente en un factor que actúa como "enhancer". La mayoría de las proteínas con este motivo son del tipo b/HLH, es decir, presentan una región básica conservada en el NH₂-term adyacente a la estructura que se requiere para el pegado al DNA. Otras proteínas presentan adicionalmente en el COOH-term un "ziper" de

leucina (Z), por lo que son denominadas b/HLH/Z. Generalmente, las proteínas con HLH se pegan como dímeros a secuencias conservadas o cajas-E (CANNTG) (46). El HLH forma una protuberancia con el arreglo de cuatro-hélices usando dos-hélices de cada monómero. La región básica forma una hélice que se ubica en el surco mayor del DNA (42).

ANTECEDENTES

Las clasificaciones en biología son muy importantes, ya que son las herramientas que nos ayudan a resumir y a 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 se han agrupado en familias cuyos miembros divergen de un ancestro común, tienen plegamientos similares y presentan por lo general secuencias y funciones similares (10). Así, las proteínas se clasifican para reflejar tanto relaciones estructurales como evolutivas. Existen diferentes jerarquías que reflejan las relaciones ya mencionadas, pero las principales son: familia, superfamilia y plegamiento en el espacio o "fold", que a continuación se definen.

Familia: Las proteínas agrupadas en una familia muestran una relación evolutiva clara. La identidad entre las secuencias es del orden de 30% o mayor. Por ejemplo, los miembros de la familia LysR (53) comparten un ancestro común y muestran porcentajes de identidad mayores al 30% y en algunos casos de hasta el 60% (por ejemplo, C1cR y TcbR que regulan el catabolismo de hidrocarburos aromáticos). Además comparten características funcionales muy similares, por ejemplo, regulan promotores divergentes. Sin embargo, también podemos encontrar proteínas con una baja identidad de secuencias, pero cuyas funciones y estructuras proveen evidencias de compartir un origen común.

Superfamilia: Proteínas que tienen una baja identidad de secuencias, pero cuyas características funcionales y estructurales sugieren un origen evolutivo común. Por ejemplo, las proteínas que unen retinol (RBP), una proteína de insectos que une el pigmento "biliverdina", y una β -lactoglobulina abundante en la leche con función desconocida, se agrupan en una superfamilia. Estas proteínas son casi de la misma longitud, pero presentan un pobre porcentaje de identidad de alrededor del 17%. Cuando se hace la comparación de las secuencias y la forma como funcionan, se observa que éstas proteínas pegan grandes moléculas hidrofóbicas, a través de una estructura denominada " β -barrel" que se enrolla de manera muy similar en las tres proteínas antes mencionadas (6), por lo que se propone que comparten un ancestro común.

Plegamiento tridimensional o "Fold": Son proteínas que comparten un plegamiento similar. Diferentes proteínas con el mismo plegamiento a menudo tienen elementos periféricos de estructuras secundarias y regiones con vueltas que difieren en tamaño y conformación. Las proteínas que se agrupan dentro de la categoría de un mismo plegamiento no comparten un origen evolutivo común: las similitudes estructurales podrían deberse solo a propiedades físicas y químicas que favorecen un arreglo y topologías muy similares.

Suzuki y Brenner (56) proponen que las proteínas que se pegan al DNA pueden ser clasificadas en subgrupos que no siempre están relacionados evolutivamente. lo cual sugiere que los plegamientos multi-hélice y muchos DBD (Dominio de pegado al DNA), son el producto de una convergencia evolutiva. Los plegamientos globulares multihélices se han encontrado en DBD's de procariontes y eucariontes.

Familias de proteínas reguladoras (Ver apéndice I)

Como ya se mencionó anteriormente, las proteínas involucradas en la regulación de la expresión génica se clasifican con base en la comparación de secuencias de aminoácidos. En la base de datos Swiss-Prot (3, 4) podemos encontrar al menos 15 familias que agrupan alrededor del 85% del total de las proteínas reguladoras de procariontes. Sin embargo, no todas las familias están igualmente estudiadas y de algunas se conoce poco, como es el caso de la familia ArsR que regula la expresión de los genes para la resistencia al arsénico (2). De los grupos más estudiados encontramos a la familia AraC/XylS (18, 48), LysR (27, 53), CRP (54), GalR/LacI (59, 43, 65), y LuxR (45) y EBP (38) de la familia de los dos componentes (55).

Las funciones reguladas por las proteínas que forman parte de una misma familia son muy homogéneas. Ver Tabla 2. Por ejemplo, la mayoría de los reguladores de la familia GalR/LacI regulan la expresión de los genes para el catabolismo de fuentes de carbono: lo interesante de esta familia, es la proteína transportadora de ribosa o RbsB asociada a membrana no se une al DNA y que forma parte del grupo (65). Para el caso de la familia LysR, sus miembros regulan principalmente genes para la biosíntesis de aminoácidos (27). La familia de los dos componentes agrupa proteínas que responden a estímulos provenientes del medio extracelular, comparten el dominio de pegado al DNA y un dominio que es fosforilado por un proteína quinasa (55). Las proteínas de la familia EBP regulan los genes para la asimilación de nitrógeno o procesos que no son totalmente requeridos para que la sobrevivencia de la célula (36), y finalmente la

familia AraC/XylS tiene reguladores involucrados en el metabolismo de fuentes de carbono y patogénesis (18).

Además de esta homogeneidad funcional se han descrito otras características conservadas entre los miembros de cada familia. Por ejemplo, los más de 50 diferentes activadores y represores de la transcripción de la familia LysR, comparten una región altamente conservada en el extremo NH₂-term que contiene el motivo de unión al DNA o HTH (44). El extremo COOH-term está menos conservado y posiblemente está involucrado en el pegado a los metabolitos (27). La gran conservación de la secuencia donde se ubica el HTH (por arriba del 30% de identidad) sugiere que este motivo estructural presenta un origen común para los miembros de la familia (53). En el caso de los miembros agrupados en la familia AraC/XylS, se presenta una región altamente conservada en el extremo COOH-term y que corresponde al HTH. El extremo NH₂-term está menos conservado y posiblemente está involucrado en la regulación alostérica y la dimerización de estas proteínas (48). Finalmente, se ha descrito en la familia GalR/LacI, que sus miembros comparten al menos un 60% de similitud y más de 30% de identidad. La región más conservada se localiza en extremo NH₂-term y corresponde al dominio de pegado al DNA (con el HTH) (65).

También encontramos familias cuyos miembros presentan un tipo de regulación predominante, como es el caso de la familia GntR que agrupa aproximadamente 25 reguladores, descritos a partir del represor del operón gluconato en *Bacillus subtilis* (26). Se ha observado que la mayoría de estas proteínas son represores de la transcripción, como por ejemplo el regulador HutC (para la utilización de histidina) de *Pseudomonas putida* y *Klebsiella aerogenes*, y el represor FadR para la utilización de ácidos grasos. También se han descrito familias como AraC/XylS donde predominan activadores de la transcripción (48) y grupos donde predominan proteínas con actividad dual, como en la familia LysR.

En general, las familias donde se agrupan los reguladores transcripcionales presentan características muy particulares en cuanto a secuencia y a funciones se refiere.

OBJETIVO

Describir la posición del motivo de unión al DNA con relación a la actividad regulatoria de represor, activador o dual en un conjunto de proteínas reguladoras de la transcripción de procariontes y eucariontes. Los motivos estructurales de unión al DNA a analizar son los del tipo hélice-vuelta-hélice (HTH) de procariontes; los homeodominios, los dedos de zinc y los hélice-loop-hélice (HLH) de eucariontes.

MATERIAL Y METODO

Selección de registros

Un total de 403 proteínas que se pegan al DNA en procariontes y 458 de eucariontes se obtuvieron de la base de datos Swiss-Prot versión 31.0 (3, 4). Los registros se evaluaron con base en los siguientes criterios:

- 1) Todas las secuencias parciales se rechazaron.
- 2) Secuencias con el motivo de unión al DNA no determinado se eliminaron también.
- 3) No se incluyeron secuencias sin función determinada.

Las proteínas procariontes a analizar se separaron en los siguientes grupos:

- 1) Proteínas reguladoras de promotores de la familia $\sigma 70$ (227 proteínas, que a su vez se subclasificaron en familias que se describen en Swiss-Prot) y de las cuales 86 son activadores, 73 represores y 62 duales;
- 2) 50 proteínas que regulan promotores del tipo $\sigma 54$. La mayoría de estas proteínas son activadoras de la transcripción;
- 3) 69 factores sigma. De los cuales 53 pertenecen a la familia $\sigma 70$ y 16 a la familia de factores $\sigma 54$.

En el caso de eucariontes, las proteínas se subdividieron en:

- 1) 95 proteínas con el motivo hélice-loop-hélice, sin descripción de su actividad regulatoria;
- 2) 223 homeoproteínas. En 198 no se describe función reguladora, 23 son activadoras y 2 son represores; y,
- 3) 140 proteínas con dedos de zinc repetidos en tandem de dos dedos.

GRAFICACION

Posición relativa del motivo de unión al DNA en la proteína

Para las proteínas con HTH de procariontes, el centro de cada motivo se identificó y se localizó con respecto a una longitud normalizada para cada proteína. El número de proteínas que tienen el motivo ubicado dentro de un intervalo definido cada 10%, se contaron y se graficaron. El 0% representa el extremo NH₂-term y el 100% el COOH-term proteico.

Los reguladores transcripcionales de eucariontes se analizaron de igual forma que las proteínas procariontes (previamente descrito). La localización y ubicación del motivo de unión al DNA dentro de rangos normalizados se realizó de forma independiente para cada una de las siguientes estructuras: homeodominios, dedos de zinc y hélice-loop-hélice.

Posición absoluta del motivo estructural

La posición central del HTH se obtuvo para cada grupo de proteínas divididas de acuerdo a su actividad (represores y activadores). El número de aminoácidos desde el centro del HTH al extremo proteico (NH₂-term o COOH-term) se calculó y se graficó con un intervalo de 5 aminoácidos.

Un análisis similar se realizó para los factores sigma de procariontes y para los reguladores transcripcionales de eucariontes, estos últimos de acuerdo a cada tipo de motivo estructural que presentan.

REVISION BIBLIOGRAFICA

Se hizo una revisión bibliográfica para determinar si una proteína presentaba actividad adicional a la descrita en Swiss Prot, además de realizar, una descripción funcional de cada grupo o familia, tomando en cuenta las siguientes características: actividad reguladora (represor, activador o dual) y autoregulación, familia (o grupo) a la cual pertenece, posición y tipo de estructura de pegado al DNA; entre otras propiedades.

ANALISIS DE SECUENCIAS

Pileup

Para la comparación de secuencias, utilizamos el programa Pileup (Genetics Computer Group, versión 8.1, Universidad de Wisconsin, Madison), que realiza un alineamiento múltiple a partir de un grupo de secuencias relacionadas, utilizando una simplificación del método de Feng y Doolittle (17). El procedimiento de alineamiento múltiple inicia con la comparación de las dos secuencias más similares generando un grupo o "cluster" de dos secuencias alineadas. Este grupo puede entonces compararse con la siguiente secuencia más similar o bien a otro grupo de secuencias alineadas. El alineamiento final se realiza por una serie de comparaciones progresivas

entre pares, que incluye un decremento en similaridad entre las secuencias y clusters, hasta que todas las secuencias estén incluidas en el alineamiento final. Antes del alineamiento, las secuencias se agrupan primero por similaridad para producir un dendograma o representación gráfica del grado de similaridad entre las distintas secuencias alineadas. La distancia del eje vertical es proporcional a la diferencia entre secuencias; mientras que la distancia del eje horizontal no tiene significado alguno.

Plotsimilarity

Este programa calcula el porcentaje de similaridad (o identidad) entre todos los miembros de un grupo de secuencias alineadas en cada posición en el alineamiento, utilizando una ventana de desplazamiento para realizar la comparación. El resultado es una gráfica en la cual se observa el porcentaje de similitud o identidad entre las secuencias.

RESULTADOS

Los reguladores de la transcripción en procariontes se dividen con base en la posición del HTH y la actividad que desempeñan en: REPRESORES-NH₂ term, ACTIVADORES-COOH term y DUALES-FAMILIA LYSR-NH₂ term.

Se analizaron 227 proteínas reguladoras de promotores sigma70 con base en la posición de la estructura de pegado al DNA (HTH) y su papel como regulador transcripcional (86 proteínas activadoras, 73 represores y 62 duales). Adicionalmente, 50 reguladores asociados a promotores σ 54 y 69 factores sigma se analizaron de forma similar a los reguladores de promotores σ 70. De este conjunto de proteínas que se unen a DNA un total de 77 se han descrito en *E. coli*, mientras que el resto, 150, se describen en bacterias como *Klebsiella pneumoniae* y *S. typhimurium*, entre otras.

Al realizar el análisis de distribución del HTH en proteínas asociadas a promotores de la familia σ 70 (Figura 8), se observa que esta estructura se localiza predominantemente hacia los extremos proteicos. En el primer 20% del extremo NH₂-term se presenta una población de aproximadamente 114 proteínas, mientras que en el último 30% COOH-term se localiza una segunda población de 71 reguladores. Al revisar la actividad reguladora de cada grupo, encontramos que las proteínas que presentan el HTH en el NH₂-term son predominantemente represoras, las proteínas con el HTH en el COOH-term son activadoras y hay un tercer grupo de proteínas con actividad dual que presentan el HTH en el NH₂-term. El motivo estructural también se presenta en la región central, aunque la predominancia es mucho menor en comparación a los dos picos de los extremos observados en la Figura.

La distribución de la figura 8, es interesante porque describe la relación entre la posición del HTH con respecto a la actividad de regulación asociada a cada proteína. Es decir, encontramos tres grupos principales de proteínas, con base a la posición del HTH: represores-NH₂-term, Activadores-COOH-term y Duales -NH₂-term. Adicionalmente, la localización del HTH en este conjunto de proteínas nos describe que este motivo estructural se restringe casi en su totalidad a los extremos proteicos y que hay restricciones para ubicarse posición central de la proteína.

La distribución observada en la figura 8 es el resultado de la localización del HTH en el total de las proteínas reguladoras asociadas a promotores σ 70, sin embargo.

esta distribución podría estar sesgada por la sobrerepresentación de una proteína, por lo que nos preguntamos si la distribución cambiaría al excluir las proteínas ortólogas (proteínas homólogas de diferentes organismos que son el resultado de un evento de duplicación génica antes de la especiación) del conjunto total de proteínas y realizamos de nuevo el análisis de localización del HTH. Lo relevante de este segundo análisis es que la distribución del HTH asociado a la actividad reguladora no se modifica (Figura 9). Esta evidencia nos muestra que la distribución originalmente obtenida no es un artefacto producto de la sobrerepresentación de un tipo particular de proteínas. Es decir, se sigue conservando la relación NH₂-term represores y duales; y activadores COOH-term.

Los grupos obtenidos a partir de la distribución del HTH y la asociación a la actividad regulatoria, nos mostró que hay regiones preferenciales de localización del motivo de unión del DNA. Así mismo, hay grupos en donde el HTH se localiza en un intervalo más estrecho y grupos donde la distribución del HTH es más amplia. Esto sugiere que los picos pueden estar formados por grupos homogéneos de proteínas o bien por varios grupos diferentes de proteínas. Para verificar esta observación analizamos el interior de cada pico. En el caso de las proteínas duales NH₂-term, el HTH se localiza en el primer 10%, sugiriendo homogeneidad funcional entre las proteínas que componen este grupo. Al verificar los miembros que componen este grupo, encontramos que en su totalidad están clasificadas en la familia de reguladores LysR (Figura 10). Este grupo es funcionalmente independiente al de los represores y activadores, aunque se comporta como los represores-NH₂ en cuanto a la ubicación del HTH. Lo interesante de esta familia de proteínas reguladoras, es que predominan los activadores transcripcionales con autoregulación negativa (lo que les da la propiedad de reguladores duales).

Continuando con el análisis del segundo pico localizado en posición NH₂-term, encontramos otro grupo funcional: Los represores transcripcionales (Figura 11a). Lo interesante de este pico, es que está formado por miembros de varias familias, entre las que tenemos: GalR/LacI, DeoR, AsnC y GntR, entre otras. Finalmente, en el grupo de activadores con el HTH en el COOH-term predominan miembros de las familias CRP, LuxR y AraC/XylS, entre otras (Figura 11b).

Para determinar con mayor precisión como estaban conformados los picos obtenidos en la distribución de las Figuras 8 y 9, calculamos la posición absoluta del HTH para cada una de las proteínas agrupadas de acuerdo a la actividad que

desempeñan. Es decir, para todas las proteínas represoras y activadoras (por separado) se calculó y graficó la posición absoluta del motivo de unión al DNA (Figuras 11a y b). De acuerdo a este análisis, cada pico está integrado por varios grupos o familias. Lo interesante, es que la posición absoluta del HTH es particular y diferente para cada familia evolutivamente relacionada. Por ejemplo, el grupo de las proteínas represoras que está conformado por miembros de la familia GalR/LacI (entre otras familias) presentan el HTH centrado aproximadamente en el aminoácido 15 del extremo NH₂-term, mientras que los miembros de la familia DeoR lo presentan centrado aproximadamente en el aminoácido 25 (Figura 11a). En el caso de las proteínas activadoras, la localización del HTH es también particular para cada familia; aunque los miembros de la familia AraC/XylS no cumplen con esta particularidad, ya que se presentan miembros con el HTH en diferentes posiciones. La familia LuxR tiene centrado su HTH aproximadamente en el aminoácido 80, tomando como número 1 el último aminoácido del extremo COOH-term (Figura 11b).

El siguiente grupo de reguladores transcripcionales que analizamos fue el que está asociado a promotores del tipo σ 54, denominados EBP o "Proteínas que actúan como enhancers". En este análisis, pensamos en la posibilidad de que el HTH estaría localizado en una posición conservada, como se encontró en los reguladores asociados a promotores σ 70. La figura 12a, muestra la distribución relativa del HTH en un total de 50 activadores transcripcionales, predominando la ubicación de este motivo estructural en el último 10% del extremo COOH-term. Al describir la posición absoluta del HTH, observamos que se presenta en los últimos 35 residuos de aminoácidos COOH-term de la mayoría de estas proteínas (Figura 12b).

Finalmente, realizamos un análisis similar en un total de 69 factores sigma, de los cuales 53 proteínas pertenecen a la familia σ 70 y 16 a la familia σ 54 (Ver apéndice II). En este caso, las proteínas presentan el HTH en posiciones claramente diferenciables: los miembros de la familia σ 70 tienen localizado el HTH en el último 20% del extremo COOH-term, mientras que las proteínas de la familia σ 54 lo presentan entre el 70-80% respecto al COOH-term (Figuras 13a). La posición absoluta nos diferencia con mayor claridad los dos grupos: el HTH de la familia σ 70 se localiza en los últimos 60 residuos de aminoácidos del extremo COOH-term, mientras que en la familia σ 54 se presenta en un intervalo mucho mayor, aproximadamente 100 residuos de aminoácidos del extremo COOH-term (Figura 13b).

Las observaciones anteriores sugieren que el HTH se localiza en regiones

preferenciales de las proteínas y que está asociado a la función que desempeñan cuando regulan la expresión genética. Además que la conservación posicional se mantiene y particulariza para cada familia de proteínas. Por lo que esta descripción de la posición del HTH vs función del regulador (repressor, activador o dual), la describiremos -a continuación- con base en la agrupación de las proteínas en familias.

Aproximadamente siete grandes familias de reguladores agrupan al 60% del total de proteínas reguladoras con HTH en procariontes; mientras que el resto podemos considerarlas como familias nuevas o bien familias con pocos miembros descritos (Tabla 2).

Al revisar la longitud de cada uno de los miembros que integran cada familia de reguladores proteicos, encontramos que pueden tener longitudes variables o conservadas (Tabla 2). La misma tabla muestra que la posición del HTH al interior de cada grupo está conservada. Hay grupos o familias de proteínas en las que el 100% de sus miembros tienen ubicado el HTH en el primer 20% NH₂-term (familia GalR) o bien en el último 30% COOH-term. (familia CRP). También encontramos que la única familia en que sus miembros presentan el HTH en posición central es la denominada "ArsR". Finalmente, dentro de cada familia proteica existe una proteína (funcional) de menos de 100-150 aa y donde se describe un HTH. Esto sugiere que la unidad mínima funcional, para el caso de estas proteínas consta de dominios con un tamaño de entre 100-150 aa.

Las características descritas en la tabla 2 muestran que las familias son conservadas en cuanto a función y posición del HTH. Aunado a estas observaciones, las funciones que son reguladas por los miembros de estas familias también se conservan cuando analizamos los genes regulados para cada grupo; por ejemplo, la familia GalR/LacI regula genes para la toma de fuentes de carbono, mientras que la familia CRP está involucrada en respuestas globales.

Después de mostrar una perspectiva de la distribución del HTH en las proteínas reguladoras de procariontes, nos preguntamos si encontraríamos una relación o distribución similar en proteínas reguladoras de eucariontes, por lo que decidimos analizar tres grupos de proteínas que se pegan al DNA: Homeodominios, Dedos de zinc y Hélice-loop-hélice (Ver apéndice III)

HOMEODOMINIOS

En 223 homeoproteínas se identificó y localizó el motivo de unión al DNA u homeodominios (variante del HTH clásico descrito en reguladores procariontes). La figura 14, muestra la distribución del homeodominio. En más del 50% de este conjunto total de proteínas el motivo estructural se ubica predominantemente hacia el último 30% del extremo COOH-term. Si a esto le agregamos que la mayoría de las proteínas en eucariontes son activadores, podemos decir que estas proteínas se comportan como las proteínas activadoras bacterianas con el HTH en el extremo COOH-term.

DEDOS DE ZINC

La posición de los dedos de zinc se obtuvo para aquellas proteínas que presentan esta estructura en repeticiones de dos dedos. Se observa que el motivo de unión al DNA se ubica preferencialmente en el extremo NH₂-term. Es probable que la actividad regulatoria de la mayoría de estas proteínas sea la de activar el inicio de la transcripción. Figura 15.

HELICE-LOOP-HELICE

En este tipo de proteínas encontramos dos poblaciones con el HLH en diferentes regiones: la primera población de proteínas lo presenta hacia la región central y la segunda población lo presenta en el último 20% COOH-term. Sin embargo, cuando vemos la distribución de grupos al interior de cada pico no encontramos proteínas relacionadas en cuanto a secuencia se refiere. Figura 16.

DISCUSION

La regulación de la expresión genética es uno de los temas centrales para tratar de entender los mecanismos de diferenciación y/o expresión celular. Para que una célula se diferencie o responda a una función definida es necesario que se presenten diferentes condiciones ambientales, por ejemplo, concentración de nutrientes, hormonas, carbohidratos, ácidos grasos, entre otros. La presencia de estos componentes en el medio extra e intracelular está íntimamente ligado a la presencia de elementos proteicos que "gobiernan" la expresión de genes-productos para la utilización de los compuestos disponibles.

Estos factores proteicos se han descrito en una serie de bacterias, y van desde represores del inicio de la transcripción (como el represor del operon *lac*) hasta proteínas activadoras (tales como MalT), pasando por proteínas con actividad dual (CRP).

Una característica generalizada en este tipo de proteínas reguladoras procariontes, es que presentan una estructura del pegado al DNA del tipo HTH (hélice-vuelta-hélice), también descrita en eucariontes (homeodomínios). Otras proteínas como MetJ de *Salmonella thymurium* y *Escherichia coli*, y ArcA de *E. coli* presentan otra estructura denominada β -plegada antiparalela.

En este contexto, el análisis realizado para detectar la posición del HTH en el conjunto de proteínas reguladoras de procariontes, dió por resultado el descubrir una relación entre la posición del HTH y la actividad regulatoria. Es decir, hay una tendencia a que los represores tengan el dominio de pegado al DNA en la región NH₂-term, y que los activadores lo presenten en el COOH-term. Los reguladores duales se comportan (en cuanto a posición) similar a los represores NH₂-term. También encontramos un pico de proteínas duales en el extremo COOH-term pero en menor proporción que los duales NH₂-term (figura 8). La descripción de los grupos al interior de cada pico obtenido es necesaria para tratar de encontrar una explicación de esta relación funcional encontrada. Así, al analizar la posición absoluta del HTH en los reguladores se obtienen grupos con el motivo de unión al DNA en una posición particular y conservada, que corresponden a cada una de las familias de reguladores proteicos descritos por análisis evolutivos (figuras 11a y 11b) y que M. Riley (50) obtiene al comparar fragmentos de 100 aa.

La comparación de la posición del HTH entre las diferentes proteínas y la actividad regulatoria desempeñada a nivel familia o grupo evolutivamente relacionados, nos describe a estos grupos como unidades funcionales homogéneas, es decir, la relación existente entre la posición del HTH con respecto a la actividad regulatoria es característica para cada familia de reguladores.

Pero ¿cómo explicar la relación del HTH vs actividad regulatoria? Una hipótesis probable, es suponer que cada familia de proteínas ha evolucionado a partir de un ancestro que se fusionó por recombinación con una estructura pequeña y monomérica que contenía el HTH: esta(s) proteína(s) a la cual se fusionó el HTH y que probablemente originó cada familia es probable que haya tenido una función previa, por ejemplo, proteínas con función enzimática o transportadoras (figura 17). Hay ejemplos de proteínas reguladoras con funciones adicionales como enzimas (PutA, BirA, Ada) o bien proteínas no reguladoras, que no presentan la estructura de unión al DNA y que son homólogas a reguladores transcripcionales (por ejemplo en la familia GalR-LacI, encontramos al transportador de ribosa: RbsB de *S. typhimurium* y *E. coli*). Esta hipótesis explicaría la ubicación del HTH en las proteínas, y su relación con la actividad regulatoria (repressor, activador o dual), ya que cada familia adquirió una actividad determinada a partir de la función inicial que definió cada grupo. Es decir, un ancestro para la familia GalR/LacI, probablemente tenía las características de regular negativamente los genes involucrados en el catabolismo de azúcares; mientras que para el caso de la familia LysR (con actividad dual), la familia diseñó el mecanismo de autoregulación negativa para impedir su propia transcripción, pero para activar la de genes involucrados en la síntesis de aminoácidos. Ahora bien, ¿cómo explicar la posición del HTH en posición central? La hipótesis de ancestros individuales para cada familia podría explicar la relación obtenida para las proteínas que presentan el HTH en posición central y que pertenecen a una sola familia (ArsR) de represores.

Evidencias a favor de esta hipótesis

Las dos zonas donde predominantemente se presenta el HTH en la proteína son fragmentos pequeños, localizados hacia los extremos proteicos (primer 20% NH₂-term o último 30% COOH-term). Lo interesante es que para cada familia de proteínas hemos encontrado al menos un miembro funcional de 100 aa de longitud (tabla 2), que se sale considerablemente del tamaño del resto de las proteínas de cada familia. Debemos agregar que el análisis de la posición absoluta del HTH muestra que la estructura de pegado al DNA se localiza en un fragmento de aproximadamente 100 aa

(longitud menor encontrada para una proteína funcional en cada familia de reguladores). Así, en cada grupo o familia proteica, existe una región preferencial donde se ubica el HTH que no rebasa un fragmento de 100 aa, por lo que podemos decir que el DBD (Dominio de pegado al DNA) ha conservado su tamaño y que preferentemente no se permiten deleciones o amplificaciones, solo sustituciones que no alteran el tamaño preferencial del dominio. Además, cuando las familias reguladoras se comparan en términos de su estructura primaria, la región más conservada se localiza en el dominio de pegado al DNA (DBD) y que presenta el HTH (figura 18). Cuando observamos el activador del regulón maltosa, MalT, perteneciente a la familia LuxR el dominio que lleva a cabo el pegado específico al DNA se localiza en los últimos 95 residuos de aminoácidos del extremo COOH-term (60), y que es donde se presenta la homología más significativa con respecto a los miembros del resto de la familia de reguladores.

En general, la mayoría de las familias proteicas presentan el mayor porcentaje de identidad en la región donde se localiza el dominio de pegado al DNA (DBD, que contiene el HTH) como ya se mencionó anteriormente. En el caso de la familia CRP/FNR, todas las proteínas son homólogas a FNR, y análogas con respecto al DBD, aunque no necesariamente con respecto a las señales que inducen la respuesta de las proteínas. Además, los sitios reconocidos en el DNA por los miembros de la familia también están conservados, sugiriendo que probablemente hay una conservación en el mecanismo de reconocimiento entre DNA-proteína, con diferencias sutiles para cada miembro (54).

Como se ha venido sugiriendo, es muy probable que los eventos de recombinación de un HTH ancestral y una proteína con actividad previa se hayan llevado a cabo antes del mecanismo de la especiación. En el caso de la familia LysR, las grandes distancias genéticas y el contenido de G+C entre las bacterias donde se presentan miembros de esta familia sugiere que el ancestro se originó en etapas muy tempranas de la evolución de los procariontes. La diversidad estructural de los diferentes coinductores que estimulan a los miembros de la familia para activar la transcripción sugiere también una divergencia muy grande en las señales de reconocimiento. Sin embargo, los plásmidos auto-transmitibles pudieron promover de forma más rápida su diseminación y evolución. El tamaño conservado, organización y diversidad de la familia sugiere que son reguladores muy antiguos, pero muy eficientes (53). Los eventos de duplicación génica posteriores a la fusión con la proteína ancestral pueden documentarse, como por ejemplo: NahR de *Pseudomonas putida* y NodD de

Rhizobium, con un 35% de identidad (y mayor conservación en la región del HTH) son transcritos divergentemente con relación a los genes que regulan y presentan sitios de pegado al DNA muy similares, inclusive se propone que ambas proteínas pudieron haber evolucionado a partir de un ancestro común (52). Este es uno de los caminos que pudieron haber seguido no sólo este grupo de proteínas, sino también otras familias proteicas descritas en los procariontes. Para la familia GalR/LacI, es probable que la mayoría de sus miembros se hayan originado por eventos de duplicación, que ocurrieron en un tiempo específico de la historia evolutiva, y después estos eventos fueron más escasos y raros, figura 19 (43). En la familia EBP el dominio central y el COOH-term (donde se localiza el HTH) tienen un origen común para toda la familia, mientras que el dominio NH₂-term parece ser un dominio común sólo para un subgrupo de los activadores de la familia de los dos componentes; además, no hay un "shuffling" extensivo de los dominios funcionales, por lo que probablemente son el producto de coevolución (38).

Por otra parte, Harrison (25) propone que el HTH es un motivo estructural originado a partir de un ancestro único y que la capacidad de pegarse a diferentes operadores específicos en el DNA se adquirió por reemplazamientos de aminoácidos dentro de la estructura (donde predominan aminoácidos polares e hidrofóbicos en la hélice de reconocimiento: Gln, Ser, Thr, Tyr, Asn; Ile, Val, Ala).

Esta estructura es exclusiva de proteínas que se pegan a DNA (reguladores transcripcionales o factores sigma), a excepción de 4 ó 5 proteínas que no se unen a DNA y que funcionan como enzimas (citocromo c peroxidasa, proteína ribosomal L7/L12) (56). La hipótesis de un HTH ancestral, con capacidad de pegado al DNA que se fusionó a otra proteína para dar origen a cada familia de reguladores, es apoyado por el trabajo de Pao, et al. (45), quienes describen que los módulos receptores (del sistema de transducción de señales en bacterias y que consisten en dominios unidos covalentemente a dominios efectores) divergieron de dominios ancestrales comunes. Mientras que el dominio efector correspondiente asociado pudo ser consecuencia del "shuffling" proteico.

La evolución de las proteínas que se pegan a DNA (reguladoras transcripcionales) por la adición de fragmentos a un HTH ancestral, hace suponer que los dominios de pegado al DNA y de otras funciones se encuentran separados, como puede verse en la figura 20. Por otra parte, hay evidencias experimentales que muestran que la función de pegado al DNA se encuentra en una unidad funcional separada a otras

funciones, por ejemplo para el caso del regulador global Lrp de *E. coli* (HTH NH₂-term), que incrementa la expresión de algunos operones y decrece la expresión de otros. Se han obtenido mutantes con cambios en los aminoácidos asociados a la respuesta a leucina, activación de la transcripción y pegado al DNA, y que se encuentran en sitios diferentes o dominios separados (47). En el caso de TyrR, los cambios en la región NH₂-term impiden la capacidad de activar pero no de pegarse al DNA, sugiriendo que ambas actividades se encuentran en dominios separados de la proteína (12). Mientras que en el represor del operón lac se han realizado análisis de los productos obtenidos por proteólisis y estudios de mutantes que no pueden pegarse al DNA, observándose que la proteína se divide en dos dominios, el NH₂-term con 59 residuos de aminoácidos ("headpiece") y que se pega al operador de forma muy similar al represor intacto y un dominio COOH-term muy largo, responsable de la tetramerización y la respuesta al inductor (65). En la familia LysR análisis mutacionales y de similitud de secuencias identifican tres regiones: (a) un dominio de pegado al DNA, con el HTH (residuos 1-65), (b) dominios involucrados en el reconocimiento y/o respuesta al coinductor (residuos 100-173 y 196-206), (c) dominio requerido tanto para el pegado al DNA como para la respuesta al coinductor (residuos 227-253) (53).

Un ejemplo entre otros que se han descrito en la literatura que apunta en la misma dirección es la de la químera realizada entre las dominios de las proteínas LexA-Gal4. El híbrido consta de la región de pegado al DNA de LexA, mientras la región de activación es de Gal4. La proteína resultante es activadora de la transcripción solo si el operador de LexA esta presente cerca del +1. Con este experimento se evidencia que la función de pegado al DNA por Gal4 puede ser reemplazada por la región de pegado del represor LexA sin pérdida de la función de activación (7)

Evidencias de proteínas con función adicional

Las proteínas reguladoras pueden presentar funciones adicionales a las de pegado al DNA, que evocan la posibilidad de que las proteínas sean el producto del intercambio de dominios con proteínas con funciones previas. A continuación se dan ejemplos de reguladores transcripcionales con funciones enzimáticas y que han acoplado sus diferentes funciones optimizando la regulación transcripcional:

El represor BirA del operón *bio* en *E. coli*, actúa adicionalmente como biotina-5'-adenilato sintetasa (EC. 6.3.4.15). El HTH se localiza en la región NH₂-term (18-39

aa), mientras que el motivo para la activación por biotina se ubica en las posiciones: 83-119 y 189-235 (8, 67).

Ada, actúa como O-6-metilguanín-DNA alquiltransferasa (EC 2.1.1.63), involucrada en la respuesta adaptativa. La función de activación transcripcional se ubica en el NH₂-term: reparación del DNA por transferencia de un grupo alquilo en posición O-6 de la guanina a un residuo cisteína-69 (39).

PutA, una prolina deshidrogenasa (EC 1.5.99.8) (prolina oxidasa)/delta-1-pirrolina-5-carboxilato deshidrogenasa (EC 1.5.1.12) (P5C deshidrogenasa). Oxida la prolina a ácido glutámico (como fuente de carbono/nitrógeno) y también actúa como un regulador transcripcional (37). Maloy (35) propone que PutA es una proteína multifuncional que pudo haber evolucionado para prevenir ciclos fútiles de síntesis endógena de prolina.

¿Tiene alguna explicación estructural la posición del HTH con respecto a su actividad?

Al parecer la relación "ubicación del HTH-actividad reguladora" no tiene una explicación aparente desde el punto de vista estructural, debido a las siguientes evidencias:

1. El dominio de pegado al DNA no determina la función de represor, activador o dual de la proteína, sino que es un componente necesario para que la proteína se pegue al DNA.

2. Aunque las estructuras de reconocimiento al DNA dentro de una proteína están relativamente dispersos en su estructura primaria o secuencia de aminoácidos, cuando se observan tridimensionalmente, sus posiciones pueden ser similares (56).

3. Al sobreponer los HTH's de LacI (localizado en el NH₂-term), GalR (localizado en el NH₂-term) y CRP (localizado en el COOH-term), se observan homologías significativas entre la estructuras que forman las tres proteínas, concluyendo que hay un mecanismo común por el cual las proteínas pueden reconocer bases específicas en el DNA-doble cadena, por medio de los aminoácidos de las α -hélices que se ubican en el surco mayor del B-DNA (64).

4. El HTH lo podemos encontrar en posición NH₂- o COOH-term, para los represores y activadores predominantemente, sin embargo, encontramos represores-HTH-COOH-term, activadores-HTH-NH₂-term, represores-HTH-posición central y HTH-duales en ambos extremos proteicos, aunque en menor proporción que los grupos principales ya descritos.

5. Los extremos NH₂-term y COOH-term no quedan cercanos al interactuar con el DNA para el caso de CRP. Cuando observamos la estructura tridimensional del represor Fis, la separación entre los extremos terminales es proporcional a una α-hélice que se encuentra entre ambas proteínas al asociarse como dímero proteico.

PROTEÍNAS REGULADORAS DE EUKARIOTES

La descripción de la relación estructura-función en los reguladores transcripcionales de procariontes y en particular, aquellos que están asociados al factor sigma54, nos planteó la necesidad de revisar los reguladores transcripcionales de eucariotes para tratar de describir una relación similar a la encontrada en proteínas con HTH. Por lo cual, se analizaron tres motivos estructurales: los homeodominios, los dedos de zinc y los hélice-loop-hélice (HLH). Como en el caso de procariontes, la ubicación de los motivos de pegado al DNA en eucariotes se conserva en los homeodominios y dedos de zinc (repetidos en "tandem" de dos dedos). Este análisis muestra que las estructuras de pegado al DNA no se encuentran organizados aleatoriamente en la proteína. Para el caso de las proteínas con HLH, el análisis mostró la existencia de dos picos predominantes hacia la región central y COOH-term de las proteínas, donde parece ser que la distribución tampoco está determinada por el azar. Ahora bien, lo que debemos determinar es la función asociada a cada proteína (represor, activador o dual) para determinar si realmente existe una relación directa entre la posición del dominio de pegado al DNA y la actividad regulatoria de la proteína, como la observada en proteínas procariontes.

CONCLUSIONES

1. De acuerdo a los resultados descritos previamente, las proteínas reguladoras de promotores $\sigma 70$ en procariontes se pueden asociar en 3 grupos funcionales: represores-NH₂-term, activadores-COOH-term y duales (familia LysR)-NH₂-term.

2. La probabilidad de encontrar un represor en el extremo NH₂-term es de 3 a 4 veces más alta que la de encontrar un activador en este mismo extremo. Sin embargo, la probabilidad de encontrar una proteína dual-NH₂-term depende del porcentaje de similitud/identidad con respecto a los miembros de la familia LysR. En el extremo COOH-term es 3 veces más probable encontrar proteínas activadoras que represoras.

3. El HTH convencional (descrito en proteínas de procariontes y fagos) es una estructura conservada en secuencia, función y localización dentro de cada familia de reguladores. Por ejemplo, todas las proteínas de la familia GalR/LacI tienen un dominio de unión al DNA (que contiene el HTH) de 69 residuos de longitud, en el primer 20% del extremo NH₂-term; mientras todas las proteínas de la familia CRP presentan el HTH en el último 30% del extremo COOH-term. Las únicas proteínas que presentan el HTH en posición central se agrupan en una sola familia de represores "ArsR". Finalmente, las proteínas de la familia "EBP" o activadores de la transcripción en promotores $\sigma 54$, presentan el HTH en los últimos 40 aa del extremo COOH-term.

4. Los factores sigma de las familias $\sigma 70$ y $\sigma 54$, presentan el HTH en una posición definida y prácticamente exclusiva para cada una de las familias analizadas. Mientras en las proteínas de la familia $\sigma 70$, el HTH se localiza en los últimos 60 aa del extremo COOH-term, las proteínas de la familia $\sigma 54$ lo presentan a una distancia de aproximadamente 100 aa con respecto al COOH-term.

5. El análisis posicional de las estructuras que se unen al DNA en las proteínas nos muestra que se pueden diferenciar perfectamente las proteínas evolutivamente relacionadas (y que en algunos casos ya han sido clasificadas en familias). Este trabajo nos sugiere además, que con la descripción y localización de un motivo estructural podemos clasificar proteínas nuevas y reforzar los criterios de clasificación proteica.

6. Es posible que el HTH sea un motivo estructural originado a partir de un

ancestro común a todos los reguladores transcripcionales y factores sigma de procariontes, así como de las homeoproteínas de eucariontes. Inclusive podemos decir que esta estructura se originó antes del evento de divergencia de los procariontes, los eucariontes y las arqueobacterias; ya que en estos organismos también se han descrito proteínas con el HTH. Es probable que cada familia de reguladores tenga un proteína ancestral con una función definida, a la cual se le fusionó el DBD (que contiene el HTH) a través de un evento de recombinación, en posición NH₂-term o COOH-term. Posteriormente, esta proteína ancestral originó el resto de los miembros en cada grupo, por duplicación génica y diversificación.

8. En el caso de los reguladores transcripcionales de eucariontes, la distribución homogénea del motivo de unión al DNA nos sugiere que el evento de diversificación de los motivos estructurales sigue un patrón no aleatorio y que puede ser diferenciable para cada grupo de proteínas. Siguiendo la tendencia de los reguladores procariontes, es probable que la función de regulación también esté asociada a la posición de cada motivo de unión al DNA estudiados.

PERSPECTIVAS

1. El análisis posicional de los motivos proteicos se puede utilizar para realizar predicciones en proteínas con función desconocida. En el caso de las proteínas que se unen al DNA con el motivo HTH, la posición de esta estructura nos puede ayudar a determinar si una proteína es reguladora y cuál es su función más probable.

A continuación describiremos un primer ejemplo de la utilización de información la posicional de los motivos estructurales (en este caso del HTH), para predecir posibles reguladores transcripcionales bacterianos:

A partir de dos conjuntos de secuencias de proteínas relacionadas que presentaban el motivo HTH, se definieron y realizaron dos matrices de peso independientes, que representan la frecuencia de un aminoácido en una posición dentro de un conjunto de secuencias alineadas. Se utilizaron 8 secuencias de *E. coli* con el HTH descrito por análisis bioquímicos y/o genéticos; y 6 secuencias de reguladores transcripcionales de fagos. Para cada una de estas matrices, se realizó una búsqueda en todas las proteínas de *E. coli* reportadas en la base de datos Swiss-Prot versión 31.0 (3, 4) para detectar todas las proteínas que presentan un posible HTH. Una vez que se localizó el patrón de este motivo estructural, verificamos su posición central con respecto al total de la proteína, obteniéndose lo siguiente (se utilizó el programa Profile de GCG (22)):

Proteína Detectada	HTH_E.coli	HTH_fago	Función Posible	Long
ALPA_E.coli	14-31(3.67)	13-31 (2.67)	Represor/Dual	70aa
FECI_E.coli	141-158(4.00)	140-158(4.20)	Activador	173aa
TTK_E.coli	48-67(3.57)	49-62(2.68)	Represor	212aa
YFHH_E.coli	62-80(2.66)	62-80(3.59)	Represor	282aa
HIPa_E.coli	28-47(6.41)	28-47(11.84)	Represor	440aa

De forma independiente, las 2 matrices reconocieron a las 5 proteínas. Los números entre paréntesis corresponden a la calificación para cada posible HTH detectado. En el caso de la matriz generada a partir de las secuencias de *E. coli*, la calificación más alta para una proteína es de 7.69 (CRP_E. coli) y la más baja de 2.52 (ORF_1877817); mientras que para la matriz generada con secuencias de proteínas de fagos, la calificación más alta es de 11.84, (HipA_E. coli) y la más baja de 2.53

(ORF_3495728). Se realizó la predicción de la función más probable para las 5 proteínas detectadas con ambas matrices con base en la posición de sus HTH's. Esta predicción funcional se comparó con la función que les asigna Swiss-Prot, observándose que las 2 proteínas Alpa y Ttk (dual y represor) coinciden con la función descrita en la base de datos; las proteínas YfhH y HipA que nosotros predécimos como represoras, SwissProt (2,3) las describe con función desconocida; finalmente FecI_E. coli a quién la asignamos la actividad de represor, SwissProt (2,3) la describe como un posible factor σ .

2. Podríamos elaborar un reconecedor para proteínas con función desconocida o bien un modelo que describa las reglas de posición del HTH con relación a otros motivos para los reguladores transcripcionales de procariontes:

 Posición central (porcentual y absoluta) del HTH para cada grupo funcional (represor, activador y dual), además para cada familia de reguladores;

 Actividad para cada familia de proteínas reguladoras;

 Posición de otros dominios involucrados en la función de las proteínas (multimerización, reconocimiento al inductor, entre otros);

3. Una alternativa interesante, es reconstruir la historia evolutiva de todas las proteínas reguladoras de la transcripción en procariontes y tratar de dilucidar las relaciones o conexiones existentes entre cada una de las familias de reguladores por arriba de este nivel de clasificación, es decir, intentar agrupar estas proteínas en superfamilias. También podríamos realizar un análisis evolutivo en las diferentes proteínas reguladoras de eucariontes y que presentan los diferentes motivos de unión al DNA: homeodominios, dedos de zinc y hélice-loop-hélice.

REFERENCIAS

1. Adhya, S. and S. Garges. 1990. Positive control. *J. Biol. Chem.* 265(19): 10797-10800.
2. Bairoch, A. 1993. A possible mechanism for metal-ion induced DNA-protein dissociation in a family of a prokaryotic transcriptional regulators. *Nucleic Acids Res.* 21(10):2515.
3. Bairoch A. and R. Apweiler. 1996. The Swiss-Prot protein sequence databank and its new supplement TREMBL. *Nucleic Acids Res.* 24(1):21-25.
4. Bairoch, A., and B. Boeckmann. 1992. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.* 20:2019-2022.
5. Beckwith, J. 1987. The Lactose operon. p. 1444-1452. In F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed). *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. Vol. II. American Society for Microbiology, Washington, D. C.
6. Branden, C. and J. Tooze. 1991. Introduction to Protein Structure. Garland Publishing, Inc. New York and London. p. 85-110.
7. Brent R. and M. Ptashne. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 43: 729-36
8. Buoncristiani, M.R., P. K. Howard & A. J. Otsuka. 1986. DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of *Escherichia coli*. *Gene* 44: 255-61
9. Busby S. and R. H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743-746.
10. Chothia, C. 1992. One thousand families for the molecular biologist. *Nature*. 357, 543-544.
11. Collado-Vides, J., B. Magasanik and J. D. Gralla. 1991. Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* 55(3):371-394.
12. Cui, J and R. L. Somerville. 1993. Mutational uncoupling of the transcriptional activation function of the TyrR protein of *Escherichia coli* K-12 from the repression function. *J. Bacteriol.* 175: 303-6.
13. Damante, G., L. Pellizzari, G. Esposito, F. Fogolari, P. Viglino, D. Fabbro, G. Tell, S. Formisano and R. D. Lauro. 1996. A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO Journal*, 15(18):4992-5000.
14. Doolittle R. F. and P. Bork. 1993. Evolutionary mobile modules in proteins. *Scientific American*. October. 34-40.
15. Ebright, R. H. and S. Busby. 1995. The *Escherichia coli* RNA polymerase α

- subunit: structure and function. *Curr. Op. Gen. and Dev.* 5:197-203.
16. Eick, D., A. Wedel and H. Heumann. 1994. From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. *Trends in Genet.* 10:292-296.
 17. Feng D-F. and R. F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25: 351-360.
 18. Gallegos, M-T., C. Michán and J. L. Ramos. 1993. The XylIS/AraC family of regulators. *Nucleic Acid Research.* 21(4):807-810
 19. Gralla, J. D. 1990. Promoter recognition and mRNA initiation by *Escherichia coli* Ec70. *Methods in Enzymology.* 185:37-54.
 20. Gralla, J. 1992. *lac* repressor. p. 629-642. In S. L. McKnight and K. R. Yamamoto (ed), *Transcriptional Regulation. Vol. 2.* Cold Spring Harbor Laboratory Press, New York.
 21. Gralla J. D. and J. Collado-Vides. 1996. Organization and function of transcriptional regulatory elements. p. 1232-1246. In F. C. Neidhart, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umberger (ed), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology, Vol.1.* American Society for Microbiology, Washington, D. C.
 22. Gribskov, M., R. Lüthy, and D. Eisenberg. 1990. Profile Analysis. In *Methods in Enzymology.* 183; 146-159.
 23. Gross, C. A., M. Lonetto, and R. Losick. 1992. Bacterial Sigma Factors. p. 129-176. In S. L. McKnight and K. R. Yamamoto (ed), *Transcriptional Regulation, Vol. 1.* Cold Spring Harbor Laboratory Press, New York.
 24. Harrison, S. C. 1991. A structural taxonomy of DNA-binding domains. *Nature* 353: 715-9.
 25. Harrison, S. C. and A. K. Aggarwal. 1990. DNA Recognition by proteins with the helix-turn-helix motif, *Annu. Rev. Biochem.* 59: 933-69.
 26. Haydon, D. and J. Guest. 1991. A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.* 79: 291-296.
 27. Henkoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace. 1988. A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. U.S.A.* 85: 6602-6.
 28. Kaptain, R. 1993. Protein-nucleic acid interactions by NMR. *Curr. Op. Struct. Biol.* 3:50-56.
 29. Klug, A. 1993. Protein designs for the specific recognition of DNA. *Gene.* 135:83-92
 30. Lewin, B. 1994. *GENES V.* Oxford University Press. 1272 p.

31. Lonetto, M., M. Gribskov and C. A. Gross. 1992. The $\sigma 70$ family: Sequence conservation and evolutionary relationships. *J. of Bacteriol.* 174(12): 3843-3849.
32. McClure, W. R. 1985. Mechanism and control of transcription initiation in prokaryotes. *Ann. Rev. Biochem.* 54:171-204.
33. Magasanik, B. 1996. Regulation of nitrogen utilization. p.1344-1356. In F. C. Neidhart, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, Vol. I. American Society for Microbiology, Washington, D. C.
34. Makino, K., M. Amemura, T. Kawamoto, S. Kimura, H. Shinagawa, A. Nakata, and M. Suzuki. 1996. DNA binding of PhoB and its interaction with RNA polymerase. *J. Mol. Biol.* 259:15-26.
35. Maloy, S. 1987. The proline utilization operon. p. 1513-1519. In F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, Vol. II. American Society for Microbiology, Washington, D. C.
36. Margalit, H., B. A. Shapiro, R. Nussinov, J. Owens, and R. L. Jernigan. 1988. Helix stability in Prokaryotic promoter regions. *Biochemistry*, 27: 5179-5188.
37. Menzel R. and J. Roth. 1981. Purification of the *putA* gene product. *J. Biol. Chem.* 256(18): 9755-9761.
38. Morett, E. and L. Segovia. 1993. The $\sigma 54$ bacterial enhancer-binding protein family: Mechanism of action and phylogenetic relationship of their functional domains. *J. of Bacteriol.* 175(19):6067-6074.
39. Myers, L. C., G. L. Verdine, and G. Wagner. 1993. Solution structure of the DNA methyl phosphotriester repair domain of *Escherichia coli* Ada. *Biochemistry*, 32:14089-14094.
40. Neidhart F. C. 1987. Multigene systems and regulons. p. 1313-1317. In F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, Vol. II. American Society for Microbiology, Washington, D. C.
41. Neidhart F. C., J. L. Ingraham and M. Schaechter. 1990. Physiology of the bacterial cell. A Molecular Approach. SINAUER, p. 351-388.
42. Nelson, H. C. M. 1995. Structure and function of DNA-binding proteins. *Curr. Op. Gen. and Dev.* 5:180-189.
43. Nguyen, C. C. and M. H. Saier. 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcriptional factors. *FEBS Lett.* 377:98-102.

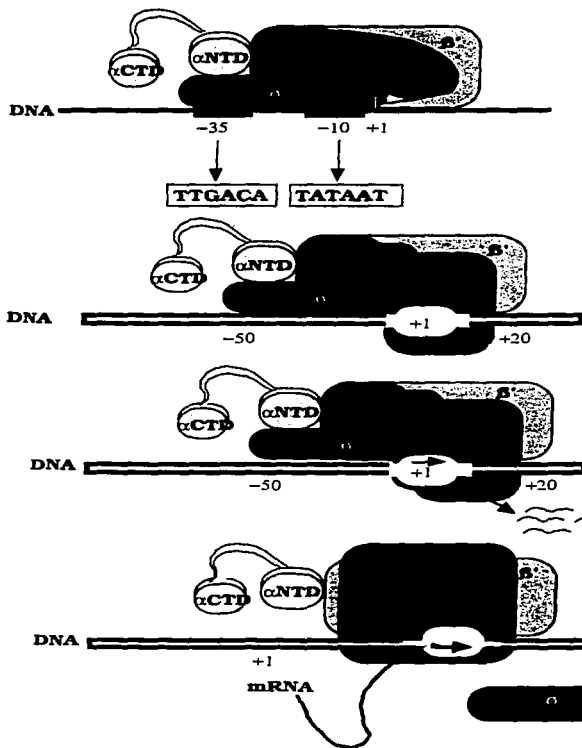
44. Ninfa, A. J. 1996. Regulation of gene transcription by extracellular stimuli. p. 1246-1262. In F. C. Neidhart, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, Vol. 1. American Society for Microbiology, Washington, D. C.
45. Pao G. M. and M. H. Saier Jr. 1995. Response regulators of bacterial signal transduction systems: Selective domain shuffling during evolution. *J. Mol. Evol* 40: 136-154.
46. Phillips, S. EV. 1994. Built association: structure and function of helix-loop-helix DNA-binding proteins. *Structure* 2:1-4.
47. Platko J. V. and J. M. Calvo, 1993. Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J. of Bacteriol.* 175: 1110-7.
48. Ramos, J. L., F. Rojo, L. Zhou and K. N. Timmis. 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and *Escherichia coli* AraC activators. *Nucleic Acids Res.* 18(8): 2149-2152.
49. Record, M. T., W. S. Reznikoff, M. L. Craig, K. L. McQuade, and P. J. Schlax. 1996. *Escherichia coli* RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation. p. 792-821. In F. C. Neidhart, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, Vol. 1. American Society for Microbiology, Washington, D. C.
50. Riley, M. and B. Labedan, 1996. *Escherichia coli* gene products: Physiological functions and common ancestries. p. 2118-2202. In F. C. Neidhart, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, Vol. II. American Society for Microbiology, Washington, D. C.
51. Sasse-Dwight S. and J. D. Gralla. 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor σ54. *Cell.* 62: 945-954.
52. Schell M. A. and M. Sukordhaman. 1989. Evidence that the transcription activator encoded by the *Pseudomonas putida nahR* gene is evolutionarily related to the transcription activators encoded by the *Rhizobium nodD* genes. *J. of Bacteriol.* 171: 1952-1959.
53. Schell, M. A. 1993. Molecular biology of the LysR family of transcriptional

- regulators. *Annu. Rev. Microbiol.* 47:597-626.
54. Spiro, S. 1994. The FNR family of transcriptional regulators. *Antoine van Leeuwenhoek* 66:23-36.
55. Stock, J. B. A. J. Ninfa and A. M. Stock. 1989. Protein phosphorylation of adaptative responses in bacteria. *Microbiol. Rev.* 53(4):450-490.
56. Suzuki, M. and S. E. Brenner. 1995. Classification of multi-helical DNA-binding domains and application to predict the DBD structures of sigma factor. *LysR, OmpR/PhoB, CENP-B, RapI, and Xy1S/Ada/AraC*. *FEBS Lett.* 372:215-221
57. Suzuki, M. 1995. DNA recognition by a β -sheet. *Protein Engineering*. 8 (1): 1-4.
58. Tullius, T. 1995. Homeodomains: together again for the first time. *Structure* 3:1143-1145.
59. Vartak, N. B., J. Reizer, A. Reizer, J. T. Gripp, E. A. Groisman, L.-F. Wu, J. M. Tomich and M. H. Saier Jr. 1991. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: 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.
60. Vidal-Ingigliardi, D., E. Richet, O. Danot and O. Ratbaud. 1993. A small C-terminal region of the *Escherichia coli* MalT protein contains the DNA-binding domain. *J. Biol. Chem.* 268: 24527-30.
61. Von Hippel, P. H., D. G. Bear, W. D. Morgan and A. McSwiggen. 1984. Protein-nucleic acid interactions in transcription: A molecular analysis. *Ann. Rev. Biochem.* 53:389-446.
62. Von Hippel, P. H., T. D. Yager and S. C. Gill. 1992. Quantitative aspects of the transcription cycle in *Escherichia coli*. p. 179-201. In S. L. McKnight and K. R. Yamamoto (ed), *Transcriptional Regulation*, Vol. I. Cold Spring Harbor Laboratory Press, New York.
63. Wanner, B. 1987. Phosphate regulation of gene expression in *Escherichia coli*. p. 1326-1333. In F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, Vol. II. American Society for Microbiology, Washington, D. C.
64. Weber, I. T., D. B. McKay and T. A. Steitz. 1982. Two helix DNA binding motif of CAP found in *lac* repressor and *gal* repressor. *Nucleic Acids Res.* 10(16):5085-5102.
65. Weickert, M. J. and S. Adhya. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* 267 (22):15869-15874.
66. Wharton, R.P., E. L. Brown and M. Ptashne. 1984. Substituting an alpha-helix

- switches the sequence-specific DNA interactions of a repressor. *Cell* 38: 361-9.
67. Wilson, K. T., L. M. Schewchuk, R. G. Brennan, A. J. Otsuka and B. W. Matthews. 1992. *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl. Acad. Sci. USA.* 89:9257-9261.
68. Yura, K. S. Tomoda and M. Go. 1993. Repeat of a helix-turn-helix module in DNA-binding proteins. *Protein Engineering.* 6(6):621-628.

Fig. 1. MODELO DE LA TRANSCRIPCION EN PROMOTORES $\sigma 70$

Modificado de Ebright and Busby, 1995 y Eick, D. et al, 1994.



Complejo RNA pol+Factor $\sigma 70$. formando el complejo cerrado. La RNAPol- $\sigma 70$ reconoce el promotor de forma específica...

Formación del complejo abierto. La RNAPol- $\sigma 70$ abre el DNA, formando la burbuja de transcripción...

Transcripción abortiva. Se sintetizan fragmentos de alrededor 9 nucleótidos de longitud...

mRNA's abortivos pequeños

Transcripción productiva. Se libera el factor sigma, para ser reutilizado en otro proceso transcripcional.

ESTRUCTURA HELICE-VUELTA-HELICE

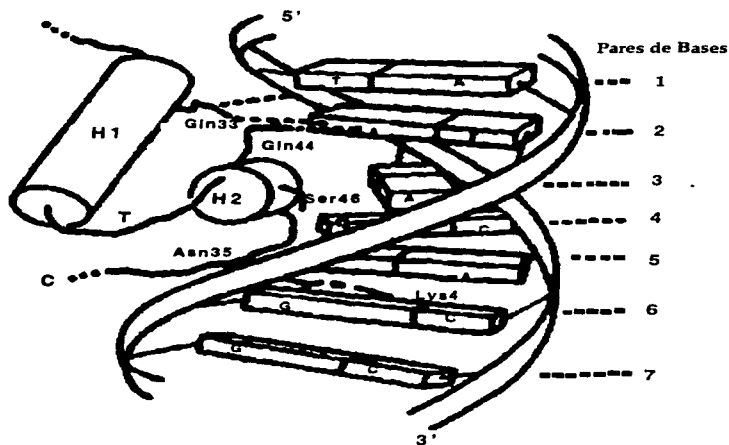


Fig. 2. Estructura de pegado al DNA, hélice-vuelta-hélice (HTH). La α -hélice 1 (NH₂-terminal) contacta con los grupos fosfato del DNA, mientras que la α -hélice 2 contacta con nucleótidos específicos del surco mayor del DNA. Esquema tomado de Adhya y Garges (1).

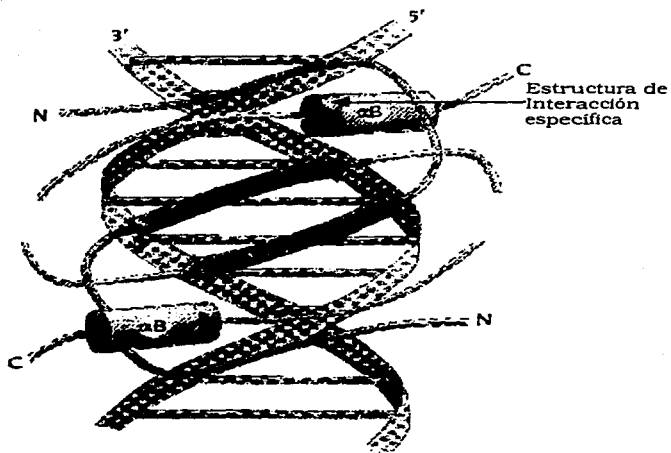


Fig. 3. Esquema de la interacción de la estructura β -plegada antiparalela del represor MetJ de *E. coli* y el DNA. Las dos cadenas β se unen a secuencias específicas en el surco mayor del DNA. Las interacciones con los grupos fosfato del DNA la realizan los "loops" del NH_2 -term.

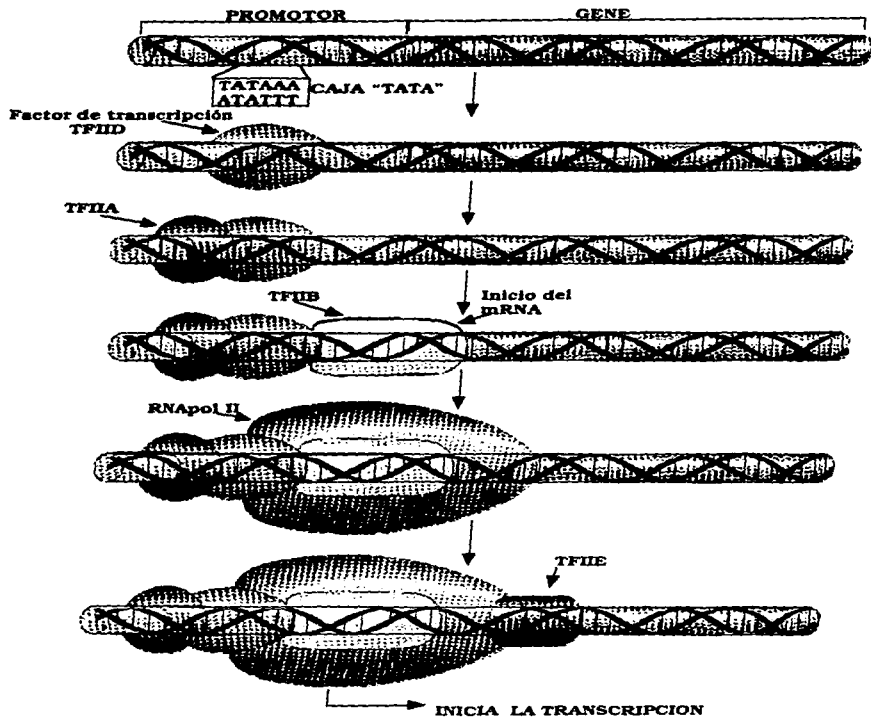


Fig. 4. Esquema del mecanismo de la transcripción en células eucariotes. La formación del complejo de transcripción se lleva a cabo por interacciones sucesivas de proteínas (factores) con la RNAPol II, que es necesario para el inicio de la transcripción. Para detalles ver el texto de la página 9.

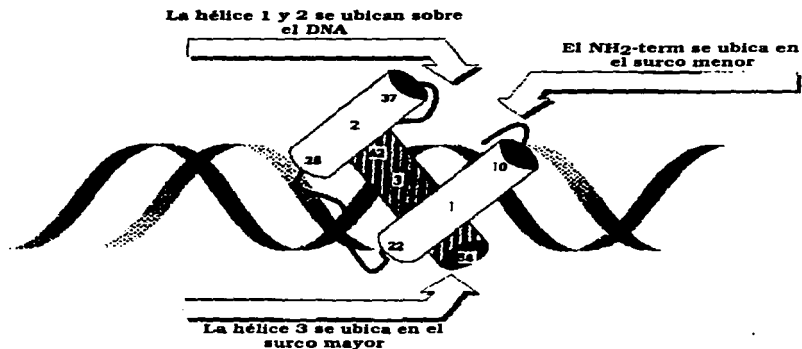


Fig. 5. La α -hélice 3 del homeodominio se pega en el surco mayor del DNA, con las α -hélices 1 y 2 ubicadas fuera de la doble hélice. La α -hélice 3 contacta tanto los grupos fosfatos y las bases específicas. El brazo NH₂-terminal, se ubica en el surco menor haciendo contactos adicionales (Lewin, 1994).

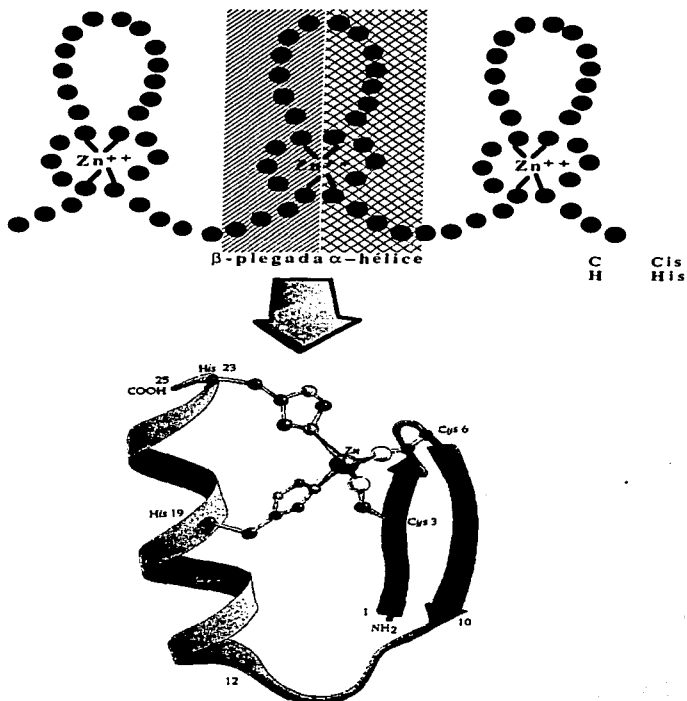


Fig. 6. Se muestra un clásico dedo de zinc. Un átomo de zinc se une a dos cisteínas y dos histidinas. Este motivo estructural se localiza preferentemente en extremo NH₂-term, sin embargo podemos encontrar casos (receptor de glucocorticoides) que lo presentan en posición central.



Fig. 7. Hélice-loop-hélice. La estructura consiste de dos α -hélices unidas por un loop o bucle. Las proteínas que presentan este motivo se asocian en dímeros.

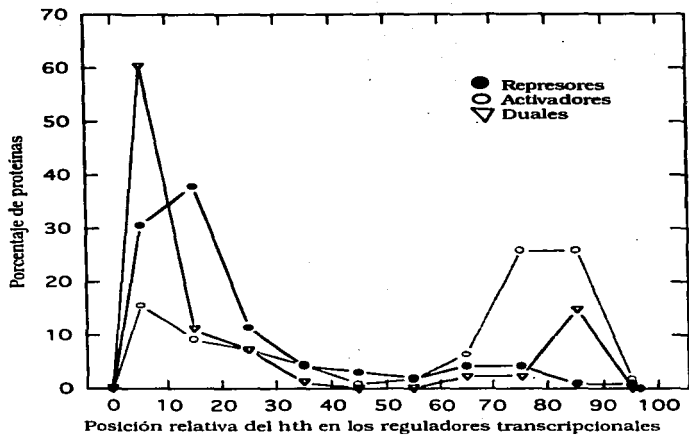


Fig. 8. Distribución del HTH en el total de proteínas reguladoras asociadas al factor $\sigma 70$. En el eje de las "X" el 0% representa el NH₂-term y el 100% el extremo COOH-term, mientras que en el eje de las "Y" se representa el % de proteínas represoras, activadoras y duales. Para detalles ver la sección de resultados, pág. 19. Para el cálculo de los % del eje de las "Y" se tomó como 100% el total de activadores, represores y duales por separado.

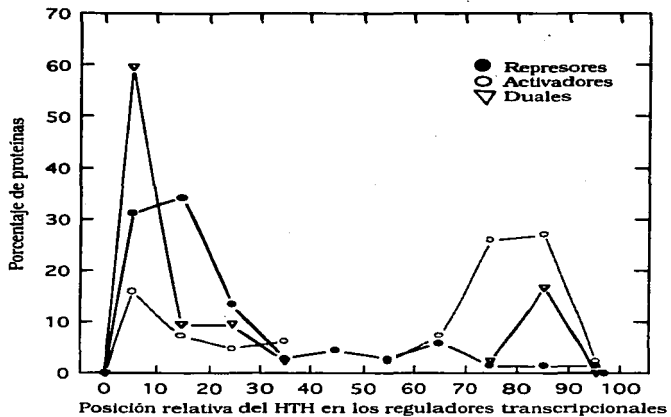


Fig. 9. Distribución del HTH en el total de proteínas reguladoras asociadas al factor $\sigma 70$, eliminando a las proteínas ortólogas. En el eje de las "X" el 0% representa el extremo NH₂-term y el 100% el extremo COOH-term.

Los cálculos para la obtención del porcentaje se realizaron de forma similar a la figura 8.

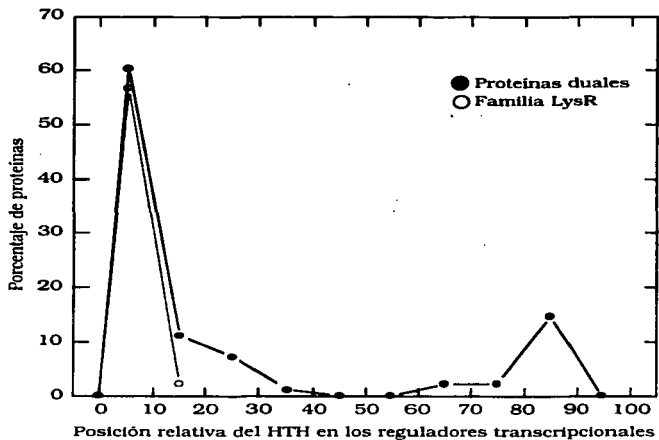


Fig. 10. Distribución del HTH en proteínas reguladoras con actividad dual, asociadas al factor $\sigma 70$, en comparación con la familia LysR. En el eje de las "X" el 0% representa el NH₂-term y el 100% el COOH-term. La distribución no cambia cuando se excluyen las proteínas ortólogas.

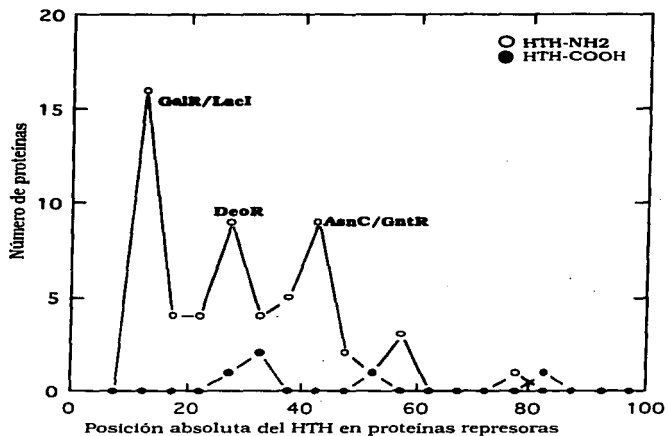


Fig. 11a. La figura muestra la posición absoluta del HTH en el total de proteínas reguladoras asociadas al factor $\sigma 70$. En el eje de las "X" se representa la distancia desde el HTH al extremo proteico más cercano (NH₂-term o COOH-term). Además, se muestran las familias que predominan en cada uno de los picos.

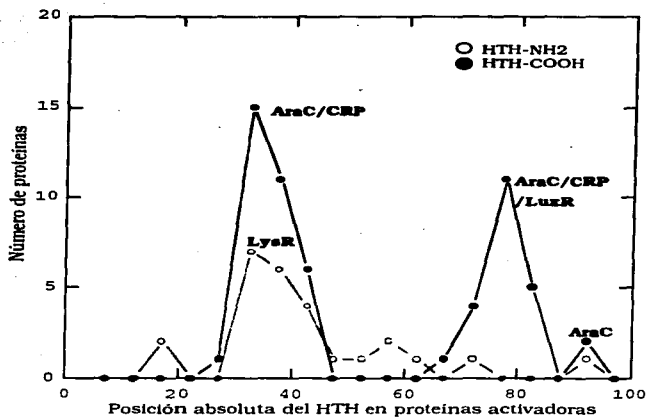


Fig. 11b. La figura muestra la posición absoluta del HTH en el total de proteínas reguladoras asociadas al factor $\sigma 70$. En el eje de las "X" se representa la distancia desde el HTH al extremo proteico más cercano (NH2- o COOH-term). También se muestran las familias que predominan en cada uno de los picos.

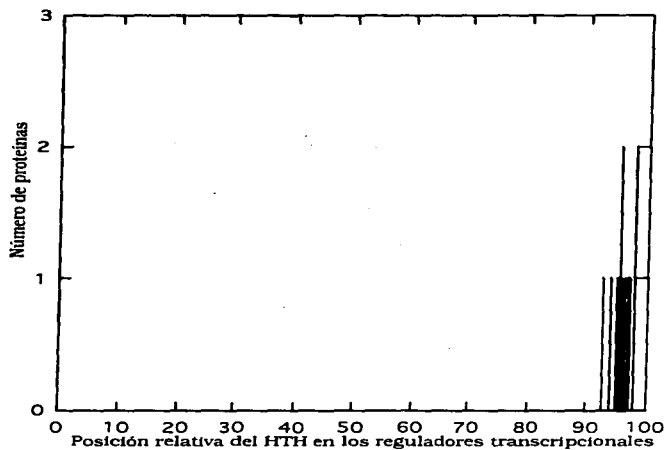


Fig. 12a. Distribución del HTH en las proteínas reguladoras asociadas al factor σ_{54} . En el eje de las "X" el 0% representa el extremo NH₂-term. y el 100% el extremo COOH-term. Sólo se graficaron 30 proteínas activadoras. Se excluyen de la gráfica 10 proteínas duales, 2 represores y 3 con función no determinada.

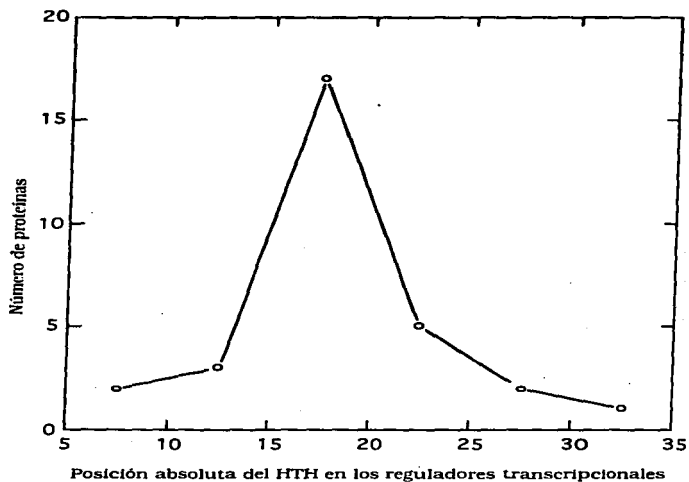


Fig. 12b. Distribución absoluta del HTH en el total de proteínas reguladoras asociadas al factor $\sigma 54$. En el eje de las "X" se representan los últimos 35 aminoácidos del extremo COOH-term de este conjunto de proteínas donde se ubica predominantemente el HTH.

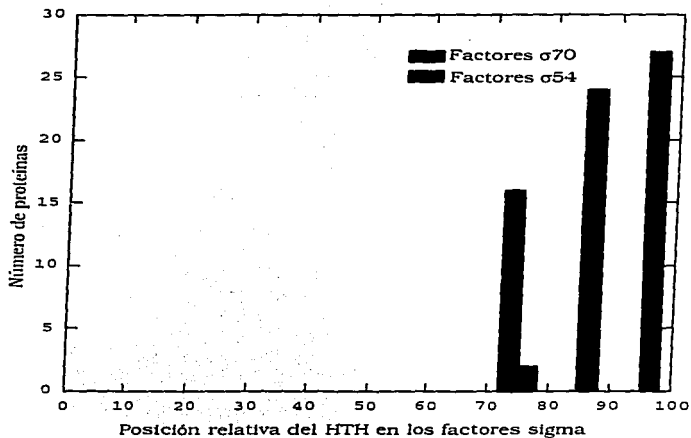


Fig. 13b. Distribución del HTH en el total de factores sigma de las familias $\sigma 70$ y $\sigma 54$. Se observa una clara diferencia entre ambas familias con base en la posición del HTH. En el eje de las "X" el 0% representa el extremo NH₂-term y el 100% el extremo COOH-term. Para detalles ver la sección de resultados, pág. 19.

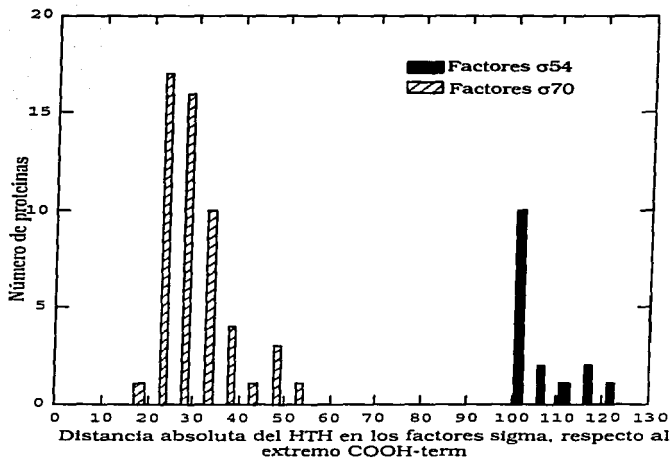


Fig. 13b. Distribución absoluta del HTH en el total de factores sigma de las familias $\sigma 70$ y $\sigma 54$. Esta gráfica muestra una clara separación de familias con base en la posición del HTH. Mientras el HTH se localiza en los últimos 60 aa del extremo COOH-term en la familia $\sigma 70$; en la familia $\sigma 54$ se ubica a una distancia cercana a los 100 aa respecto al extremo COOH-term.

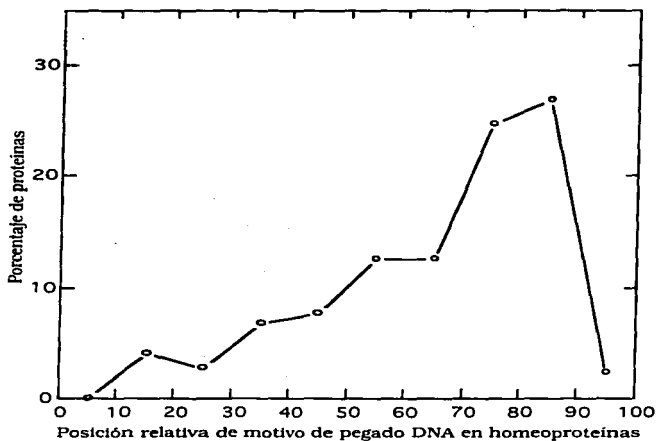


Fig. 14. Distribución del HTH en proteínas reguladoras de eucariontes (homeodominios). En el eje de las "X" el 0% representa el NH₂-term y el 100% el COOH-term. Se observa la tendencia de esta estructura de pegado al DNA a ubicarse hacia el COOH-term de las proteínas.

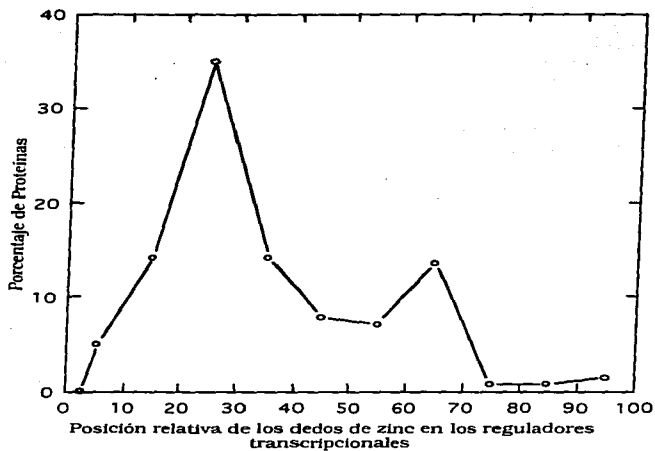


Fig. 15. Distribución de los dedos de zinc en reguladores transcripcionales de eucariotes. Se representa la estructura repetida dos veces en tandem. En el eje de las "X" el 0% representa el NH₂-term. y el 100% el COOH-term.

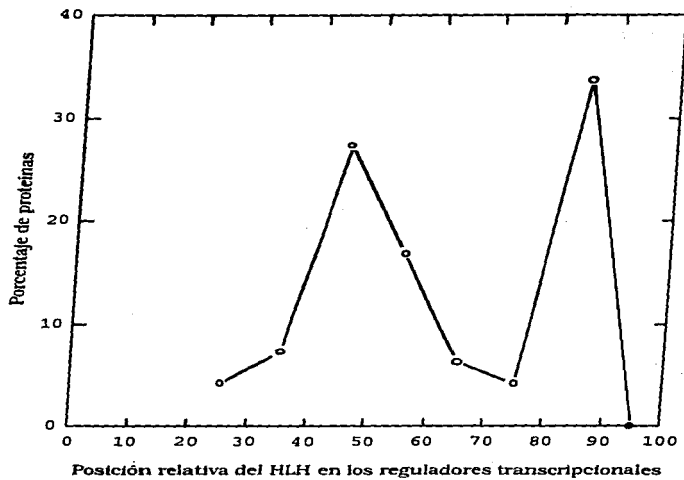


Fig. 16. Distribución del hélice-loop-hélice (HLH) en proteínas reguladoras de eucariotes. En el eje de las "X" el 0% representa el NH₂-term y el 100% el COOH-term.

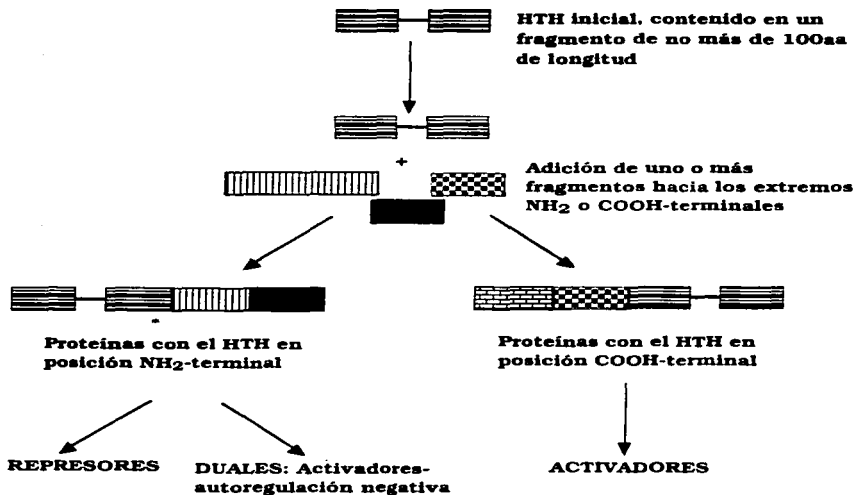


Fig. 17. Esquema hipotético que muestra el posible camino evolutivo para los grupos proteicos que presentan el HTH en posición NH₂-term o COOH-term, y que fue probablemente debida a la adición de fragmentos con funciones previas a un HTH ancestral. La divergencia de actividades regulatorias (+, - o +/-) asociado a la posición del HTH, se produjo como una consecuencia directa de estas adiciones.

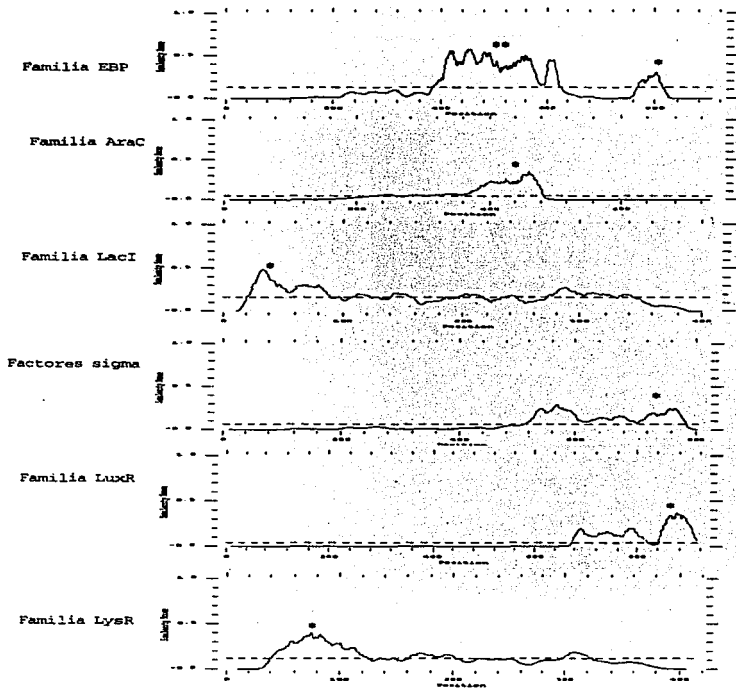


Fig. 18. Perfil de identidad (Plotsimilarity-GCC) para las familias de proteínas reguladoras de procariontes. La región más conservada está localizada en el DBD (*) que contiene el HTH. Para el caso de la familia EBP, se presenta otra región conservada, que corresponde al probable sitio de interacción con el factor sigma (*) y de fosforilación (**).

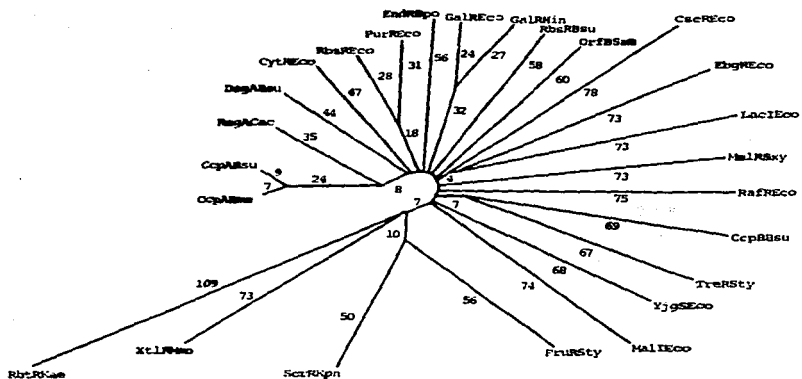


Fig. 19. Análisis evolutivo de la familia GalR/LacI. La longitud de las ramas es proporcional a la distancia filogenética. Se observa casi una perfecta simetría radial, lo que implica que todas las proteínas resultaron de eventos de duplicación ocurridos en un mismo momento de la historia evolutiva (43).

ORGANIZACION DEL DOMINIO DE PEGADO AL DNA EN LAS FAMILIAS REGULADORAS DE PROCARIONTES

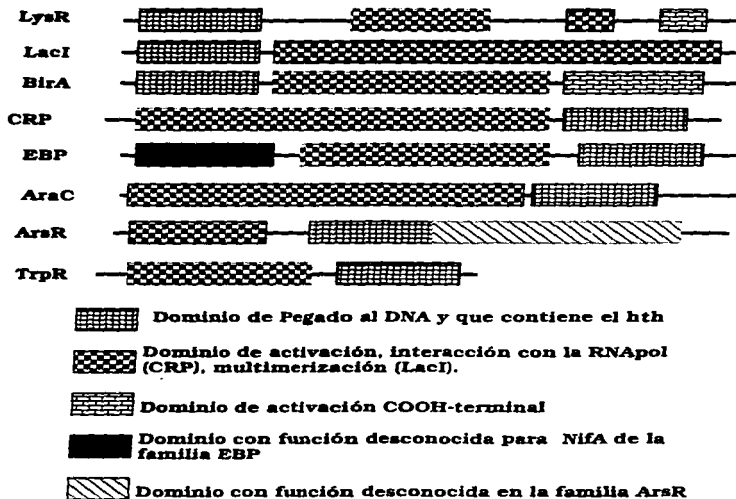


Fig. 20. Esquema de la organización modular de las proteínas reguladoras asociadas a promotores $\sigma 70$ y $\sigma 54$. El Dominio de pegado al DNA esta generalmente conservado en tamaño (aproximadamente 100 aa).

Sigma	Organismo	Región -35	Región -10
$\sigma 70$	<i>E. coli</i>	TTGACA	TATAAT
$\sigma 32$	<i>E. coli</i>	TCTC-CCCTTGAA	CCCCAT-TA
σA	<i>B. subtilis</i>	TTGACA	TATAAT
σB	<i>B. subtilis</i>	AGGTTTAA	GGGTAT
σD	<i>B. subtilis</i>	CTAAA	CCGATAT
σE	<i>B. subtilis</i>	ATATT	ATACA
σG	<i>B. subtilis</i>	TGAATA	CATACTA
σK	<i>B. subtilis</i>	AC	CATA--T
σH	<i>B. subtilis</i>	CAGGA	GAATT--T
ogp28	fago SPO1	AGGAGA	TTT-TTT ^a
ogp55	Fago T4	Ninguna	TATAAATA
Sigma	Organismo	Región -24	Región -12
$\sigma 54$	<i>E. coli</i> , <i>S. typhimurium</i> <i>K. pneumoniae</i> <i>Rhizobium sp.</i>	CTGGA-A	TTGCA

Tabla 1. Secuencias consenso de promotores reconocidos por la RNA polimerasa conteniendo varios factores sigma (23). T^a de SPO1 representa hidroximetiluracil que reemplaza a la Timina.

FAM	PRINCIPALES FUNCIONES	REG	NH2-	COOH-	RANGO	MEDIA	TAM
LysR	Síntesis de aminoácidos	+/-	96.4%	122-344	297.46	100	
GalR	Catabolismo de azúcares	-	100%	94-360	318.08	36	
GntR	Catabolismo de ácidos grasos e histidina	-	84.6%	109-493	258.62	27	
MerR	Resistencia a mercurio	-	95%	122-273	164.20	20	
DeoR	Metab. de fucosa, glucitol y deoxiribosa	-	100%	189-269	248.83	18	
TetR	Resistencia a tetraciclina	-	76.47%	103-226	200.76	17	
AsnC	Sint. de aminoácidos y metab. de glicerol	-	100%	120-254	169.84	13	
AraC	Virulencia, transposición, metab. de azúcares	+	60.8%	106-529	279.49	57	
LuxI	Biosint. germinación, metab. del glicerol	+	100%	74-901	230.02	37	
CRP	Respuestas globales	+	100%	206-257	232.95	23	
ArsR	Resistencia a arsénico. Sist. de flujo de cadmio	-	Zona media	104-122	114.88	9	
IcIR	Síntesis de aminoácidos	-	100%	247-306	271.87	8	
TrpR	Síntesis de aminoácidos	-	100%	107-109	107.75	4	
BirA	Síntesis de biotina	-	100%	240-325	297.00	4	
TWO	Forofilitación y respuesta adaptativa	+	-----	-----	-----	---	
EBP	Asimilación de nitrógeno, sint. de alginatos	+	100%	98-692	445.04	49	
SIG F	Retocamiento específico del promotor	-	100%	60-708	352.61	69	

Tabla 2. Principales funciones asociadas a cada familia de reguladores transcripcionales en procariontes. También se incluyen los factores sigma.

FAM	Familia de proteínas reguladoras.
PRINCIPALES FUNCIONES	Asociadas a los genes regulados.
REG	Tipo de regulador: Represores (-), Activadores (+) o Duales (+/-)
NH2-	Indica el % de proteínas que tienen el HTH en el primer 20% NH2-term. Por ejemplo: el 94% de las proteínas de la familia LysR tienen el HTH en el primer 20% NH2-term.
COOH-	Indica el % de proteínas que tienen el HTH en el primer 30% COOH-term. Por ejemplo: el 100% de las proteínas de la familia CRP tienen el HTH en el último 30% COOH-term.
RANGO	Intervalo de tamaño entre la proteína funcional de longitud más corta y la más larga.
MEDIA	Longitud promedio de los miembros de la familia.
TAM	Número de proteínas que componen a la familia.
SIG-F	Factores sigma de las familias $\sigma 70$ y $\sigma 54$.

APENDICE I

Proteínas reguladoras de promotores $\sigma 70$ agrupadas de acuerdo a su función y familia.

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>Rhodospseudomonas palustris</i>	AADR	+	CRP
<i>Pseudomonas aureoginosa</i>	ANR	+	CRP
<i>Bradyrhizobium japonicum</i>	FIXK	+	CRP
<i>Rhizobium leguminosarum</i>	FX24	+	CRP
<i>Actinobacillus pleuropneumoniae</i> HLYX		+	CRP
<i>Anabaena sp.</i>	NTCA	+	CRP
<i>Synechococcus sp</i>	NTCA	+	CRP
<i>Synechocystis sp.</i>	NTCA	+	CRP
<i>Pseudomonas stutzeri</i>	FNRA	+	CRP
<i>Bacillus subtilis</i>	FNR	+	CRP
<i>E. coli</i>	CRP, FNR	+/-	CRP
<i>S. typhimurium</i>	FNR	+/-	CRP
<i>Haemophilus influenzae</i>	FNR	+/-	CRP
<i>Xanthomonas campestris</i>	CLP	+/-	CRP
<i>Azorhizobium caulinodans</i>	FIXK	+/-	CRP
<i>Rhizobium meliloti</i>	FIXK	+/-	CRP
<i>S. typhimurium</i>	CRP	N/D	CRP
<i>Haemophilus influenzae</i>	CRP	N/D	CRP
<i>Lactococcus lactis</i>	FLP	N/D	CRP
<i>Bordetella pertussis</i>	BTR	N/D	CRP
<i>Synechococcus sp</i>	CYSR	N/D	CRP
<i>Shewanella putrefaciens</i>	ETRA	N/D	CRP
<i>E. coli</i>	CFAD, ENVY, FAPR, MELR, RHAR, RHAS, URER, RNS, AGGR, CSVR, PERA	+	ARAC
<i>Pseudomonas aureoginosa</i>	PCHR, EXSA, MMSR	+	ARAC
<i>Pseudomonas putida</i>	XVLS, XYS1, XYS3, XYS4, XYS2	+	ARAC
<i>Bacillus subtilis</i>	ADAA	+	ARAC
<i>S. typhimurium</i>	INVF, RHAS	+	ARAC
<i>Burkholderia solanacearum</i>	HFPB	+	ARAC
<i>Yersinia enterocolitica</i>	LCRF	+	ARAC
<i>Pedococcus pentosaceus</i>	PAFF	+	ARAC
<i>Streptococcus mutans</i>	MSMR	+	ARAC
<i>Vibrio cholerae</i>	TCPN	+	ARAC
<i>Proteus mirabilis</i>	URER	+	ARAC
<i>Shigella dysenteriae</i>	VIRF	+	ARAC
<i>Yersinia pestis</i>	VIRF	+	ARAC
<i>Providencia stuartii</i>	AARP	+	AFAC
<i>Erwinia chrysanthemi</i>	ARAC	+/-	ARAC
<i>Citrobacter freundii</i>	ARAC	+/-	ARAC
<i>E. coli</i>	ARAC, MARA	+/-	ARAC
<i>S. typhimurium</i>	ARAC	+/-	ARAC
<i>E. coli</i>	APPY, CELD	-	ARAC

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>E. coli</i>	ADIY, ROB, SOXS, TETD, XYLR	N/D	ARAC
<i>Shigella flexneri</i>	MXIE	N/D	ARAC
<i>Haemophilus influenzae</i>	XYLR, YA52	N/D	ARAC
<i>Bacillus subtilis</i>	YZBC	N/D	ARAC
<i>Streptomyces antibioticus</i>	ARAL	N/D	ARAC
<i>Streptomyces lividans</i>	ARAL	N/D	ARAC
<i>Rhodococcus sp</i>	THCR	N/D	ARAC
<i>Streptomyces lavendulae</i>	YMCR	N/D	ARAC
<i>E. coli</i>	TRPR	-	TRP
<i>S. typhimurium</i>	TRPR	-	TRP
<i>Enterobacter aerogenes</i>	TRPR	-	TRP
<i>Enterobacter cloacae</i>	TRPR	-	TRP
<i>E. coli</i>	FUR	-	FUR
<i>E. coli</i>	UHPA, EVGA, SDIA, MALT,		
	RCSA, RC5B	+	LUXR
<i>S. typhimurium</i>	UHPA	+	LUXR
<i>Bacillus subtilis</i>	COM1, CMA1	+	LUXR
<i>Pseudomonas aureginosa</i>	LASR, GLPR	+	LUXR
<i>Bordetella pertussis</i>	BVGA	+	LUXR
<i>Bradyrhizobium japonicum</i>	FIXJ	+	LUXR
<i>Rhizobium meliloti</i>	FIXJ	+	LUXR
<i>Azorhizobium caulinodans</i>	FIXJ	+	LUXR
<i>Rhodobacter capsulatus</i>	DCTR	+	LUXR
<i>Agrobacterium tumefaciens</i>	TRAR	+	LUXR
<i>Agrobacterium vitis</i>	TRAR	+	LUXR
<i>Rhizobium leguminosarum</i>	FMH	+	LUXR
<i>Vibrio fischeri</i>	LUXR, LUXS	+	LUXR
<i>Erwinia amylovora</i>	RCSA	+	LUXR
<i>Erwinia stewartii</i>	RCSA	+	LUXR
<i>Klebsiella pneumoniae</i>	RCSA	+	LUXR
<i>E. coli</i>	NARP	+/-	LUXR
<i>Bacillus subtilis</i>	DEBJ	+/-	LUXR
<i>Haemophilus influenzae</i>	NARP	-	LUXR
<i>Bradyrhizobium japonicum</i>	NOCW	N/D	LUXR
<i>E. coli</i>	YHJB, FIMZ, YHIF, UVRY	N/D	LUXR
<i>Bacillus subtilis</i>	CEBE	N/D	LUXR
<i>S. typhimurium</i>	FIMZ	N/D	LUXR
<i>Pseudomonas fluorescens</i>	GACA	N/D	LUXR
<i>Aeromonas salmonicida</i>	YEXN	N/D	LUXR
<i>E. coli</i>	ICLR	-	ICLR
<i>S. typhimurium</i>	ICLR	-	ICLR
<i>Erwinia chrysanthemi</i>	KDGR	-	ICLR
<i>Bacillus subtilis</i>	YOSO	N/D	ICLR
<i>E. coli</i>	YJHI, YIAJ	N/D	ICLR
<i>Haemophilus influenzae</i>	YIAJ	N/D	ICLR

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>Streptomyces lividans</i>	TIPA	+	MERR
<i>Bacillus subtilis</i>	BLTR	+	MERR
<i>Bacillus sp.</i>	MERR	+/-	MERR
<i>Shigella flexneri</i>	MERR	+/-	MERR
<i>Staphylococcus aureus</i>	MERR	+/-	MERR
<i>Pseudomonas aureginosa</i>	MERR	+/-	MERR
<i>Serratia marcescens</i>	MERR	+/-	MERR
<i>Bacillus cereus</i>	GLNR	-	MERR
<i>Bacillus subtilis</i>	GLNR	-	MERR
<i>Streptomyces lividans</i>	MERR	-	MERR
<i>Haemophilus influenzae</i>	Y293	N/D	MERR
<i>E. coli</i>	YHDM, YEHV, SOXR	N/D	MERR
<i>Haemophilus influenzae</i>	YHDM, Y186	N/D	MERR
<i>Bacillus subtilis</i>	YYAN	N/D	MERR
<i>Bradyrhizobium japonicum</i>	NOLA	N/D	MERR
<i>Bradyrhizobium sp.</i>	NOLA	N/D	MERR
<i>E. coli</i>	ADA	+/-	ADA
<i>E. coli</i>	PAPB	+/-	PAPB
<i>E. coli</i>	BIRA	-	BIRA
<i>Bacillus subtilis</i>	BIRA	-	BIRA
<i>Rhodobacter capsulatus</i>	PETR	+	TWO
<i>Providencia rettgeri</i>	LEXA	-	LEXA
<i>Pseudomonas aureginosa</i>	LEXA	-	LEXA
<i>Pseudomonas putida</i>	LEXA	-	LEXA
<i>Bacillus subtilis</i>	YQCJ	N/D	ARSR
<i>Bacillus firmus</i>	CADC	N/D	ARSR
<i>Staphylococcus aureus</i>	CADC, CADF	N/D	ARSR
<i>Staphylococcus aureus</i>	ARSR	-	ARSR
<i>Staphylococcus xylosum</i>	ARSR	-	ARSR
<i>E. coli</i>	ARSE, ARSR	-	ARSR
<i>Synechococcus sp</i>	SMTB	-	ARSR
<i>E. coli</i>	TER1, TER2, TER3, TER4, TER5,	-	TETR
	ACRR, BETI, ENVR, TETC	-	TETR
<i>Bacillus megaterium</i>	BMSR	-	TETR
<i>Haemophilus influenzae</i>	Y893, TTK	-	TETR
<i>Staphylococcus aureus</i>	YP23	-	TETR
<i>Neisseria gonorrhoeae</i>	MTRR	-	TETR
<i>Streptomyces glaucoscens</i>	TCMR	-	TETR
<i>Bacillus subtilis</i>	YIXD	N/D	TETR
<i>Pseudomonas sp.</i>	TCBR	+	LYSR
<i>Pseudomonas putida</i>	CLCR	+	LYSR
<i>Lactococcus lactis</i>	MLER	+	LYSR

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>Salmonella dublin</i>	VRPR	+	LYSR
<i>S. typhimurium</i>	VRPR	+	LYSR
<i>Pseudomonas aureaginosa</i>	AMPR, TRP1	+	LYSR
<i>Rhodobacter capsulatus</i>	AMPR	+	LYSR
<i>Pseudomonas syringae</i>	TRP1	+	LYSR
<i>E. coli</i>	TDCA, XAPR	+	LYSR
<i>Rhizobium meliloti</i>	SYRM	+	LYSR
<i>Rhizobium leguminosarum</i>	SYRM	+	LYSR
<i>Streptomyces coelicolor</i>	MPRR	+	LYSR
<i>Haemophilus influenzae</i>	OXYR, CYSB	+	LYSR
<i>Klebsiella pneumoniae</i>	CYSB	+	LYSR
<i>Streptomyces lividans</i>	SNPR	+	LYSR
<i>Yersinia pestis</i>	AMPR	+/-	LYSR
<i>Agrobacterium thumefaciens</i>	NOCR-5, NOCR-7	+/-	LYSR
<i>Chromatium vinosum</i>	RECR	+/-	LYSR
<i>E. coli</i>	ICIA, LYSR, NHAR, CYNR, ILVY, CYSB, METR, OXYR	+/-	LYSR
<i>Citrobacter freundii</i>	AMPR	+/-	LYSR
<i>Enterobacter cloacae</i>	AMPR	+/-	LYSR
<i>Pseudomonas putida</i>	NAHR	+/-	LYSR
<i>Bradyrhizobium japonicum</i>	NOD1, NOD2	+/-	LYSR
<i>Rhizobium leguminosarum</i>	NOD1, NOD2, NOD3, NODD-1, NODD-v, NODD	+/-	LYSR
<i>Rhizobium meliloti</i>	NGD1, NOD2, NOD3	+/-	LYSR
<i>Rhizobium tropici</i>	NOD1, NOD2	+/-	LYSR
<i>Azorhizobium caulinodans</i>	NODD	+/-	LYSR
<i>Thiobacillus ferrooxidans</i>	RECR	+/-	LYSR
<i>Bacillus subtilis</i>	CLTC	+/-	LYSR
<i>Haemophilus influenzae</i>	ILVY, METR	+/-	LYSR
<i>Alcaligenes eutrophus</i>	TFDR	+/-	LYSR
<i>Bradyrhizobium elkanii</i>	NOD1, NOD2	+/-	LYSR
<i>Bradyrhizobium sp.</i>	NOD1, NOD2	+/-	LYSR
<i>Rhizobium galegae</i>	NOD1	+/-	LYSR
<i>Bradyrhizobium sp.</i>	NODD	+/-	LYSR
<i>Rhodobacter capsulatus</i>	HVRB	+/-	LYSR
<i>Agrobacterium thumefaciens</i>	GBPR, OCCR	+/-	LYSR
<i>Pseudomonas putida</i>	CATR	+/-	LYSR
<i>Klebsiella pneumoniae</i>	NAC	+/-	LYSR
<i>Streptomyces cacaoi</i>	BLAA	+/-	LYSR
<i>S. typhimurium</i>	CYSB, METR	+/-	LYSR
<i>Vibrio cholerae</i>	IRGB	+/-	LYSR
<i>Acinetobacter calcoaceticus</i>	CATM	-	LYSR
<i>Bacillus subtilis</i>	CITR	-	LYSR
<i>Rhizobium meliloti</i>	NOUR	-	LYSR
<i>E. coli</i>	GCVA, GENR	-	LYSR
<i>Bacillus subtilis</i>	YR26, YYBE, ALSR, YWBI, YWFK N/D	-	LYSR
<i>E. coli</i>	YFIE, PSSR, YDHB, YHCS, YAFJ, YEIE, YIAU, YGIP, YFEB, YHAJ, YIDZ, LRHA, LEUO, YHJC, YJIE	N/D	LYSR
<i>Alcaligenes eutrophus</i>	TFDT	N/D	LYSR

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>Klebsiella pneumoniae</i>	YNA2	N/D	LYSR
<i>Haemophilus influenzae</i>	YHCS, GCVA, Y775	N/D	LYSR
<i>S. typhimurium</i>	SINR	N/D	LYSR
<i>E. coli</i>	FRFR, FRUR, RBSR, ASCG, EBGR GALR, GALS, PURR, CSCR, LACI	-	-
	MALI, CYTR	-	LAC
<i>S. typhimurium</i>	FRUR, SCRR	-	LAC
<i>Haemophilus influenzae</i>	RBSR, GALR, PURR	-	LAC
<i>Bacillus subtilis</i>	RBSR, CCPA	-	LAC
<i>Bacillus subtilis</i>	YR45, YYAG	N/D	LAC
<i>Bacillus subtilis</i>	GRAR	+/-	LAC
<i>Bacillus subtilis</i>	DEGA	+	LAC
<i>E. coli</i>	YJGS, GNTR	N/D	LAC
<i>Vibrio alginolyticus</i>	SCRR	-	LAC
<i>Streptococcus pneumoniae</i>	MALR	-	LAC
<i>Bacillus polymyxa</i>	BNDR	N/D	LAC
<i>Klebsiella pneumoniae</i>	RBTR	-	LAC
<i>Klebsiella pneumoniae</i>	SCFR	-	LAC
<i>Pseudomonas aureoginosa</i>	NFXB	-	LAC
<i>E. coli</i>	PDHR	-	GNTR
<i>Bacillus licheniformis</i>	GNTR	-	GNTR
<i>Bacillus subtilis</i>	GNTR	-	GNTR
<i>Pseudomonas putida</i>	HJTC	-	GNTR
<i>E. coli</i>	FADR	+/-	GNTR
<i>Streptomyces lividans</i>	KORA	-	GNTR
<i>Klebsiella pneumoniae</i>	HJTC	-	GNTR
<i>Haemophilus influenzae</i>	FADR	+/-	GNTR
<i>Streptomyces ambofaciens</i>	YINI	N/D	GNTR
<i>Rhizobium meliloti</i>	MCCR	N/D	GNTR
<i>Rhodobacter sphaeroides</i>	YRDX	N/D	GNTR
<i>E. coli</i>	YIDP, YIDW, FARR, YIHL, LCTR, LLDR UXUR, PHNF, EXUR, YHCK, YHFR	N/D	GNTR
<i>S. typhimurium</i>	PTSJ	N/D	GNTR
<i>Bacillus subtilis</i>	YCBG	N/D	GNTR
<i>Haemophilus influenzae</i>	UXUR	N/D	GNTR
<i>Pyrococcus furiosus</i>	YGDH	N/D	ASNC
<i>Haemophilus influenzae</i>	ASNC	+	ASNC
<i>Pseudomonas putida</i>	SKDR	+	ASNC
<i>E. coli</i>	ASNC, LRP	+/-	ASNC
<i>Streptomyces coelicolor</i>	GYLR	+	ASNC
<i>Streptomyces griseus</i>	GYLR	+/-	ASNC
<i>Klebsiella pneumoniae</i>	LRP	+	ASNC
<i>S. typhimurium</i>	LRP	+	ASNC
<i>Serratia marcescens</i>	LRP	+	ASNC
<i>Haemophilus influenzae</i>	LRP	+	ASNC
<i>Haemophilus influenzae</i>	Y224	N/D	ASNC

ORGANISMO	PROTEINA	FUNCION	FAMILIA
<i>E. coli</i>	FUCR, GLPR, GUTR, SRLR, DEOR, AQAR	-	DEOR
<i>Lactococcus lactis</i>	LACR	-	DEOR
<i>Staphylococcus aureus</i>	LACR	-	DEOR
<i>Streptococcus mutants</i>	LACR	-	DEOR
<i>Haemophilus influenzae</i>	GLPR	-	DEOR
<i>Aerobacterium thumefaciens</i>	ACCR	-	DEOR
<i>Haemophilus influenzae</i>	FUCR	+	DEOR
<i>E. coli</i>	YIHW, YJFQ, YJHJ	N/D	DEOR
<i>Bacillus subtilis</i>	KLR	N/D	DEOR
<i>Haemophilus influenzae</i>	YGBI	N/D	DEOR
<i>Azotobacter vinelandi</i>	ANFA, NIFA, VNFA	+	EBP
<i>E. coli</i>	HYDG	+	EBP
<i>Rhizobium mellioli</i>	DCTD, NIFA	+	EBP
<i>Alicyclopes eutrophus</i>	HOXA	+	EBP
<i>Bradyrhizobium japonicum</i>	HOXA, NIFA	+	EBP
<i>Pseudomonas syringae</i>	HRPR, HRPS-s, HRPS-p	+	EBP
<i>Rhodobacter capsulatus</i>	HUPR, NIFA	+	EBP
<i>S. typhimurium</i>	HYDG	+	EBP
<i>Azospirillum brasilense</i>	NIFA	+	EBP
<i>Rhodobacter capsulatus</i>	NIFA	+	EBP
<i>Herbaspirillum seropedicae</i>	NIFA	+	EBP
<i>Klebsiella pneumoniae</i>	NIFA	+	EBP
<i>Rhizobium leguminosarum</i>	NIFA	+	EBP
<i>S. typhimurium</i>	PGTA	+	EBP
<i>Pseudomonas aureoginosa</i>	PILR	+	EBP
<i>Bacillus subtilis</i>	RCOR	+	EBP
<i>Erwinia stewartii</i>	WTSA	+	EBP
<i>Pseudomonas putida</i>	XYLR	+	EBP
<i>E. coli</i>	ATOC, ORD1	-	EBP
<i>Azospirillum brasilense</i>	NTRC	+/-	EBP
<i>Klebsiella pneumoniae</i>	NTRC	+/-	EBP
<i>Proteus vulgaris</i>	NTRC	+/-	EBP
<i>Rhizobium mellioli</i>	NTRC	+/-	EBP
<i>Rhodobacter capsulatus</i>	NTRC	+/-	EBP
<i>Azorhizobium caulinodans</i>	NTRX	+/-	EBP
<i>E. coli</i>	FIS, NTRC, TYRR	+/-	EBP
<i>E. coli</i>	YCJB, YGAA	N/D	EBP
<i>Clostridium beijerinckii</i>	STC	N/D	EBP
<i>Haemophilus influenzae</i>	TYRR	N/D	EBP
<i>Pseudomonas sp.</i>	YHYA	N/D	EBP

Nota: Los símbolos +, - y +/- representan a las proteínas activadoras, represoras y duales respectivamente.

APENDICE II

Factores sigma analizados, pertenecientes a las familias $\sigma 70$ y $\sigma 54$ de procariontes.

ORGANISMO

Escherichia coli
Salmonella typhimurium
Pseudomonas aureaginosa
Myxococcus xanthus
Bacillus subtilis
Streptomyces coelicolor
Alcaligenes eutrophus
Pseudomonas syringae
Streptomyces aureofaciens
Citrobacter freundii
Bacillus thuringiensis
Bacillus licheniformis
Agrobacterium tumefaciens
Vibrio parahaemolyticus
Buchnera aphidicola
Chlamydia trachomatis
Rickettsia prowazekii
Salmonella dublin
Shigella flexneri
Clostridium acetobutylicum
Lactococcus lactis
Stigmatella aurantiaca
Bacillus megaterium
Staphylococcus aureus

FACTOR

FlIA, FlhC, RP32, RP70, RpoE, RpoS
FlIA, RpoE, RpoS, RP70
FlIA, RP70, RpsH
CarQ, RP80, RPSB, RpsC
RpsA, RpsB, RpsD, RpsE, RpsF, RpsG, RpsH, RpsK, SigX
HrdA, HrdB, HrdC, HrdD, RpoE, RpoF, RpsW
CnhI
HrL
RpoX, RpsA
RP32
RP28, RP35
RpsH, RpsF
RpsA
LAFS
RP70
RP70, RpsA
RP70
RpoS
RpoS
RpsA, RpsG
RpsA
RpsB
RpsF
RpsP

FACTORES SIGMA DE LA FAMILIA $\sigma 54$

ORGANISMO

E. coli
S. typhimurium
B. subtilis
Alcaligenes eutrophus
Pseudomonas putida
Rhizobium meliloti
Klebsiella pneumoniae
Bradyrhizobium japonicum
Acinetobacter calcoaceticus
Azorhizobium caulinodans
Azotobacter vinelandii
Rhodobacter capsulatus
Rhodobacter sphaeroides
Thiobacillus ferrooxidans

FACTOR

RP54
RP54
RP54
RP54
RP54
RP55, RP54
RP54
RP54, RP55
RP54
RP54
RP54
RP54
RP54
RP54
RP54

APENDICE III

A continuación se listan todas las proteínas analizadas en este trabajo, de acuerdo al tipo de motivo estructural que presentan. En algunos casos, se adiciona el nombre de la familia en la cual han sido clasificadas previamente.

III. PROTEINAS CON HOMEODOMINIOS

ORGANISMO	PROTEINA	FAMILIA
<i>Drosophila melanogaster</i>	HMCU, HMES, HMH2, HMN1, HMOC, HMRO	
<i>Ustilago maydis</i>	HMT1, HMZ1, HMZ2, PRO, ZFH1, ZFH2	
<i>Homo sapiens</i>	B1, B2, B3, B4, B5, B6, B7	
<i>Mus musculus</i>	HMPH, HX11	
<i>Shizophyllum commune</i>	MHCX, TES1	
<i>Saccharomyces cerevisiae</i>	MAY1, MAY3, MAY4, MAZ3, MAZ4	
<i>Gallus gallus</i>	PHO2, YOX1	
<i>Zea mays</i>	DEF1, HMPH	
<i>Caenorhabditis elegans</i>	HKN1	
<i>Dugesia tigrina</i>	HM23	
<i>Drosophila virilis</i>	HMH1, HMH2	
<i>Drosophila melanogaster</i>	HME0	
<i>Mus musculus</i>	HMAA, HMAN, HMFT, HMSC	HOMEODOMINIO TIPO-ANTP
<i>Homo sapiens</i>	HXA5	HOMEODOMINIO TIPO-ANTP
<i>Mus musculus</i>	HXA5, HXB3, HXB5, HXB6, HXB7, HXC5, HXC6, HXD3	HOMEODOMINIO TIPO-ANTP
	HXB3, HXB6, HXB7, HXB8, HXC6, HXC8, HXD8, HXA7, HXB5	HOMEODOMINIO TIPO-ANTP
<i>Coturnix coturnix japonica</i>	HXA7	HOMEODOMINIO TIPO-ANTP
<i>Xenopus laevis</i>	HXA7, HXB7, HXC6	HOMEODOMINIO TIPO-ANTP
<i>Brachydanio rerio</i>	HXB5, HXB8	HOMEODOMINIO TIPO-ANTP
<i>Notophthalmus viridescens</i>	HXC6	HOMEODOMINIO TIPO-ANTP
<i>Gallus gallus</i>	HXD8	HOMEODOMINIO TIPO-ANTP
<i>Caenorhabditis elegans</i>	LI39	HOMEODOMINIO TIPO-ANTP
<i>Drosophila melanogaster</i>	HMLA	SUBFAM. TIPO ANTP-LABIAL
<i>Gallus gallus</i>	HXB1	SUBFAM. TIPO ANTP-LABIAL
<i>Homo sapiens</i>	HXB1	SUBFAM. TIPO ANTP-LABIAL
<i>Mus musculus</i>	HXA1, HXB1, HXD1	SUBFAM. TIPO ANTP-DEFORMADO
<i>Gallus gallus</i>	HXA4, HXB4, HXD4	SUBFAM. TIPO ANTP-DEFORMADO
<i>Mus musculus</i>	HXA4, HXB4, HXC4, HXD4	SUBFAM. TIPO ANTP-DEFORMADO
<i>Homo sapiens</i>	HXA4, HXB4, HXC4, HXD4	SUBFAM. TIPO ANTP-DEFORMADO
<i>Drosophila melanogaster</i>	HME0	SUBFAM. TIPO ANTP-DEFORMADO
<i>Xenopus laevis</i>	HXB4	SUBFAM. TIPO ANTP-DEFORMADO
<i>Drosophila melanogaster</i>	HMPB	SUBFAM. TIPO ANTP-PROBOSCIPEDIA
<i>Mus musculus</i>	HXA2	SUBFAM. TIPO ANTP-PROBOSCIPEDIA
<i>Gallus gallus</i>	HXA2	SUBFAM. TIPO ANTP-PROBOSCIPEDIA
<i>Rattus norvegicus</i>	HXA2	SUBFAM. TIPO ANTP-PROBOSCIPEDIA
<i>Homo sapiens</i>	HXB2	SUBFAM. TIPO ANTP-PROBOSCIPEDIA
<i>Caenorhabditis elegans</i>	HML16	HOMEODOMINIO ENGRAILED
<i>Brachydanio rerio</i>	HME1, HME2, HME3	HOMEODOMINIO ENGRAILED

ORGANISMO

Gallus gallus
Homo sapiens
Mus musculus
Artemia salinfranciscana
Bombyx mori
Drosophila virilis
Drosophila melanogaster
Brachydanio rerio
Gallus gallus
Colurnix colurnix japonica
Homo sapiens
Mus musculus
Drosophila melanogaster
Mus musculus
Xenopus laevis
Gallus gallus
Brachydanio rerio
Dugesia japonica
Caenorhabditis elegans
Drosophila melanogaster
Homo sapiens
Gallus gallus
Xenopus laevis
Mus musculus
Notoponthalmus visidescens
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Mus musculus
Homo sapiens
Mus musculus
Mus musculus
Drosophila melanogaster
Mus musculus
Homo sapiens
Xenopus laevis
Rattus norvegicus
Brachydanio rerio
Homo sapiens
Drosophila melanogaster
Caenorhabditis elegans
Brachydanio rerio
Drosophila melanogaster
Mus musculus
Xenopus laevis
Caenorhabditis elegans
Caenorhabditis vulgarensis
Homo sapiens
Mus musculus
Homo sapiens

PROTEINA

HME1, HME2
HME1, HME2
HME1, HME2
HMEN
HMEN, HMIN
HMEN
HMIN
HMMC, HMMD
HMX1
HMX1
HMX1, HMX2
HMX1, HMX2
GSBD, GSBP, HMBC, HMPR
HMGC, PAX3
HMIX
HPR1
PAX6
POU1
EGL5
HMAB
HXAA, HXD9, HXDA
HXAB, HXDB, HXDC
HXB9
HXC9, HXCA, HXD9, HXDA, HXDB, HXDC
HXDB
HMB1
HMB2
HMB3
OC11
OC3A, OC3B
OCT3
HMC1, HMC4, HMCA
PDM1, PDM2
BRN1, OC3N, OCT6
OC3N, OCT6
POU1, POU2, POU3
BRN4
POU1
HMPR
IPOU
UN86
POUC
APTE
LH2, ISL1, LIM1
LIM1
MEC3
MEC3
OTX1
OTX1
OTX2

FAMILIA

HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX ENGRAILED
HOMEBOX TIPO MSH
HOMEBOX TIPO MSH
HOMEBOX TIPO MSH
HOMEBOX TIPO MSH
HOMEBOX TIPO MSH
HOMEBOX TIPO MSH
HOMEBOX TIPO PAIRED
HOMEBOX TIPO PAIRED
HOMEBOX TIPO PAIRED
HOMEBOX TIPO PAIRED
HOMEBOX TIPO PAIRED
HOMEBOX TIPO PAIRED
FACTORES DE TRANSCRIPCIÓN POU
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
FAMILY ABD-B
ATBH-2 Y ATBH-3
ATBH-1 Y ATBH-3
ATBH-1 Y ATBH-2
PERTENECE A LA CLASE 2-POU
PERTENECE A LA CLASE 5-POU
PERTENECE A LA CLASE 5-POU
HOMEBOX TIPO CAUDAL
CLASE-2 POU
CLASE-3 POU
CLASE-3 POU
CLASE-3 POU
CLASE-3 POU
CLASE-3 POU
CLASE-4 POU
CLASE-4 POU
CLASE-4 POU
CLASE-4 POU
CLASE-6 POU
HOMEBOX
DOMINIO LIM
DOMINIO LIM
DOMINIO LIM
DOMINIO LIM
SIMILAR A OTD DE *Drosophila* y OTX2
SIMILAR A OTD DE *Drosophila* y OTX2
SIMILAR A OTD DE *Drosophila* y OTX1

ORGANISMO

Mus musculus
Homo sapiens
Mus musculus
Drosophila melanogaster
Caenorhabditis elegans
Mus musculus

Drosophila melanogaster

Saccharomyces cerevisiae
Drosophila melanogaster
Halocynthia roretzi
Homo sapiens
Rattus norvegicus
Mus musculus
Rattus norvegicus
Caenorhabditis elegans

ORGANISMO

Drosophila melanogaster
Mus musculus
Homo sapiens
Rattus norvegicus
Mesocricetus auratus
Bos taurus
Meleagris gallopavo
Oncorhynchus mykiss
Sus scrofa
Gallus gallus
Xenopus laevis
Caenorhabditis elegans
Drosophila melanogaster
Saccharomyces cerevisiae
Rattus norvegicus

PROTEINA

OTX2
 PAX6
 PAX6
 E2D
 MAB5
 EVX1

HMEV

MTA0, MTA1
 ZFH2
 HOX1
 GCP
 TTF1
 MOX1, MOX2
 MOX2
 YM61

PROTEINA

CF1A
 HMP1, HNFA, HNF6, OCT1
 HMP1, HNFA, HNF6, OCT1, OCT2
 HMP1, HNFA, HNF6
 HMCS, LMX1
 HMP1
 HMP1
 HMP1
 HNF6
 OCT1
 OCT1
 OCT1
 PAL1
 HMEN
 MAT2
 OCT6

FAMILIA

SIMILAR A OTD DE *Drosophila* y OTX1
 DOMINIO PAIRED BOX
 DOMINIO PAIRED BOX
 FAMILIA PBX
 SIMILAR A OTROS HOMEODOMINIO TIPO-ANTP
 SUBFAM. QUE INCLUYE EVE de
Drosophila, EVX-1 y EVX-2 de *Raton*, y
 XHOX-3 de *Xenopus*
 SUBFAM. QUE INCLUYE EVE de
Drosophila, EVX-1 y EVX-2 de *Raton*, y
 XHOX-3 de *Xenopus*
 Homologo a homeoproteinas
 Con tres homeobox
 Similar a H2 de *Drosophila*
 Similar a CUT de *Drosophila*
 Similar a NK-2 de *Drosophila*
 SIMILAR A OTROS HOMEODOMINIO
 SIMILAR A OTROS HOMEODOMINIO
 PROSPERO

FUNCION

ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 ACTIVADOR
 REPRESOR
 REPRESOR
 REPRESOR

IIIB. PROTEINAS CON DEDOS DE ZINC

En este trabajo se analizaron proteínas con dedos de zinc repetidos en tandem de dos dedos.

PROTEINAS CON 2 DEDOS DE ZINC Y QUE SE PRESENTAN EN ESTE TRABAJO.

ORGANISMO	PROTEINA
<i>Saccharomyces cerevisiae</i>	ADR1, MIG1, MSN2, MSN4, NOT4, RGM1, YEL8
<i>Homo sapiens</i>	ANDR, ARP1, COTF, EAR1, EAR2, ERR1, ERR2, ESTR, GAT1, GAT2, GAT3, GCRA, GCR, HNF4, MCR, NAK1, PPAR, PPAS, PPAT, PRGR, ROR1, ROR2, ROR3, RRA1, RRB2, RRG1, RRG2, RRXA, RRXB, RXXC, THA1, THA2, THB1, THB2, TR2, VDR, ZEP2, ZN46
<i>Mus musculus</i>	ANDR, ELP, ESTR, GAT1, GAT3, GAT4, GCR, NUR1, NUR7, PPAR, PPAS, PPAT, PPAU, PRGR, RRA1, RRA2, RRB1, RRB2, RRB3, RRG1, RRG2, RRXA, RRXB, RXXG, THA1, THB1, THB2
<i>Drosophila melanogaster</i>	7UP1, 7UP2, DHR3, E75A, E75C, EGR, EGON, FTF1, KNIR, KNRL, TLL, TTK, USP
<i>Xenopus laevis</i>	GAT2, GAT3, GATA, GATB, NGFI, PPAR, PPAS, PPAT, RRG2, THAA, THAB, THB1, THB5, THB6, THB7
<i>Bos taurus</i>	AD4B
<i>Aspergillus niger</i>	CREA
<i>Emericella nidulans</i>	CREA
<i>Caenorhabditis elegans</i>	ELT1
<i>Avian erythroblastosis virus</i>	ERBA
<i>Salmo irideus</i>	ESTR
<i>Manduca sexta</i>	MHR3
<i>Oryctolagus cuniculus</i>	FRQR
<i>Notophthalmus viridescens</i>	RRA, RRD
<i>Rana catesbeiana</i>	THA
<i>Ustilago maydis</i>	URB1
<i>Gallus gallus</i>	ESTR, GAT1, GAT2, GAT3, PRGR, RRB1, RRB2, RRXA, THA, THB
<i>Rattus norvegicus</i>	ANDR, ESTR, GCR, HNF4, MCR, NGFI, PPAR, RNR1, RRXA, THA1, THA2, THB1, THB2, VDR

III.C. PROTEINAS CON HELICE-LOOP-HELICE (HLH)

ORGANISMO	PROTEINA	FAMILIA
<i>Mus musculus</i>	ASH1	MYC
<i>Rattus norvegicus</i>	ASH1, ASH2	MYC
<i>Xenopus laevis</i>	ASH1	MYC
<i>Homo sapiens</i>	ID1, ID2, ID3, ID4	Subfam. ID
<i>Mus musculus</i>	ID1, ID2, ID3, ID4	Subfam. ID
<i>Rattus norvegicus</i>	ID1, ID2, ID3	Subfam. ID
<i>Drosophila melanogaster</i>	ATO	Subfam. Atonal
<i>Mus musculus</i>	MTH1, MTH2	Subfam. Atonal
<i>Drosophila melanogaster</i>	AST3, AST4, AST5, AST8, DEI, TWST, MYOD	
<i>Homo sapiens</i>	HEN1, HEN2, LYL1, MAD, MAX, MYCL, MYCN, MYCN, MYC, MYF3, MYF4, MYF5, MYF6, SCL	
<i>Xenopus laevis</i>	MF25, MYC2, MYCL, MYCN, MYCN, MYOD, MYF5, TWST	
<i>Mus musculus</i>	HEN1, LYL1, MYCN, MYC, MYF5, MYOD, MYOG, TWST, MYF6, MYCL, SCL	
<i>Rattus norvegicus</i>	MYCS, MYC, MYF6, MYOD, MYOG	
<i>Gallus gallus</i>	MYCN, MYC, MYF5, MYF6, MYOD, MYOG, SCL	
<i>Marmota monax</i>	MYC2, MYC1	
<i>Serinus canaria</i>	MYCN	
<i>Avian myelocytomatosis virus cmii</i>	MYC	
<i>Avian myelocytomatosis virus mc29</i>	MYC	
<i>Avian myelocytomatosis virus hb1</i>	MYC	
<i>Avian retrovirus mh2e21</i>	MYC	
<i>Avian retrovirus ok10</i>	MYC	
<i>Callithrix jacchus</i>	MYC	
<i>Felis catus</i>	MYC	
<i>Feline leukemia virus</i>	MYC	
<i>Feline leukemia provirus fit</i>	MYC	
<i>Hylobates lar</i>	MYC	
<i>Pan troglodytes</i>	MYC	
<i>Coturnix coturnix japonica</i>	MYOG	
<i>Lytechinus variegatus</i>	SUM1	
<i>Caenorhabditis elegans</i>	MYOD, YLB7, YMH7	
<i>Schizosaccharomyces pombe</i>	ESC1	
<i>Bos taurus</i>	MYF5	
<i>Coturnix coturnix japonica</i>	MYOD	
<i>Ovis aries</i>	MYOD	