

64
zey



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ZARAGOZA**

FALLA DE ORIGEN

**CARACTERIZACION DE UN GEN CON HOMOLOGIA
A ONCOGENES CELULARES TIPO MyB AISLADO
DE LA PLANTA *Craterostigma plantagineum*.**

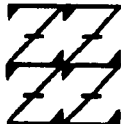
T E S I S

**QUE PARA OBTENER EL TITULO DE
QUIMICO FARMACEUTICO BIOLOGO**

P R E S E N T A

VILLEGAS SANCHEZ JOSE ANTONIO

UNAM
PES
ZARAGOZA



LO HUMANO EJE
DE NUESTRA REFLEXION

México, D.F.

1995



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Dedicatoria

A mis padres:

Ya que simplemente, sin su inmenso cariño, su completo apoyo y enorme comprensión no hubiera podido alcanzar esta meta.

A mi hermano:

A quien agradezco sus consejos, su invaluable amistad y su apoyo, todo lo cual me ha ayudado a seguir adelante.

A Adriana:

Quien en base a su amor, su amistad, su apoyo, sus críticas y su comprensión, me ha impulsado siempre, sin ninguna condición, a trabajar arduamente para conseguir lo que me propongo.

Quiero dedicar muy especialmente este trabajo a mi querida abuelita, quien a pesar de que ya falleció, se que estaria muy orgullosa de mí.

Agradecimientos

Al Doctor Gabriel Iturriga de la Fuente por su asesoría durante todo el desarrollo de mi trabajo.

Al Doctor Luc Leyns por el soporte teórico y práctico, así como las sugerencias que me brindó durante su estancia en el laboratorio.

Al Doctor Jorge Nieto por su asesoría en aspectos fundamentales de mi trabajo.

A la Doctora Alejandra Covarrubias por sus valiosos consejos.

De manera especial a Rodolfo, por ser un gran amigo con el que siempre pude contar.

Tabla de contenido

Introducción	v
Fundamentación del tema	vi
Planteamiento del problema	viii
1. Generalidades	1
Factores de transcripción	1
Antecedentes	1
Región codificante	1
Región promotora	2
Aumentadores de la transcripción	3
Unión específica al ADN	7
Estructura del dominio de unión al ADN	7
Dominio hélice-giro-hélice	8
Homeodominio	8
Dominio Myb	10
Dedos de zinc	11
Hélice anfipáticas	15
Cierres de leucina	15
Dominio hélice-vuelta-hélice	17
Dominio de transactivación	18
Familias de factores de transcripción	21
Regulación de los factores de transcripción	21
Regulación de la translocación hacia el núcleo	22
Regulación de la actividad de unión al ADN	22
Regulación de la transactivación	23
Transducción de señal	23
Factores de transcripción en plantas	25
Gen Knotted-1	26
Gen Opaque-2	26
Genes Myb	26
Gen C1	27
Gen Gl-1	27
Biología molecular de la tolerancia a la sequía	28
Sistemas de estudio	28
Embriones de semillas	28

Plantas de resurrección	28
Genes aislados	29
Factores de transcripción y sequía	31
Gen GC 19	31
Gen AtMyb2	31
2. Objetivos	33
Objetivos general	33
Objetivos particulares	33
3. Hipótesis	33
4. Material	34
Cepas	34
Enzimas de restricción y modificación	34
Reactivos y solventes	34
5. Métodos	34
1. Aislamiento de ADN	34
1.1 Preparación de plásmido a gran escala	34
1.2 Preparación de plásmido a pequeña escala	35
2. Digestión de ADN con endonucleasas de restricción	36
3. Separación de fragmentos de ADN por electroforesis	36
3.1 Geles de agarosa	36
4. Aislamiento de fragmentos de ADN a partir de geles de electroforesis	37
4.1 Electroelución en bolsas de diálisis	37
5. Clonación de fragmentos de ADN en plásmido	37
5.1 Ligación del ADN	37
5.2 Transformación de células competentes de E coli	38
6. Preparación de sonda radiactiva	39
6.1 Marcaje del ADN con oligos (Oligolabelling)	39
7. Transferencia Southern	39
7.1 Transferencia al gel	39
7.2 Hibridización	40
8. Secuenciación del ADN	41
8.1 Reacciones de secuencia	41
8.2 Geles de acrilamida	42
9. Analisis de secuencia por computadora	42

6. Resultados	43
7. Discusión	64
8. Referencias	67

Índice de figuras

Generalidades

Figura 1 Organización de un promotor de un gen transcrito por la ARN polimerasa II	4
Figura 2 Método denominado cambio en la movilidad del ADN o retardamiento en gel	5
Figura 3 Dominio Hélice-giro-hélice	9
Figura 4 Esquema de la interacción propuesta entre el dominio Myb y el ADN	12
Figura 5 Primera estructura propuesta para los dedos de zinc	13
Figura 6 Estructura más probable de los dedos de zinc de tipo cisteína-histidina	14
Figura 7 Dominio de cierre de leucina	17
Figura 8 Dímero de un proteína con dominio de unión Hélice-vuelta-hélice	18
Figura 9 Diferentes estructuras para los dominios de transactivación	20
Figura 10 Mecanismo de activación de un gen por una hormona liposoluble	24

Resultados

Mapa de restricción de la clona genómica 5	44
Secuencia del gen CpM5	46
Comparación de secuencia entre CpM5, CpM10 y CpM7	50
Diferentes marcos de lectura del gen CpM5	55
Comparación de proteínas de CpM5 y CpM10 y CpM7	59
Alineamiento del dominio de unión al ADN de CpM5	60
Dendograma	62
Comparación entre las proteínas de los genes CpM5 y Atmyb2	63

INTRODUCCION

En este trabajo, se obtuvo la secuencia de un gen aislado de la planta de resurrección *Craterostigma plantagineum*. Esta planta posee la capacidad de sobrevivir en un estado de deshidratación casi total, durante largos periodos de tiempo, y después de rehidratarse, recuperar completamente su actividad fisiológica, en un lapso de tiempo muy corto. Este gen denominado CpM5 presenta homología a nivel de secuencia de ADN con el oncogen *v-myb* (Klempner et al., 1982). Se sabe que los genes relacionados con este oncogen, están involucrados en la regulación de diferentes procesos en las plantas. Entre estos procesos se encuentran la diferenciación de células específicas (Oppenheimer, et al., 1991) la respuesta a determinadas condiciones ambientales adversas como la sequía (Urao et al., 1993) y la regulación de vías biosintéticas (Paz-Arez et al., 1987).

A pesar de que aún no se conoce la función del gen CpM5 dentro de la planta, se encontró, por el análisis de su secuencia, que presenta las características típicas de un factor de transcripción. Además, al comparar su secuencia con la de otros genes similares, se encontró que se está muy relacionado con otro gen regulador, el *AtMyb2* que se expresa durante el estrés hídrico en la planta *Arabidopsis thaliana*, de esta manera, este trabajo dará la pauta para que se pueda esclarecer la función de este gen dentro de *Craterostigma plantagineum*.

FUNDAMENTACIÓN DEL TEMA

Definida de la manera más simple, la transcripción es el paso de la información genética contenida en el ADN hacia el otro ácido nucleico el ARNm, el cual, después de la transcripción, actuará como "mensajero" para llevar dicha información hacia donde se procesará y será traducida a una proteína que tendrá una función específica, por ejemplo una enzima, o una proteína constitutiva de un tejido en especial.

La transcripción representa uno de los aspectos más importantes de la biología a nivel molecular, ya que representa un punto de control atractivo de la regulación de la expresión de los genes en tipos particulares de células o en respuesta a una señal particular (Latchman, 1990).

Se ha demostrado mediante una gran cantidad de experimentos bioquímicos y genéticos que los organismos controlan la mayoría de sus genes, casi todo el tiempo regulando la transcripción (Beardsley, 1992). De esta manera, la comprensión de el proceso de transcripción y de su regulación es indispensable para entender como funciona la expresión genética y porque un tipo de célula expresa un patrón de genes específico, en el momento preciso, y otro tipo celular expresa otro patrón de genes distinto. El discernir este fenómeno es de fundamental importancia para entender más a fondo el mecanismo de procesos tan importantes como la respuesta de los organismos a condiciones ambientales desfavorables.

Dentro de este complejo esquema de regulación, existen genes que codifican proteínas clave denominadas factores de transcripción. Estos genes son los que a fin de cuentas controlan la transcripción, ya sea interaccionando directamente con el ADN en la región promotora del gen al cual regulan o interactuando conjuntamente con otros factores de transcripción para activar el gen bajo su control (Daniell, 1990). Estos factores de transcripción pueden estar a su vez regulados por otros factores, por una señal química específica, o responder directamente a una condición extrema en particular. De esta manera, se debe hablar de una jerarquía de regulación y no de una sola clase o nivel de regulación dentro de la transcripción.

A pesar de que los factores de transcripción más estudiados han sido los que se han encontrado en animales e insectos, también existen proteínas con características similares en plantas (Brunelle & Chua, 1993).

Dentro de este tipo de genes, existe un grupo muy importante de genes reguladores denominados proto-oncogenes, estos generalmente son genes normales que se encargan de controlar algún aspecto de la división celular, sin embargo, cuando sufren alguna mutación, o una desregulación de su expresión, pueden producir cáncer el cual se manifiesta por un crecimiento anormal de la célula. Esta desregulación puede ser ocasionada por un virus al integrarse al genoma de la célula.

Existe una familia de oncogenes denominada myb (Avian myeloblastosis virus) (Klempnauer et al., 1982). Estos genes se expresan en tejido hematopoyético inmaduro y aparentemente participan en la diferenciación de ese tejido. Los genes relacionados con los genes myb parece que realizan una serie de funciones en diversos organismos a través de la activación transcripcional de diferentes genes (Sablowski et al., 1994).

PLANTEAMIENTO DEL PROBLEMA

Los genes relacionados con el oncogen Myb codifican a factores de transcripción que se encuentran involucrados en la regulación de procesos muy diferentes.

En plantas, se ha encontrado que este tipo de genes participan en procesos tan trascendentes como la regulación de la biosíntesis de antocianinas en maíz (Paz-Ares et al., 1987); la síntesis de pigmentos flavonoides en flores de *Antirrhinum* (Sablowski et al., 1994); la diferenciación de células específicas (Oppenheimer et al., 1991; Noda et al., 1994); así como la respuesta a determinadas condiciones ambientales adversas como el estrés hídrico (Urao et al., 1993). De esta manera debido a que un solo gen puede controlar a toda una serie de genes que se encargan de una función particular, es muy importante entender el mecanismo así como las condiciones bajo las cuales un factor de transcripción regula dicho proceso. Una vez que se logra esto, entonces es posible pensar en una aplicación o en el aprovechamiento de dicha función.

Para entender como funciona un factor de transcripción, sobre que genes actúa, en que condiciones se expresa, como es regulado, etc, es necesario, primeramente, caracterizar tanto a la proteína en sí como al gen que la codifica. Existen diversas estrategias para caracterizar un gen, sin embargo, cuando ya está aislado, una de las estrategias más directas consiste en obtener su secuencia nucleotídica. Mediante esta información es posible establecer definitivamente la similitud de dicho gen con otros genes ya conocidos, permitiendo en una primera instancia, proporcionar una idea de la posible función del factor de transcripción codificado por el gen.

En el caso específico del gen CpM5, no se conoce su función dentro de la planta *Craterostigma plantagineum* ya que se aisló directamente por homología a nivel de ADN con una sonda que posee una región homóloga a Myb. Además, no se tiene ningún antecedente bioquímico o genético de este gen. De esta forma, lo que se pretende en este trabajo es obtener la secuencia del gen CpM5 para poder comprobar, a nivel de secuencia de ADN, que se trata de un factor de transcripción con homología a Myb, además de sugerir una posible función al comparar su secuencia con la de otros genes similares.

GENERALIDADES

Factores de transcripción

Antecedentes

Para poder entender el proceso de transcripción en organismos eucariotes, y la función de los factores de transcripción, es necesario primero tener una idea clara de la estructura de los genes en estos organismos.

En general, un gen está constituido por dos regiones principales. Una de ellas es la región codificante, la cual se traducirá posteriormente a una proteína; la otra región, la región promotora se encuentra orientada hacia el extremo 5' del gen y contiene la información necesaria tanto para iniciar como para regular la transcripción de dicho gen.

Región codificante

En un gen procarionte, casi invariablemente, se obtiene un ARN mensajero funcional a partir de la región codificante, sin embargo, en organismos eucariotes, el transcrito inicial debe de sufrir una serie de modificaciones en el núcleo, a fin de obtener el ARNm funcional. Esto se debe a que la región codificante de un gen eucariote contiene secuencias que no se traducen, estas secuencias denominadas intrones se encuentran intercaladas entre las secuencias que sí se traducen, llamadas exones y la unión de estos, después del procesamiento del precursor del ARNm, es la secuencia que se traduce a una proteína. Los intrones se presentan mayoritariamente en genes de plantas y animales y en los virus que los infectan. De manera general son más largos que los exones pudiendo medir desde 100 hasta 10000 pares de bases. Los exones, en contraste, pueden medir de 50-300 pares de bases.

Al eliminarse los intrones de un gen eucariote, se deben realizar "cortes" muy precisos, ya que un error en una sola base puede ocasionar un cambio en el marco de lectura y por consecuencia una proteína no funcional. La especificidad de estos sitios de corte, está dado por las secuencias que se encuentran en la unión exón-intrón. Así, se sabe que el par de nucleótidos guanina y uracilo (timina en el ADN), se encuentran casi siempre, en el extremo 5' del intrón, seguido de una secuencia consenso:

A/G A G U

En donde: A = Adenina
 G = Guanina
 U = Uracilo

mientras que en el extremo 3' se encuentra el dinucleótido adenina, guanina, precedido al menos en mamíferos, por una región rica en pirimidinas (Darnell et al., 1990).

Región promotora

Dentro de la región promotora, existen secuencias conservadas las cuales especifican los sitios que la ARN polimerasa y otras proteínas, reconocen para llevar a cabo la activación de la transcripción del gen. Una de las secuencias más importantes es la denominada caja TATA o caja de Hogness, que consiste en la secuencia consenso:



En donde T = timina y A = adenina y los números denotan la frecuencia de las bases en cada posición.

Esta región generalmente se encuentra a 25-35 bases del sitio de iniciación de la síntesis del ARNm, en el extremo 5' del gen. Se ha demostrado que la sustitución de una sola base en esta secuencia produce un drástico efecto negativo en la transcripción del gen. En una de estas sustituciones, se invirtió simplemente el orden de un par de bases A-T, demostrándose que la simple composición de bases no resulta suficiente para lograr la función de la secuencia. Además, mutaciones en la región que se encuentra entre la caja TATA y el sitio de inicio no afectan significativamente el nivel de la transcripción. Se ha propuesto que la función de esta caja, consiste en permitir que la ARN polimerasa reconozca el sitio exacto de iniciación de la transcripción.

Se han encontrado diversos ejemplos de genes que no tienen una caja TATA, pero que en cambio, poseen una región rica en guaninas y citosinas que generalmente se encuentra en las primeras 200 bases hacia el extremo 5' del sitio de inicio de la síntesis del ARNm con una longitud entre 20-50 bases, y que puede presentar una función similar a la de dicha caja.

El análisis de una gran cantidad de genes de animales y plantas, ha revelado que existen además dos secuencias conservadas en la región proximal, las cuales se presentan con una frecuencia aproximada del 10-15 %. Estas secuencias, o cajas denominadas CAT y CG, poseen, respectivamente, las secuencias consenso.

CCAAT y CCCCCG

Estas cajas, se localizan generalmente a 60-120 bases del sitio de inicio de la transcripción y producen también un decremento en el nivel de transcripción cuando sufren una mutación en alguna de sus bases, lo cual indica que estas secuencias, al igual que la caja

TATA, poseen un importante papel regulatorio. Sin embargo, a diferencia de la caja TATA, se ha observado que estas cajas son capaces de funcionar en cualquier orientación. De esta manera, se ha propuesto que la función de estas secuencias consiste en controlar la unión inicial de la ARN polimerasa al iniciar la transcripción.

Ahora bien, la ARN polimerasa además de estas secuencias, necesita de la presencia de otras proteínas, denominadas factores generales de transcripción, para poder formar un complejo de preiniciación de la transcripción. La formación de este complejo es un proceso secuencial, en donde primeramente, uno de dichos factores, el $TF_{II}D$, el cual se puede unir por sí mismo a la caja TATA, forma un complejo con el ADN. Esta unión provoca un plegamiento del ADN el cual puede tener como función el "acercamiento" de proteínas regulatorias unidas a los elementos proximales del promotor (Travers et al., 1993). Posteriormente otro factor el $TF_{II}B$ se une a la polimerasa y solo después de esto, la enzima se une al ADN. Una vez que se realiza la interacción entre el ADN y la polimerasa, el factor $TF_{II}B$ actúa como una ATPasa, liberando energía a partir de la hidrólisis de una molécula de ATP. Aún no se sabe si la energía liberada se utiliza para modificar la conformación del ADN o del complejo de proteínas, pero sí se sabe que es un paso necesario para que la ARN polimerasa comience la transcripción en presencia de ribonucleótidos y de otro factor más, el $TF_{II}E$. Una vez que se llevó a cabo el inicio de la transcripción, se necesita un último factor, el $TF_{II}S$ denominado factor de elongación, el cual tiene como función evitar que la polimerasa termine prematuramente la cadena de ARN, antes de alcanzar el sitio de terminación (Lewin, 1994).

Aumentadores de la transcripción

En algunos casos, el nivel de transcripción de un gen dado, se ve aumentada en enorme medida por la presencia de secuencias denominadas aumentadores (enhancers). Estas secuencias no necesariamente deben de conservar una posición fija para llevar a cabo su función, ya que pueden localizarse tanto en el extremo 5' del gen, como en el extremo 3' y a diferente distancia del gen. Los aumentadores pueden potenciar la transcripción estando tan cerca como a 100 pares de bases y tan lejos como a 5000 pares de bases con respecto al gen. Además estas secuencias pueden funcionar en cualquier orientación y no necesariamente actúan sobre un solo gen, ya que pueden estimular a cualquier gen que se encuentre cercano a ellos.

El principal método que se ha utilizado para identificar tanto las secuencias reguladoras, como los aumentadores, consiste en mutagenizar o deletar progresivamente regiones de ADN en la región promotora, para posteriormente, en un sistema de transcripción, ya sea *in vitro* o *in vivo*, medir el efecto que dichos cambios provocan sobre

la transcripción del gen que se está ensayando. Este procedimiento permite identificar las secuencias o elementos, que ejercen un efecto sobre la transcripción, el cual puede ser tanto positivo como negativo. Estos elementos, resultan ser los sitios de unión de proteínas regulatorias denominadas factores de transcripción, los cuales pueden ser activadores o represores (Aitchison, 1988) de este proceso (figura 1).

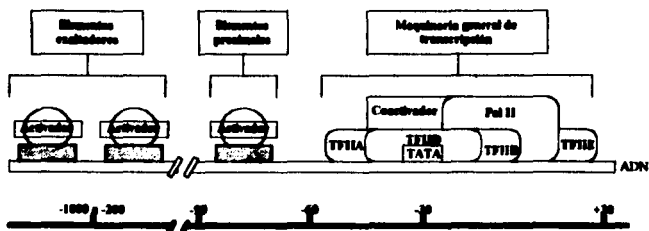


Figura 1. Organización del promotor de un gen transcrito por la ARN polimerasa II (Tomado de Travers, 1993)

Existen dos procedimientos principales para identificar a los factores de transcripción. Uno de ellos es el ensayo denominado cambio en la movilidad del ADN, en el cual un fragmento de ADN marcado, que contiene un posible elemento regulatorio, se incuba con extractos de proteínas celulares. Posteriormente, mediante una electroforesis en un gel no desnaturalizante de poliacrilamida se separan el complejo ADN-proteína, del ADN sin proteína unida. De esta manera, dicho complejo al migrar más lento, aparece como una banda retardada en la autoradiografía del gel (figura 2).

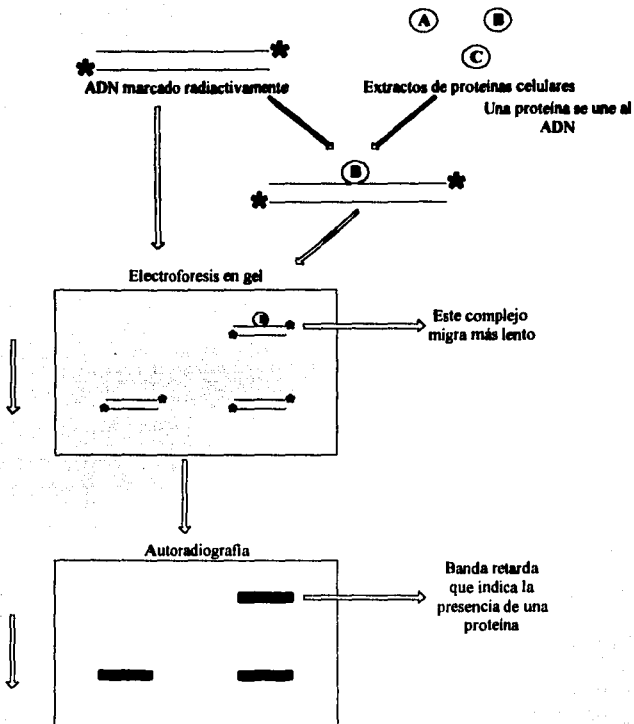


Figura 2. Método denominado cambio en la movilidad del ADN o retardamiento en gel (Tomado de Latchman, 1992)

El otro método utilizado para identificar factores de transcripción es el ensayo de la huella digital de la DNasa I. En esta técnica, también se incuban extractos de

proteínas con el ADN, pero en este caso solo se marca el final de una de las cadenas del ácido nucleico. Una vez que la proteína se une al ADN, este se digiere con pequeñas cantidades de la enzima DNasa I. Después de este tratamiento, los fragmentos obtenidos se separan por electroforesis, obteniéndose un patrón de bandas definido. Sin embargo, en la zona en la que se encuentra unida la proteína, se presenta una zona sin bandas en el gel, ya que la presencia de la proteína impide la digestión del ADN por la DNasa I (Darnell et al., 1990, Lewin, 1994).

Los métodos mencionados anteriormente, permitieron aislar una gran cantidad de secuencias y proteínas regulatorias relacionadas, que no se encontraban de manera regular en todos los promotores, sino que solo estaban presentes en las regiones regulatorias de genes específicos, infiriendo que estas proteínas activaban la transcripción, mediante las secuencias a las que se unían, de genes específicos bajo una condición dada, en un momento preciso.

En esta forma, se pudo establecer que los factores de transcripción, son proteínas que resultan ser esenciales dentro del esquema de regulación de la transcripción, ya que su presencia o ausencia, determina en última instancia que un gen determinado se exprese o no.

Este control sobre la transcripción se ejerce, por una parte a través de la interacción directa entre el ADN y una región específica del factor, el dominio de unión al ADN, y por otra parte, por la interacción del dominio de transactivación de dicho factor, con otros factores de transcripción y/o con la ARN polimerasa.

A pesar de que los factores de transcripción no son los únicos elementos de control, si son muy importantes ya que pueden regular, tanto positiva como negativamente, toda una serie de genes responsables de una función particular, por ejemplo, la diferenciación de una célula específica (Oppenheimer et al., 1991), la respuesta a determinadas condiciones adversas, como la sequía en una planta (Urao et al., 1993), la regulación de una vía biosintética (Hattori et al., 1992), etc.

Durante los últimos diez años, se han encontrado una gran cantidad de genes que codifican para factores de transcripción en levaduras, insectos y mamíferos, y se ha demostrado que varias de sus características estructurales, incluyendo parte de su secuencia de aminoácidos, se ha conservado durante el curso de la evolución. Si bien los genes de los factores de transcripción más estudiados han sido los de animales e insectos, también se han descubierto varios de ellos en plantas (Katagiri & Chua, 1992; Brunelle & Chua, 1993). El hecho de que este tipo de proteínas se haya conservado en distintos organismos a lo largo de la escala filogenética, indica que juegan un papel fundamental en la organización molecular de estos. Hasta la fecha se han encontrado diversas clases de factores de

transcripción (Pabo & Sauer, 1992), los cuales presentan una serie de características comunes. Entre estas características se encuentran:

UNIÓN ESPECÍFICA AL ADN

Una de las principales características de los factores de transcripción, es que se unen a una secuencia específica de ADN. Ahora bien, para que se lleve a cabo dicha interacción, existen dos aspectos determinantes principales, uno de ellos es la secuencia nucleotídica de la región del ácido desoxirribonucleico con la que interacciona, y el otro es la conformación tridimensional tanto del ADN, como de la proteína. La conformación está determinada en primera instancia también, por la secuencia de bases en el caso del ADN y por la de aminoácidos en el caso de la proteína. Las principales interacciones involucradas en el contacto directo entre la cadena polipeptídica de la proteína y las bases expuestas que conforman la doble hélice del ADN, son los puentes de hidrógeno que forman dichas bases y las fuerzas de van der Waals. Sin embargo, estas interacciones no proporcionan la energía suficiente para mantener de manera estable el complejo formado entre la proteína y el ADN. Para lograr esta estabilidad, la energía necesaria es proporcionada por las interacciones electrostáticas entre los residuos de los aminoácidos básicos y la estructura cargada negativamente que forman los grupos fosfato y los azúcares del ADN (Travers, 1993).

Se han encontrado factores de transcripción que a pesar de que no se unen al ADN, sí pueden activar otros genes. Tal es el caso de la proteína VP1 que se encuentra involucrada en la expresión de proteínas de almacenamiento de la semilla del maíz. Mediante varios experimentos, se ha demostrado que el dominio ácido de esta proteína puede activar la transcripción de otros genes (Ptashe, 1988), se ha propuesto que, debido a que VP1 carece de un dominio obvio de unión, puede actuar sobre otras proteínas que se unen al ADN (Brunelle et al., 1993). Sin embargo la mayoría de las proteínas regulatorias hasta ahora descritas interactúan con el ADN.

ESTRUCTURA DEL DOMINIO DE UNIÓN AL ADN

Generalmente, el dominio de unión al ADN está compuesto por aminoácidos que contienen residuos básicos como la arginina, lisina, histidina, etc. esto como ya se mencionó anteriormente, permite que al interaccionar estos grupos con el ADN que posee una carga neta negativa, se logre una estabilidad del complejo proteína-ADN.

La mayoría de las secuencias de los factores de transcripción que se han descrito hasta ahora, poseen dominios de unión al ADN que pueden agruparse en tres familias principales:

DOMINIO HELICE-GIRO-HELICE

El dominio de unión al ADN denominado Hélice-giro-hélice fue la primera estructura de este tipo que se descubrió, por lo tanto, es una de las que se ha estudiado más arduamente. Se identificó por primera vez en proteínas regulatorias de organismos procariotes las cuales forman dímeros para unirse al ADN, estos primeros ejemplos fueron las proteínas regulatorias del fago λ denominadas CRO y C1 y la proteína activadora de *E. coli* llamada CAP (Johnson and McKnight, 1989).

Gracias a estudios cristalográficos de rayos x, tanto de las proteínas aisladas como de los complejos proteína-ADN, se estableció que este dominio está constituido por dos α -hélices relativamente cortas, que se encuentran separadas por un giro de tipo β (figura 3). Este giro en la estructura es posible debido a la presencia de un residuo de glicina entre las dos hélices, este aminoácido actúa como un eje flexible que permite que la cadena polipeptídica se doble entre las dos hélices de manera que puedan establecer contactos hidrofóbicos entre ellas (Travers, 1993).

La hélice que se localiza más cercana al extremo carboxilo-terminal, denominada hélice 3 se encuentra orientada de manera antiparalela a la otra hélice, llamada hélice 2. La hélice 3 interactúa directamente, mediante algunos de sus aminoácidos, presumiblemente glicinas (Schleif, 1988), con las bases expuestas en el surco mayor del ADN. Por esta razón se denomina también hélice de reconocimiento, mientras que la hélice 2 se sitúa a través del surco mayor interactuando inespecíficamente con el ADN. De esta manera, la hélice de reconocimiento es la que controla la especificidad de unión de este dominio.

Al estudiar varias proteínas que presentaban este dominio, se observó que a pesar de que la geometría del dominio estaba altamente conservada, existía una variabilidad en la secuencia de aminoácidos. Sin embargo, se observó también, que existían una serie de aminoácidos que se conservaban en ciertas posiciones de la cadena polipeptídica. De manera que en la hélice 2, el cuarto residuo posee una cadena lateral hidrofóbica, mientras que el quinto residuo generalmente es alanina o glicina. También el giro de tipo β se inicia con una glicina e invariablemente es seguido por un residuo que posee una cadena hidrofóbica, además, los residuos cuatro y siete que constituyen la hélice 3 están ocupados por aminoácidos básicos (Johnson and McKnight, 1989).

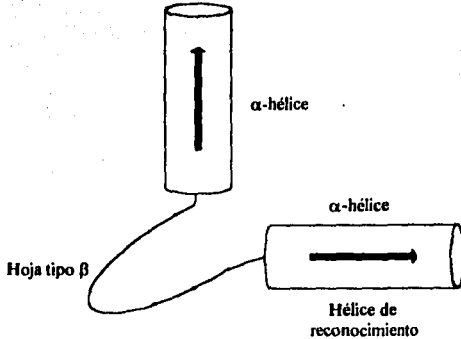


Figura 3. Dominio Hélice-giro-hélice (Tomado de Latchman, 1990)

Homeodominio

La conservación de estas posiciones dentro de la secuencia de aminoácidos del dominio hélice-giro-hélice en procariotes, permitió que se pudieran reconocer proteínas similares en organismos eucariotes. De esta manera, este dominio se identificó en una familia de proteínas de *Drosophila* que controlan una serie de etapas clave dentro de la diferenciación temprana de este organismo.

Estas proteínas presentan una región muy conservada de aminoácidos, alrededor de 60, los cuales fueron denominados Homeodominio. Al realizar estudios estructurales del mismo, se encontró que estaba constituido por 4 α -hélices, de las cuales, dos de ellas, las hélices 2 y 3 estaban separadas un giro de tipo β , presentando una estructura muy similar a la del dominio hélice-giro-hélice encontrado en procariotes (Pabo & Sauer, 1992).

Esta similitud sugería que el homeodominio interaccionaba con el ADN de la misma manera, siendo la secuencia de aminoácidos dentro de la hélice de reconocimiento la que determinaba el sitio de unión. De tal manera, que si dicha secuencia variaba, también sería diferente el sitio de unión. Para comprobar esto, se realizaron experimentos con las

proteínas de *Drosophila* en los cuales se intercambiaron estas hélices, dando como resultado que efectivamente una proteína reconocía el sitio de unión de la otra, ya que poseía su hélice de reconocimiento.

Continuando con estos experimentos, al intercambiar residuos de aminoácidos dentro de esta hélice del homeodominio, se identificó también que el residuo ubicado en la posición número 9, era indispensable para el reconocimiento específico, ya que al sustituirlo se perdía dicha especificidad, cosa que no sucedía si se cambiaban otros aminoácidos y se conservaba el noveno. Este último descubrimiento estaba en cierta manera en desacuerdo con los datos obtenidos al estudiar el dominio hélice-giro-hélice de los organismos procariotes, en donde se había identificado a los aminoácidos que se encontraban en el extremo amino-terminal, como los más importantes para la unión específica al ADN. Sin embargo, gracias a los estudios estructurales del gen de *Drosophila* denominado *Antennapedia*, se comprobó la importancia del aminoácido número nueve para la unión específica del homeodominio, de esta manera se concluyó que la hélice de reconocimiento en este dominio se orientaba de manera diferente en el surco mayor del ADN, de tal forma que el extremo carboxilo, en donde se encuentra el aminoácido número nueve, es el responsable de la interacción específica con las bases expuestas de la estructura del ácido nucleico (Pabo & Sauer, 1992).

Dominio Myb

El dominio Myb, se identificó primero en la proteína codificada por el oncogen Myb (*v-myb*) del retrovirus AMV (Avian Myeloblastosis Virus) (Klemmner et al., 1982). Este virus tiene la capacidad de producir leucemia mieloblástica o monocítica en el pollo, y de transformar cultivos de células de tipo mielomonocíticas.

Para identificar este dominio, se probó la capacidad de dicha proteína para unirse al ADN al removerle diferentes regiones. Esto permitió establecer que la región indispensable para dicha unión se encontraba en el extremo amino-terminal (Klemmner & Sippel, 1987).

Posteriormente, se comprobó que una gran cantidad de genes que codificaban proteínas relacionadas con el gen *v-myb* se encontraban en especies diferentes como mamíferos, insectos y plantas. Todas estas proteínas presentaban homología con el gen *v-myb* en la región amino terminal, es decir en el dominio de unión al ADN. Esto sugería que dicho dominio resultaba importante para la función de la proteína.

De esta manera, una vez identificado, se comprobó que la unión del dominio

Myb resulta específica (Biedenkapp et al., 1988), ya que las proteínas que lo poseen se unen a la secuencia consenso:

py A A C G T G

En donde py implica una pirimidina

A pesar de esto, no se conocía con certeza la estructura del dominio Myb, e inclusive en un principio se pensó que no presentaba ninguna homología con el dominio hélice-giro-hélice (Biedenkapp et al., 1988). Sin embargo al estudiar más a fondo la estructura, se encontró que el dominio Myb esta conformado por una serie de repeticiones imperfectas de una secuencia de aproximadamente 50 aminoácidos. En plantas y levaduras se encuentran presentes únicamente dos repeticiones de esta secuencia, mientras que en animales estos genes poseen tres repeticiones.

Los aminoácidos en cada repetición, forman tres α -hélices, las cuales se encuentran agrupadas alrededor de un núcleo hidrofóbico constituido por tres triptofanos altamente conservados (Frampton et al., 1991). Estos triptofanos están separados por una cadena de 18-19 aminoácidos, y entre las α -hélices se encuentran intercalados giros muy similares a los encontrados en el homeodominio, por lo que estos dominios son estructuras altamente relacionadas.

Se ha identificado que tanto las repeticiones 2 y 3 del dominio Myb son necesarios para que se lleve a cabo la interacción con el ADN. Asimismo se sabe que la α -hélice 3 de cada repetición, es la responsable de la unión específica del dominio, ya que los residuos que se encuentran en una de las caras de esta hélice (asparagina-179, lisina-182, asparagina-183, asparagina-186 y serina-187), son cruciales para el reconocimiento de las bases en el surco mayor del ADN (Ogata et al., 1992).

Si bien el homeodominio y el dominio Myb son muy parecidos, existen varias características que los distinguen, entre ellas se encuentra el hecho de que en el dominio Myb, el giro que separa a las α -hélices, es un aminoácido más largo y contiene un residuo de prolina, además la orientación relativa de las hélices 2 y 3 es distinta (figura 4). Dentro de la repetición 3 del dominio Myb, las hélice 1 y 2 están alineadas oblicuamente, mientras que la hélice 3 se encuentra casi perpendicular a la hélice 1 (Ogata et al., 1992).

Finalmente, se ha demostrado que la proteína del gen *v-myb*, produce un plegamiento en el ADN cuando se une a su secuencia consenso (Saikumar et al., 1994), esto puede representar una característica fundamental en el mecanismo de activación de las proteínas que contienen este dominio, ya que el plegamiento del ADN es uno de los mecanismos propuestos para explicar el hecho de que una proteína regulatoria pueda influir en la transcripción de un gen que se encuentra a cientos de bases de distancia.

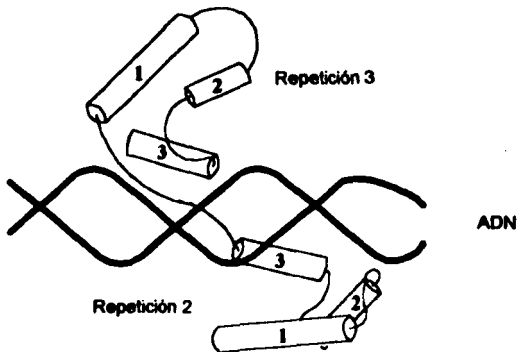


Figura 4. Esquema de la interacción propuesta entre el dominio Myb y el ADN (Tomado de Frampton et al., 1991)

DEDOS DE ZINC

Este dominio de unión al ADN, fue descubierto cuando se estudiaba la expresión de los genes encargados de la síntesis del ARN 5S en la rana *Xenopus laevis*. La transcripción de estos genes se encuentra regulada por el factor de transcripción TFIIIA, el cual junto con otros dos factores, TFIIIB y el TFIIIC, forman un complejo que interacciona con la ARN polimerasa III.

Cuando se intentaba purificar el factor TFIIIA, se descubrió que la presencia de zinc era indispensable para obtener cantidades aceptables de proteína. Una vez que esta se obtuvo pura, se observó que existían una serie de subestructuras, de tamaño similar, que se resistían al ataque de una enzima proteolítica. Además se comprobó por absorción atómica que efectivamente, la proteína TFIIIA contenía alrededor de 10 moléculas de zinc dentro de su estructura (Rhodes & Klug, 1993).

Estas observaciones sugerían que el zinc representaba una parte fundamental dentro de la estructura de la proteína y que además, esta se encontraba constituida de

subunidades repetidas, dentro de las cuales, de alguna manera, los aminoácidos que las constituían, se enlazaban fuertemente entre ellos para formar una estructura bastante estable.

Al analizar la secuencia de aminoácidos del factor TFIIA, se encontró que efectivamente, dentro de una región de la cadena polipeptídica se encontraba una secuencia de aminoácidos que se repetía nueve veces:



T= Tirosina

L= Leucina

F= Fenilalanina

H= Histidina

C= Cisteína

X= Cualquier aminoácido

Las letras en negritas indican los aminoácidos conservados.

Dentro de esta secuencia repetida se conservaban invariablemente, en cuanto a su posición, un par de cisteínas, separadas por una cadena de 12 aminoácidos dentro de la cual también se conservan una leucina y una fenilalanina y una serie de aminoácidos básicos, seguidos por un par de histidinas. Estos datos, junto con las observaciones acerca del zinc, llevaron a proponer una estructura dentro de la cual el átomo de zinc formaba un compuesto de coordinación tetraédrico al enlazarse con las dos cisteínas y las dos histidinas, mientras que los residuos entre estos aminoácidos formaban una espiral que se proyectaba fuera del plano de dicho compuesto, simulando un "dedo" (figura 5), de aquí el nombre del dominio. La interacción entre el zinc, las histidinas y cisteínas fue comprobada después por espectroscopia de absorción de rayos x (Lewin, 1994).

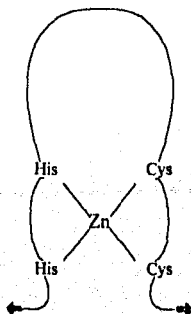


Figura 5. Primera estructura propuesta para los "dedos" de zinc (Tomado de Struhl, 1989)

A pesar de que este modelo concordaba con los descubrimientos hechos hasta ese tiempo, quedaba la duda si efectivamente los 12 residuos formaban simplemente una espiral. De esta manera, considerando también estos datos, se propuso un modelo alternativo, basándose en la estructura ya conocida que presentaban ciertas metaloproteínas, esta estructura puede dividirse en dos mitades, en una de ellas, se encuentra una α -hélice formada por la cadena de residuos que se encuentran entre las cisteínas e histidinas, mientras que en la otra mitad se forma una hoja tipo β , constituida por dos cadenas β antiparalelas. Las dos cisteínas se encuentran flanqueando el giro que se forma en la hoja tipo β , mientras que las dos histidinas se encuentran en la "cara interior" de la hélice, de frente hacia la hoja tipo β (figura 6). Así, el átomo de zinc puede formar un compuesto de coordinación con estos aminoácidos y estabilizar la estructura, manteniendo unidas las dos mitades (Darnell et al., 1990).

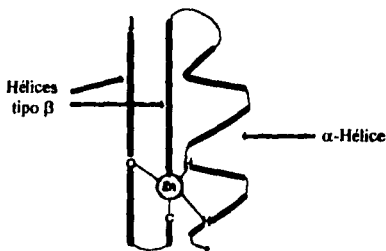
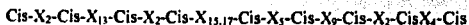


Figura 6. Estructura mas probable de los "dedos" de zinc tipo cisteína-histidina (Tomado de Latchman, 1992)

Existe otro dominio de unión al ADN relacionado al de los "dedos" de zinc, pero estructuralmente distinto, este dominio, denominado por algunos autores como "dedos" de zinc de multicisteínas (Latchman et al., 1990), esta conformado, por la secuencia conservada de aminoácidos:



En donde Cis representa una cisteína y X implica cualquier aminoácido.

Dentro de esta estructura, el zinc también forma un compuesto de coordinación con aminoácidos de la cadena polipeptídica, pero en este caso es con cuatro cisteínas en vez de con dos histidinas y dos cisteínas. Esta interacción también se ha comprobado por espectroscopia de absorción de rayos x. De esta manera, se podría pensar que la única diferencia entre este dominio y el de los "dedos" de zinc es la sustitución de dos cisteínas por un par de histidinas. Sin embargo, se comprobó que cuando se sustituían dichas cisteínas, ya sea por histidinas o por cualquier otro aminoácido, el dominio perdía su capacidad de unirse al ADN. Además se encontró que la secuencia de aminoácidos que podría formar un "dedo", no se repeta invariablemente, sino que solo lo hacía dos veces formando una sola estructura y no unidades independientes como en el caso de los "dedos" de zinc. Por estas razones, se pudo inferir que estos dominios son funcionalmente distintos.

Una de las proteínas mejor estudiadas, desde el punto de vista estructural, que presentan este dominio, es el receptor de la hormona glucocorticoide, dicho dominio esta constituido por dos espirales y dos α -hélices, estas hélices se encuentran perpendiculares una con respecto a la otra, mientras que el zinc se une a un par de cisteínas que se encuentran al comienzo de una de las espirales y a otras dos que se encuentran en el extremo N-terminal de la α -hélice. En esta proteína, la primera hélice interacciona específicamente con el ADN colocándose en el surco mayor, mientras que el resto de la estructura, es decir, las dos espiras y la otra α -hélice interaccionan con la estructura de soporte del ADN constituida por las azúcares y los grupos fosfato (Lewin, 1990).

HELICES ANFIPATICAS

A) Cierres De Leucina

Este dominio se descubrió cuando se intentaba conocer más a fondo el papel que desempeñaba la proteína regulatoria C/EBP. Se sabía que esta proteína que se encuentra en hígado de rata y es resistente al calor, se unía a secuencias regulatorias presentes en los promotores de varios genes, tanto virales como de maníferos. Cuando se purificó y posteriormente se aisló el gen, entonces fue posible comparar la secuencia de aminoácidos con las secuencias de las demás proteínas reportadas hasta entonces (Mcknight, 1991) En base a esta comparación, se pudo encontrar que una región de la proteína C/EBP, rica en el aminoácido leucina, también se encontraba en las proteínas codificadas por los oncogenes Fos y Jun.

C/EBP	LTSDNDRLRKRVEQLSRELDTLRGIFRQL
Fos	LQAETDQLEDEKSALQTEIANLLKEKEL

Jun LEEKVKTLKAQNSELASTANMLREQVAQL

L= Leucina

En estas regiones se encontró que la leucina se encontraba siempre en la séptima posición, formando un hepteto que se repetía 4 veces dentro la cadena de aminoácidos. Esta región se puede esquematizar como una α -hélice en la cual se presenta una leucina cada dos vueltas, conservándose de un solo lado de la hélice, formando una especie de "cordillera" de leucinas.

Además de esta secuencia, todas las proteínas de este tipo presentaban una región adyacente muy rica en aminoácidos cargados, principalmente lisina y arginina, conformando una región básica con carga positiva orientada hacia el extremo amino terminal.

La secuencia conservada de leucinas resultaba muy similar a la de las proteínas fibrosas como las miosinas y queratinas. Estas proteínas forman dímeros dentro de los cuales dos hélices se enrollan sobre sí mismas para formar una espiral doble, por lo que esta era la estructura más probable del cierre de leucina.

De esta forma se postuló que el dominio que constituía el cierre de leucina permitía la dimerización de la proteína mediante la interacción de los residuos de leucina sobresalientes de las dos proteínas, simulando los dientes de un cierre.

Posteriormente gracias a estudios estructurales y a mutaciones dirigidas en la proteína C/EBP, se pudo comprobar que, en efecto, el dominio del cierre de leucina era indispensable para conformar un dímero, ya que la sustitución de este aminoácido impedía que se unieran las dos proteínas. Al no producirse la dimerización, la proteína era incapaz de unirse al ADN. Sin embargo se comprobó que no era el dominio del cierre de leucina el que establecía contacto con el ADN, sino la región básica contigua, en la cual se realizaron mutaciones y se comprobó que a pesar de que la proteína no se unía al ADN, si formaba dímeros.

Por lo tanto se llegó a la conclusión de que este dominio está constituido por un dímero que se forma cuando dos α -hélices anfipáticas, una de cada proteína, se enrollan sobre sí mismas gracias a las interacciones entre las leucinas, para formar una espiral doble (Pabo & Sauer, 1992). Esta conformación permite que el resto de la estructura, es decir las regiones básicas, se encuentren en la posición correcta para interactuar con el ADN. Estas regiones básicas también forman una α -hélice, la cual sufre una ruptura por la presencia del aminoácido conservado aspargina, esta ruptura le da a la hélice un aspecto de gancho que le permite interactuar adecuadamente con el ADN (figura 7).

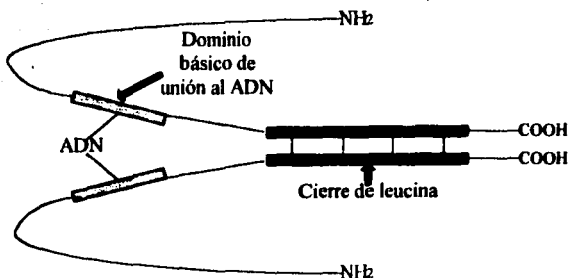


Figura 7. Dominio de cierre de leucina (Tomado de Latchman, 1990)

B) Helice-Vuelta-Helice

La comparación de secuencias de aminoácidos de diversas proteínas regulatorias tales como la proteína muscular MYOD, las proteínas E12 y E47 que aumentan la afinidad de unión de las inmunoglobulinas, proteínas tipo MYC codificadas por el oncogen *c-myc*, etc, llevaron a la proposición de la existencia de otro dominio de unión al ADN denominado hélice-vuelta-hélice. Este dominio, al igual que en el caso de las proteínas con cierres de leucina, presenta una región de aproximadamente 15 aminoácidos, con carácter básico que se une al ADN. Sin embargo, la estructura del cierre de leucina es substituida por una región capaz de formar dos hélices anfipáticas, que se encuentran separadas por una vuelta de tipo no helicoidal de longitud variable. Este tipo de hélices tienen la particularidad de tener todos sus aminoácidos cargados en un solo lado de la hélice (Lewin, 1994).

El análisis de deleciones de la proteína MyoD, demostró que la ausencia del dominio hélice-vuelta-hélice inhibe la dimerización de la proteína y por consecuencia su unión al ADN, de tal forma que este dominio actúa de manera similar al de cierre de leucina.

En el caso de las proteínas tipo MYC, la organización de su dominio de unión es todavía más complejo, ya que el dominio hélice giro hélice, se encuentra situado en el extremo N-terminal después del dominio de cierre de leucina en una posición equivalente a la ocupada por el dominio básico de las proteínas C/EBP, FOS, JUN y GCN4 (Johnson

and McKnight, 1989). De esta manera, se puede inferir que este último dominio es el que le confiere a la proteína su capacidad de reconocer secuencias específicas de ADN (figura 8).

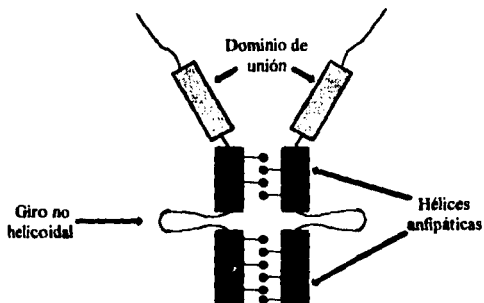


Figura 8. Dímero de una proteína con dominio de unión hélice-vuelta-hélice (Darnell, 1990)

DOMINIO DE TRANSACTIVACIÓN

El dominio de activación es la otra región fundamental de un factor de transcripción, ya que mediante esta, puede realizar una interacción con otros factores de transcripción y establecer una regulación mucho más compleja. Así entonces, una combinación de estas proteínas, pueden regular la transcripción no solo activándola, si no también pueden tener un efecto negativo, o incluso pueden afectar los niveles de transcripción del gen bajo su control. Esta región puede tener una longitud desde 30 hasta 100 aminoácidos.

Una de las características importantes del dominio de activación, es que puede funcionar aún con diferentes dominios de unión al ADN. Por ejemplo, se ha demostrado que una proteína reguladora que posee un dominio de unión al ADN con una estructura de "dedos" de zinc, la proteína JUN, puede actuar como activador transcripcional aún cuando dicho dominio se cambie por el dominio de unión del represor de *E. coli* LexA, el cual forma una estructura de hélice-vuelta-hélice (Struhl, 1989).

El primer dominio de activación que se caracterizó, fue el del gen de levadura GAL4 (Ptashne, 1988). Este dominio presenta un carácter ácido, es decir posee una carga neta negativa y en algunos casos presenta también regiones ricas en residuos de

serina y treonina. Dentro de este dominio no solo es importante su carácter ácido, sino que también se debe considerar su estructura. El análisis de delecciones y el mapeo proteolítico, indican que esta región es una serie de estructuras repetidas compuestas por α -hélices de carácter anfipático, proponiéndose que esta superficie es la que interacciona con otras proteínas dentro de la maquinaria de la transcripción (Struhl, 1989). Estas estructuras repetidas, además de no presentar similitud en su secuencia de aminoácidos, resultan ser funcionalmente redundantes (Mitchell & Tjian, 1989).

Se han encontrado varias proteínas regulatorias que poseen características similares a las de este dominio de activación ácido, pero no se han podido identificar secuencias conservadas como en el caso de los dominios de unión. Algunos ejemplos de estas proteínas son el receptor de la hormona glucocorticoide, el factor JUN, el producto del gen AP-1, etc.

El dominio de activación de carácter ácido, no es el único que se ha encontrado, pues existen por lo menos otros dos tipos de dominios. Uno de ellos, se encontró en el factor Sp1. Esta proteína posee una región rica en glutaminas (aproximadamente el 20 %) la cual se ha demostrado que posee un efecto activador sobre la transcripción de genes reporteros. Se han encontrado otros ejemplos de un dominio de activación rico en glutaminas en proteínas de *Drosophila*, como *Antennapedia*, *Zeste*, en proteínas de levadura como GAL11 y HAP1; y en mamíferos como OCT-1 y OCT-2 (Mitchell & Tjian, 1989). Al igual que en el dominio ácido, cuando se compararon las secuencias de esta regiones, no se pudieron encontrar secuencias conservadas.

El tercer tipo de dominio de activación que se ha encontrado es el de la proteína CTF. Este dominio es rico en prolina (entre un 20 y 30 %), estos aminoácidos impiden que dentro de la estructura se formen hélices. Por lo tanto, este dominio contrasta con el dominio de carácter ácido, en donde se encuentran α -hélices anfipáticas (figura 9).

Debido que existen diversos dominios de activación con características tan diferentes, no ha sido posible hasta el momento, proponer un mecanismo general mediante el cual se lleve a cabo la activación de la transcripción de un gen. Algunos autores como Tjian, proponen que el dominio ácido de activación es la estructura principal mediante la cual se realiza la activación por parte de los factores de transcripción, mientras que las proteínas Sp1, Oct1, etc., cuya región de activación no posee esta característica, actúan a través de otros factores o intermediarios que si poseen dicho dominio (Ptashne and Gann, 1990). Existen evidencias de que los dominios de transactivación ricos en glutaminas o en prolina, estimulan la transcripción proximal. Esto es, a partir de la caja TATA o de sitios relativamente cercanos a esta, mientras que los dominios ácidos y/o ricos en serina/treonina

estimulan la transcripción distal, o sea a partir de una secuencia aumentadora (Scipel et al., 1992).

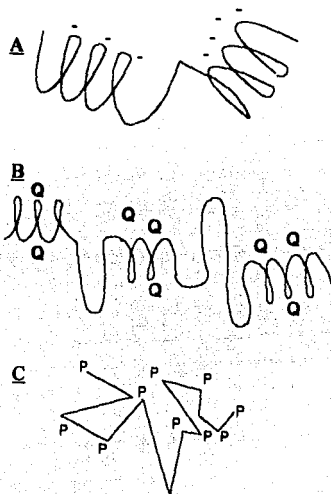


Figura 9. Diferentes estructuras para los dominios de activación. A) Dominio de activación básico del gen GAL 4; B) Dominio rico en glutamina encontrado en el factor Sp1; C) Dominio rico en prolina característico del factor CTF. (Tomado de Mitchell et al., 1989)

El hecho de que un dominio de activación de un factor determinado pueda activar un gen aún con otro dominio de unión, sugiere, que los factores de transcripción actúan sobre elementos que se conservan dentro del proceso de transcripción en organismos eucariotes, como son los factores generales que se unen a la caja TATA, como la proteína TFIID, o sobre la ARN polimerasa II directamente.

FAMILIAS DE FACTORES DE TRANSCRIPCIÓN

A diferencia de los organismos procariotes, en donde generalmente una proteína reguladora solamente reconoce una secuencia específica, en organismos eucariotes se ha encontrado que existe una gran variedad de familias de proteínas regulatorias que reconocen sitios de ADN muy similares. Estas familias se han originado, al igual que muchos de los genes que controlan, a través de la evolución, por procesos como duplicación de genes, divergencia, etc. (Mitchell and Tjian, 1989). La relación que existe entre los genes de estas familias, se presenta principalmente a nivel de su dominio de unión al ADN, ya que la divergencia en la secuencia de aminoácidos que presentan fuera de ese dominio, dos genes de una misma familia, da lugar a funciones muy diferentes.

Algunos ejemplos de estas familias son los genes de *Drosophila* que contienen el dominio de unión al ADN denominado homeodominio, todos estos genes, a pesar de ser diferentes a nivel de secuencia, reconocen una misma región, la cual es rica en adeninas y timinas. Los oncogenes *v-jun* y *v-fos*, así como también algunas proteínas de levadura y *Drosophila*, interaccionan con una secuencia común denominada sitio AP-1.

Este hecho se puede explicar debido a la complejidad de los organismos eucariotes, en donde se deben expresar una gran cantidad de genes en diferentes tejidos, en diferentes tiempos o en respuesta a determinados estímulos. De esta manera, una familia de genes regulatorios puede regular por ejemplo, a un grupo de genes relacionados en tejidos diferentes. Esto representa un sistema de regulación flexible y lo suficientemente preciso para que una pequeña diferencia en las secuencias de reconocimiento, o en la afinidad de los factores por el sitio, permita activar diferencialmente a un gen específico.

REGULACIÓN DE LOS FACTORES DE TRANSCRIPCIÓN

Desde hace mucho tiempo, se sabe que aún cuando una célula se encuentra inhibida en su síntesis de proteínas, existe una inducción de la transcripción dentro de la célula en respuesta a algún estímulo, por ejemplo al calor, esto sugiere, que los elementos que regulan la transcripción, como las proteínas regulatorias, no se sintetizan de "novo", sino que ya están presentes inclusive antes de que se les necesite. Este hecho, implica que deben existir una serie de mecanismos precisos que controlen la actividad de los factores de transcripción preexistentes (Darnell et al., 1990)

Se han descubierto varios mecanismos mediante los cuales los factores de transcripción son regulados, la mayoría de estos mecanismos implican la fosforilación o

defosforilación de ciertos residuos en las cadena polipeptídicas de los factores. Lo cual no resulta extraño debido a que se ha observado que una gran cantidad de estímulos que afectan la expresión de los genes, también resultan en la activación de enzimas que fosforilan proteínas, denominadas *protein-kinasas*.

La regulación por fosforilación puede implicar tanto la activación como la inhibición de un factor de transcripción dado. Esto depende del factor mismo y del nivel al cual se de la regulación. De esta manera, existen tres niveles a los cuales puede presentarse entre control:

Regulación de la translocación hacia el núcleo

Este es uno de los principales mecanismos que se han encontrado para regular la actividad de un factor, ya que este generalmente ejerce su acción en el núcleo, y si se encuentra fuera de este, es decir en el citoplasma, no podrá llevar a cabo su función. Se ha encontrado que un gran número de factores son proteínas fosforiladas, las cuales se encuentran inactivas en el citoplasma en su forma no fosforilada y cuando se presenta un estímulo determinado, se activan enzimas que fosforilan al factor permitiendo esto su transporte al núcleo donde se encuentran sus genes blanco (Hunter & Karin, 1992).

Así, la regulación del transporte al núcleo de una proteína regulatoria generalmente es positiva, ya que la fosforilación permite que el factor lleve a cabo su activación. Entre las proteínas que se encontrado que se regulan de esta manera están la proteína CREB, el receptor de la hormona glucocorticoide, la proteína C/EBP, etc.

Regulación de la actividad de unión al ADN

Este mecanismo, en contraste con el primero, generalmente posee un efecto negativo, ya que la fosforilación de residuos como la serina y treonina dentro de la cadena de aminoácidos del factor resulta en una disminución de su afinidad o en la inhibición total de la unión al ADN. Casi siempre, los aminoácidos que se fosforilan se encuentran muy cerca del dominio de unión al ADN del factor, por lo que se ha propuesto que la causa de esta inhibición es consecuencia de la repulsión electrostática entre los grupos fosfato presentes en la estructura del ADN y los grupos fosfatos añadidos a la proteína

Un ejemplo de este mecanismo se presenta en el producto del gen *c-Myb*, el cual se sabe que se une a secuencias relacionadas con la secuencia consenso:



(Lüsscher et al., 1990). La fosforilación de esta proteína provoca un decremento en la actividad de unión, y este efecto es reversible después de un tratamiento con fosfatasa. Se sabe también que los residuos que se fosforilan son serinas que se encuentran

aproximadamente a 50 bases del dominio de unión al ADN. Otros factores de transcripción que se regulan de esta manera, son el factor jun, la proteína Oct-1, etc.

Regulación de la transactivación

Otro de los mecanismos mediante los cuales se regula a los factores de transcripción es activando su transactivación. Al estudiar la fosforilación de ciertos factores se ha encontrado que los residuos modificados (serina y treonina) se localizan en el dominio de activación del factor, tal es el caso de la proteína CREB. Se ha propuesto que el efecto de la fosforilación sobre el factor, que contiene carga negativa, es el de permitirle a este incrementar su habilidad para interactuar ya sea directa o indirectamente con la maquinaria de iniciación de la transcripción (Seipel, et al., 1992). Otros ejemplos de este tipo los representan el producto del gen GAL, del gen c-myc, c-jun, etc.

TRANSDUCCION DE SEÑAL

En un organismo multicelular, existen etapas específicas como la diferenciación, el desarrollo, o la respuesta a determinados estímulos, dentro de las cuales, una serie de señales diferentes fuera de la célula pueden modificar tanto el número como el tipo de factores de transcripción con los que dicha célula cuenta.

Uno de los ejemplos más claros, se ilustra con la respuesta que presentan casi todos los organismos (bacterias, levaduras, insectos y mamíferos) a temperaturas más altas de lo normal. Bajo estas condiciones, se suspende la transcripción de la mayoría de los genes, para transcribir una serie de genes que codifican a las proteínas denominadas de choque térmico. Estas proteínas, ayudan a la supervivencia del organismo a temperaturas elevadas (por ejemplo, en humanos esta temperatura es de 39-40 °C, mientras que 30 °C es suficiente para inducir estas proteínas en *Drosophila*). El análisis de la región regulatoria de los genes que codifican este tipo de proteínas, ha permitido identificar algunos factores de transcripción que se unen reversiblemente a los promotores de dichos genes, activando su transcripción. De esta manera, en *Drosophila* se identificó el factor de choque térmico HSF, este factor es una proteína de 110 kilodaltones la cual sufre una modificación postraduccional para cambiar de su forma inactiva a su forma activa (Darnell, et al., 1990).

Uno de los ejemplos de transducción de señal mejor entendidos, es el caso de los genes activados por hormonas de tipo esteroide. Estas hormonas tienen la propiedad de ser solubles en lípidos por lo que pueden atravesar fácilmente la membrana celular para unirse a un factor de transcripción, denominado en este caso receptor de la hormona. Este

último forma un complejo con la hormona, lo cual le permite unirse al ADN y activar los genes bajo su control (figura 10). El primer receptor que se caracterizó fue el de la hormona glucocorticoide, la cual se produce principalmente en la glándula adrenal. También se han aislado los genes de los receptores específicos del estrógeno y de la progesterona.

Los sitios de unión al ADN de estos receptores, poseen una longitud promedio de 15 bases, y son de naturaleza simétrica, esto se debe a que estos factores generalmente se unen al ADN como dímeros.

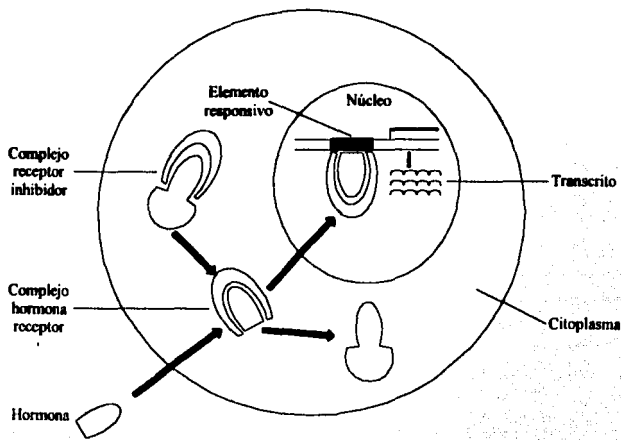


Figura 10. Mecanismo de activación de un gen por una hormona liposoluble (Tomado de Darnell, 1990)

Existe otro mecanismo de activación de genes, en el cual, la molécula que produce la señal no atraviesa la membrana, sino se une a un receptor en la superficie de la célula. Una vez que se produce esta interacción, se desencadenan una serie de eventos intracelulares que involucran la actividad de los llamados mensajeros secundarios como el Ca^{2+} , el diacilglicerol y nucleótidos cíclicos. Estos mensajeros provocan la fosforilación de

proteínas regulatorias a través de la activación de cinasas dependientes de calcio, de la proteinasa C, o de la acción del AMP cíclico.

Entre los ejemplos mejor estudiados de este tipo de activación, se encuentra el caso del factor NF- κ B. Este factor estimula la transcripción de los genes que codifican las inmunoglobulinas clase kappa producidas por los linfocitos B. Se sabe que esta proteína existe en forma inactiva en el citoplasma debido a que se encuentra acoplada a un inhibidor, sin embargo, después de que, por ejemplo, se une un lipopolisacárido bacteriano al receptor celular, se produce la fosforilación del factor NF- κ B, lo que permite que se dirija hacia el núcleo y active la transcripción de sus genes blanco (Darnell et al., 1990).

En otro ejemplo de este mecanismo, el interferon α , que es un polipéptido de 150 aminoácidos, se une, en respuesta a la infección de diferentes virus, a un receptor de superficie celular. Dicha unión activa a un factor de transcripción, el IFN α , el cual normalmente está inactivo en el citoplasma. Como resultado de esto, dicho factor se transporta hacia el núcleo para activar la transcripción.

Factores De Transcripción En Plantas

EL GEN COP 1

El desarrollo de la semilla de una planta puede seguir 2 rutas morfológicas diferentes, una en presencia de luz (fotomorfogénesis) y otra en ausencia de la misma (skotomorfogénesis).

Estudios mutacionales realizados en la planta *Arabidopsis thaliana*, condujeron al aislamiento del gen Cop 1 (Deng, et al., 1992), el cual se propone que tiene la función de reprimir la ruta fotomorfogénica en la oscuridad.

El análisis de secuencia de este gen, reveló que posee un dominio de unión al ADN de la clase de los "dedos" de zinc, similar al que presenta el receptor de la hormona glucocorticoide. Otra característica interesante de esta proteína radica en el hecho de que en el extremo carboxilo, tiene un dominio similar a los que presentan proteínas regulatorias de carácter negativo que se han encontrado en levaduras. Esto último refuerza la afirmación de que la función de esta proteína es la de ser un represor maestro, inactivable por la luz, de la ruta fotomorfogénica (Deng, et al., 1992).

GEN KNOTTED-1

El gen *Knotted-1* (*Kn1*), se aisló a partir de mutantes de maíz obtenidas por inserción de transposones. El fenotipo observado en estos mutantes fue el de la alteración del desarrollo normal de la hoja, ya que las venas laterales de la hoja no se diferencian adecuadamente y forman "nudos" (Vollbrecht et al., 1991). La secuencia de *Kn1* revela la presencia de un homeodominio similar al de proteínas de humano (*Pri*) y de levadura (*MATPi*), presentando un 35% de homología con estas proteínas. Las características que presentan las plantas mutantes en el gen *Kn1*, así como el hecho de que la proteína posea un homeodominio, sugieren que dicho gen regula a genes involucrados en el desarrollo de la hoja del maíz.

GEN OPAQUE 2

Este gen se identificó primeramente en base a estudios genéticos en plantas de maíz, sin embargo, se aisló a partir de mutantes producidas por la inserción de transposones (Hartings et al., 1989). Los estudios genéticos indicaron que la mutación del gen *Opaque 2*, ocasionaba una disminución de la acumulación de proteínas de almacenamiento en la semilla. Estas proteínas constituyen un grupo de polipéptidos solubles en alcohol, que se sintetizan en el endospermo, entre 15 y 40 días después de la polinización. Estos estudios demostraron que la disminución de las proteínas de almacenamiento, se encontraba correlacionada con la disminución de sus ARN mensajeros, sugiriendo que dicha mutación alteraba la transcripción de los genes de dichas proteínas.

El gen *Opaque 2* presenta un dominio de unión al ADN similar al de los oncogenes *fos* y *jun* es decir contiene una estructura de cierre de leucina característica de dichos factores de transcripción. La región adyacente a dicho dominio es rica en aminoácidos básicos y también es similar a la que se encuentra en los oncogenes mencionados y en el factor de levadura *GCN4*.

GENES MYB

Se han encontrado una gran cantidad de genes en plantas que presentan homología a los oncogenes de tipo *myb*.

Un oncogen es una forma mutada de un gen normal (protooncogene), el cual generalmente se encarga de controlar algún aspecto de la división celular. Cuando este gen sufre alguna mutación, o una desregulación de su expresión, puede producir cáncer, el cual se manifiesta por una división celular no controlada y pérdida de la diferenciación. Esta desregulación puede ser ocasionada por un virus al integrarse al genoma de la célula o por mutaciones ocasionadas por agentes físicos o químicos. De esta manera, se sabe que los

oncogenes *myb* se expresan en tejido hematopoyético inmaduro y aparentemente participan en la diferenciación de ese tejido (Sheng-Ong, 1990).

Los genes relacionados con los genes *myb* realizan una serie de funciones en diferentes organismos a través de la activación transcripcional de diferentes genes (Sablowski et al., 1994). De tal forma, estos genes pueden participar en la diferenciación celular (Oppenheimer et al., 1991), en la regulación de vías biosintéticas (Hattori et al., 1992), o en la respuesta a determinadas condiciones ambientales extremas como el estrés hídrico (Urao et al., 1993).

Gen C1

El gen *C1*, aislado de la planta del maíz por inserción de transposones, fue el primer gen con homología a oncogenes celulares de tipo *myb* que se describió en plantas. Mediante estudios genéticos se determinó que este gen interviene en la regulación de la biosíntesis de pigmentos púrpura denominados antocianinas en el grano del maíz. El gen *C1* posee una homología de alrededor del 40% con proteínas de tipo *Myb* encontradas en animales (Paz-Ares, 1987). Esta homología se presenta principalmente en la región del dominio de unión al ADN, es decir en la región que contiene el dominio *myb*, fuera de él, no se encuentra una similitud significativa.

Sin embargo a pesar de la similitud entre dichas proteínas, es improbable que posean una función similar en sus respectivos organismos, ya que el gen *c1*, regula una ruta biosintética que no esta presente en animales (Paz-Ares, 1987).

Gen Gl-1

Este gen es necesario para el inicio de la diferenciación de estructuras denominadas tricomas en las plantas. Estas estructuras se localizan en la superficie de la hoja, el tallo, y en los petiolos y sépalos de la flor. Las mutaciones en el gen *gl-1*, afectan la formación de tricomas, pero no el crecimiento normal de la planta, por lo que se puede afirmar que este gen se encuentra involucrado específicamente en la regulación del desarrollo de dichas estructuras (Oppenheimer, et al., 1991). El gen *gl-1* presenta un dominio de unión al ADN tipo *myb*, similar al que presenta el gen *c1* de maíz, sin embargo, al igual que este, fuera de dicha región no presenta homología. El análisis de la expresión de este gen, indica que se expresa en estípulos, estas estructuras carecen de tricomas, lo que sugiere que *gl-1* controla la síntesis de una señal que activa la diferenciación de los tricomas en la planta (Oppenheimer, et al., 1991).

Biología Molecular De La Tolerancia A La Sequía

Uno de los principales objetivos del estudio de la tolerancia a la sequía a nivel molecular, consiste en elucidar el mecanismo mediante el cual una planta regula y envía las señales necesarias para que en respuesta a la sequía, se induzcan grupos de genes, cuya expresión, producirá una serie de cambios tanto fisiológicos como bioquímicos en la planta que le permitan adaptarse al estrés hídrico.

Sistemas de estudio

Existen dos modelos principales mediante los que se ha abordado el estudio de la tolerancia a la sequía a nivel molecular, estos modelos son:

Embriones de semillas

En casi todas las plantas superiores, los embriones de la semilla, pueden soportar potenciales de agua muy bajos, de hasta el 50% de humedad relativa (Bartels et al., 1988). Así, en cuanto las semillas se ven en presencia de agua, estas son capaces de germinar. El embrión de la semilla resulta un modelo conveniente para estudiar los eventos metabólicos, así como los cambios a nivel molecular que se presentan durante este proceso. Además, se puede evaluar el efecto en estos parámetros, de la adición de diferentes hormonas tales GA (ácido giberélico) y ABA (ácido abscísico) que se sabe que juegan un papel preponderante en la embriogénesis. Una de las desventajas de este modelo es que la tolerancia a la sequía no la presenta la planta en su conjunto, sino que se encuentra restringida al embrión.

Plantas de resurrección

Las plantas de resurrección, constituyen un pequeño grupo de plantas representadas a todo lo largo de la escala filogenética, que habitan generalmente en zonas áridas. Se han encontrado principalmente en países como Sudáfrica, Australia y también en México. Existen especies de plantas de resurrección que pertenecen tanto a la familia de las monocotiledoneas, como a la de las dicotiledoneas. Estas plantas ejercen una función primaria en la colonización de las rocas. Sin embargo, su particularidad más sorprendente, radica en el hecho de que poseen la capacidad de sobrevivir en un estado de deshidratación casi total, durante largos periodos de tiempo, y después de rehidratarse, recuperar completamente su actividad fisiológica, en un lapso de tiempo muy corto.

Un ejemplo de este tipo de modelo, lo constituye la planta de resurrección sudáfricana, *Craterostigma plantagineum* (Gaff, 1971), la cual pertenece a la familia *Scrophulariaceae*. Las hojas de esta planta pueden llegar a soportar una humedad relativa del 1% y aún conservan su viabilidad. Estos tejidos recuperan sus propiedades normales en un tiempo aproximado de 24 horas después del contacto con el agua. Esta planta posee otra característica importante, ya que el conjunto de sus células indiferenciadas, que constituyen el "callo", no es tolerante a la desecación, sin embargo, si se trata previamente con la hormona ABA (ácido abscísico), el callo si resulta viable al ser rehidratado. De esta manera, *Craterostigma plantagineum*, representa un modelo de estudio muy adecuado para identificar productos de genes que son importantes para conferir la tolerancia a la sequía, ya que se pueden analizar dos tipos de tejidos que a pesar de ser metabólicamente distintos poseen la capacidad de recuperar sus funciones normales al ser rehidratados después de sufrir una desecación (Bartels et al., 1990; 1991).

Genes aislados

Los experimentos realizados en estos modelos, han permitido aislar una gran cantidad de genes que codifican proteínas inducidas durante la sequía.

Entre los genes que mejor se han estudiado, se encuentran los genes LEA (late embryogenesis abundant). Se han identificado principalmente como genes que se expresan en las etapas de maduración y de desecación durante el desarrollo de la semilla, pero se sabe también que se expresan en tejido vegetativo en respuesta a diferentes condiciones adversas, como la pérdida de agua, el choque osmótico y las bajas temperaturas, así, se han identificado este tipo de genes en las hojas deshidratadas de las plantas de resurrección (Piatkowski et al., 1990). Las proteínas LEA son de naturaleza hidrofílica en base a su composición de aminoácidos, con regiones ricas en lisina. Se han clasificado en diferentes grupos o familias (tabla 1). Se han propuesto distintos mecanismos mediante los cuales actúan para responder a la pérdida de agua, como la sustitución del agua que mantiene solvatados a las proteínas y/o la estabilización de las membranas durante la desecación (Bray, 1993).

Grupo	Mecanismo propuesto
1 (familia D-19)	Alta capacidad para enlazarse con el agua
2 (familia D-11)	Posibles chaperoninas de proteínas que ayudan a conservar la estructura celular
3 (familia D-7)	Secuestro de iones que se concentran durante la pérdida del agua
4 (familia D-113)	Reemplazamiento del agua para conservar la estructura de la membrana
5 (familia D-29)	Secuestro de iones

Tabla 1. Diferentes grupos de proteína LEA (Tomado de Bray, 1993)

Para demostrar que estos genes son necesarios para tolerar la sequía en una planta, se sobreexpresó de forma independiente, proteínas de tipo LEA aisladas de la planta de resurrección *Craterostigma plantagineum* (estas proteínas se expresan durante la sequía), en plantas transgénicas de tabaco (Iturriaga et al., 1992). Sin embargo, aún no se ha logrado mejorar la tolerancia a la sequía de una planta sobreexpresando genes que se inducen en esta condición.

Ahora bien, la pérdida de agua en una planta, provoca también un desajuste en el balance osmótico de la célula, en el cual, el equilibrio en el potencial de agua dentro y fuera de la célula se ve modificado. Para poder compensar este efecto, la célula expresa una serie de genes que codifican enzimas involucradas en la síntesis de solutos compatibles con el metabolismo u osmolitos, tales como la prolina, la glicin-betaina, la sacarosa, trehalosa, sorbitol, etc. (Yancey et al., 1992). La presencia de estos osmolitos en el citoplasma,

provoca una diferencia de potencial de agua, y debido a que el potencial en el interior de la célula es menor que el externo, es posible que el agua se mueva hacia el interior de la célula.

Factores de transcripción y sequía

Gen GC 19

En el trigo, el gen denominado GC 19, codifica al factor de transcripción EmBP-1, el cual se une a una secuencia conservada de 8 bases (CACGTGGC) que se denomina ABRE (Elementos responsivos a la hormona ABA) (Guiltinan et al., 1990). Esta secuencia se localiza en la región regulatoria del gen Em, el cual se expresa durante la maduración del embrión. Las secuencias ABRE, se encuentran en la mayoría de los promotores de los genes que son regulados por ABA, en diversas plantas como el arroz, el algodón y otros genes de trigo. Esto sugiere que durante la sequía, ABA funciona a través de estas secuencias. Al analizar la estructura de la proteína EmBP-1 se encontró que posee un dominio de unión del tipo de cierre de leucina. Asimismo, se comprobó que una mutación de dos bases en la secuencia que reconoce, no solo impide que el factor se una, sino que también reduce la capacidad de la secuencia ABRE para influir en la transcripción del gen Em. Todos estos datos sugieren que el factor EmBP-1 se encuentra involucrado en la respuesta por parte de diversos genes a la hormona ABA (Guiltinan et al., 1990).

Gen Atmyb2

El gen Atmyb2 de la planta *Arabidopsis thaliana*, es el primer factor de transcripción con el dominio de unión al ADN de tipo Myb que se sabe se induce durante la sequía (Urao et al., 1993). Este gen no solo se induce por sequía y por altas concentraciones de sales, sino que también disminuye su nivel de ARNm al rehidratar la planta.

La estructura del factor Atmyb2 consiste en dos repeticiones imperfectas de aproximadamente 50 aminoácidos, ubicadas en el extremo amino terminal, las cuales conforman el dominio de unión Myb. En el extremo carboxilo terminal presenta una región ácida de aproximadamente 30 aminoácidos. La comparación de la secuencia del dominio de unión de este gen con la de otros dominios Myb, demostró que el gen Atmyb2 es mucho más homólogo al gen AM340 de *Anthrhumum* (Jackson et al., 1991) que a cualquier otro gen con este dominio.

Una característica más de este gen es que su expresión se induce al aplicar la hormona ABA, al igual que otros genes que responden a la sequía.

De esta manera, las características que presenta este factor de transcripción, sugieren que controla la expresión de genes que se inducen por sequía o por una alta concentración de sales (Urao et al., 1993).

En base a lo anterior, se plantearon los siguientes objetivos:

OBJETIVO GENERAL

Caracterizar un gen aislado de la planta tolerante a la sequía *Craterostigma plantagineum*, el cual presenta homología por hibridación de ADN, a los oncogenes celulares de tipo Myb.

OBJETIVOS PARTICULARES

Establecer el mapa de restricción del gen denominado CpM5.

Subclonar los fragmentos necesarios de este gen para poder obtener su secuencia.

Obtener la secuencia de nucleótidos de este gen para comprobar que se trata de un gen con homología a Myb y comparar su secuencia con secuencias reportadas de otros genes de la misma clase.

HIPOTESIS

Una vez que se obtenga la secuencia de nucleótidos del gen CpM5 y se conozca su estructura, es decir, en donde se localiza el posible promotor del gen, sus intrones y exones, y las secuencias consenso que pueda contener, se podría sugerir la función y la regulación de este gen dentro del organismo del que proviene.

MATERIAL

Cepas

Las cepas microbianas utilizadas para amplificar el ADN de plásmido, son derivadas de *E. coli*, en especial se utilizaron las cepas DH5 α y Top 10 para obtener ADN para secuenciar, y la cepa HB101 para transformar, debido a su alta eficiencia.

Enzimas de restricción y modificación

En general, se utilizaron enzimas de la marca Boehringer-Mannheim, o bien, New England Biolabs y GIBCO-BRL. Para obtener la secuencia se usó la enzima Sequenase de la marca United States Biochemical.

Reactivos y solventes

Todos los reactivos utilizados fueron de la marca GIBCO-BRL, Sigma, Merck o Baker, grado analítico.

MÉTODOS

1 Aislamiento de ADN

1.1 Preparación de plásmido en gran escala

El método utilizado para la extracción de ADN de plásmido fue una modificación del método de lisis por detergente (Iturriaga, 1989). A partir de un cultivo de toda la noche de 400 ml. de medio de cultivo LB, se separaron las células por centrifugación a 5000 rpm durante 10 minutos a 4 °C en un rotor Beckman JA 10. La pastilla obtenida se secó y se reuspendió en 5 ml de una solución de 25% de sacarosa y Tris-HCl 50 mM pH 8.0 en hielo. Se adicionó 1 ml de una solución de lisozima de 10 mg/ml en Tris-HCl 250 mM pH 8.0 y se incubó 5 minutos en hielo, en seguida, se agregaron 2.8 ml de EDTA 250 mM pH 8.0 incubándose durante otros 5 minutos. Después de la adición de 8.8 ml de una solución de Tris-HCl 50 mM pH 8.0, EDTA 60 mM y Triton X-100 al 10%, el lisado se centrifugó a 18000 rpm durante 1 hora a 4°C en un rotor Beckman JA 20. El sobrenadante se removió cuidadosamente y se precipitó con 6 ml de una solución de una solución de PEG 8000 al 30% y NaCl 1.5 M incubándose durante 15 minutos en hielo. Se recuperó la pastilla después

de centrifugar a 5000 rpm durante 10 minutos a 4°C. Esta pastilla se secó muy bien y se resuspendió en 2 ml de Tris-HCl 50 mM, EDTA 10 mM y se incubó con RNAsa A (10 µg/ml) a 37°C durante 30 minutos.

Posteriormente el ADN se purificó por ultracentrifugación en gradiente de cloruro de cesio, de esta forma, después de la adición de 3.5 g de CsCl sólido y 120 µl de solución de bromuro de etidio 10 mg/ml, el volumen se completó hasta 4.2 ml. con solución de rebando (1.55 g de CsCl, 600 mg de bromuro de etidio por ml de solución Tris-HCl 50 mM pH 8.0, EDTA 10 mM). El material insoluble se removió por centrifugación a 9000 rpm durante 15 minutos a 20°C en un rotor Beckman JA 20. El sobrenadante se llevó a un volumen de 5.3 ml con solución de rebando y se centrifugó hasta el equilibrio a 50,000 rpm durante toda la noche a 20°C en un rotor Beckman NVT 65.

Se obtuvieron dos bandas después del gradiente, la superior que corresponde al ADN linear bacteriano y al ADN circular de plásmido pero con "nicks" mientras que la banda inferior corresponde al plásmido circular cerrado (Sambrook et al., 1989). Esta banda se extrajo con una solución saturada de SSC 20X (SSC 1X = 0.15 M de NaCl, 0.015 M citrato de sodio) en isopropanol, después se precipitó con 3 volúmenes de etanol a -20°C durante toda la noche.

1.2 Preparación de plásmido a pequeña escala

Las minipreparaciones de plásmido se realizaron de acuerdo al método alcalino (Sambrook et al., 1989) en el cual las colonias se crecieron en 2 ml de medio de cultivo LB con ampicilina durante toda la noche a 37°C. Un volumen de 1.5 ml de este cultivo se centrifugó a 6000 rpm durante 3 minutos en tubos eppendorf a 4°C, se eliminó el sobrenadante y la pastilla se resuspendió en 50 µl de una solución Tris 50 mM pH 8.0 y 25% de sacarosa dejándose 5 minutos en hielo. Después se agregaron 200 µl de solución de NaOH 0.2 N y SDS al 1%, se mezcló por inversión y se dejó en hielo 5 minutos. Posteriormente se agregaron 150 µl de acetato de potasio 5 M se mezcló por inversión y se esperó 5 minutos para centrifugar a 13000 rpm durante 5 minutos a 4°C. Se recuperó el sobrenadante en un nuevo tubo eppendorf y el ADN se extrajo con una mezcla de fenol:cloroformo:alcohol isoamílico (25:24:1) y se precipitó con un volumen de isopropanol, 0.1 volúmenes de acetato de sodio 3 M pH 5.2, e incubándolo durante 3 minutos en nitrógeno líquido. Después de centrifugar el ADN a 13 000 rpm durante 30 minutos a 4°C, la pastilla se lavó con etanol al 70 % y se secó en una centrifuga Speed-Vac. Posteriormente se resuspendió en 50 µl de buffer TE (Tris-HCl 10 mM pH 8.0 y EDTA 1

mM). Se utilizaron 5 μ l de este volumen para realizar digestiones con enzimas de restricción y así entonces poder seleccionar los plásmidos recombinantes mediante el análisis de los fragmentos resultantes separados por electroforesis en geles de agarosa. Este procedimiento se utilizó siempre para confirmar la identidad de las clonas recombinantes antes de subclonarias o de obtener preparaciones de plásmido a gran escala.

2. Digestión de ADN con endonucleasas de restricción

Se utilizaron los buffers en los que la enzima presentaba su máxima actividad de acuerdo a la información del fabricante, excepto en los casos en los que se usaron mezclas de enzimas y se usó el buffer más adecuado para ambas. Cuando la digestión no se pudo realizar con las dos enzimas juntas, primero se digirió con una, se extrajo el ADN con fenol, se precipitó con etanol y se digirió con la otra enzima. La concentración final de la enzima nunca excedió del 10% del volumen total para que la cantidad de glicerol no afectara su actividad. En general, la concentración del ADN en los volúmenes de digestión fue de 0.05-0.10 μ g / μ l. Las digestiones pequeñas generalmente se llevaron a cabo usando 1 unidad de enzima por μ g de ADN en un intervalo de tiempo de 1-3 horas. En el caso de las digestiones preparativas se llevaron a cabo durante toda la noche usando 0.1-0.5 unidades de enzima por μ g de ADN. Las digestiones se detuvieron inactivando la enzima a 65°C por 10 minutos o mediante extracción con fenol:cloroformo:alcohol isoamílico.

3. Separación de fragmentos de ADN por electroforesis

3.1 Geles de agarosa

Las fragmentos de ADN de 0.3 a 4.0 Kb se separaron mediante geles horizontales de agarosa al 1%, utilizando el buffer TBE (Tris-borato 134 mM, ácido bórico 89 mM, EDTA 2.5 mM). Estos geles dependiendo del tamaño se corrieron a un voltaje aproximado de 90-130 volts. Tanto al buffer de corrida como al gel se le añadió bromuro de etidio a una concentración de 0.5 μ g / μ l de modo que los geles ya no necesitaban una tinción posterior. El buffer de muestra consistió en 30% de Ficoll, EDTA 10 mM pH 8.0, de Azul de bromofenol 0.025% y de Cianol de xilano 0.025%. Finalmente, las bandas de

ADN fueron visualizadas en un transiluminador de luz UV a 254 nm y fotografiadas con película Polaroid tipo 667.

4. Aislamiento de fragmentos de ADN a partir de geles de electroforesis

4.1 Electroelución en bolsas de diálisis.

Este método se utilizó para recuperar fragmentos de ADN a partir de geles de agarosa, siguiendo una modificación del método original (Iturriaga, 1989). Después de identificar las bandas de ADN en el gel, estas se cortaron con una navaja y los fragmentos del gel se depositaron en bolsas de diálisis con 300 μ l de buffer TBE. Esta bolsa se colocó en una cámara de electroforesis con buffer TBE y el ADN se electroeluyó durante 1 hora a un voltaje de 90 V aproximadamente. El ADN se liberó de la bolsa aplicando un pulso de corriente de 10 segundos al mismo voltaje pero invirtiendo la polaridad. El contenido de la bolsa se vació en un tubo eppendorf y la bolsa se lavó con 300 μ l más de buffer, el ADN se extrajo con mezcla de fenol:cloroformo:alcohol isoamílico y se precipitó con 1 volumen de isopropanol y 0.1 volúmenes de NaAcO 3 M pH 5.2. Finalmente el ADN se resuspendió en 30 μ l de buffer TE.

5 Clonación de fragmentos de ADN en plásmidos

5.1 Ligación del ADN

Los vectores más utilizados para realizar subclonaciones fueron el plásmido vector pUC 19 (Yanisch-Perron et al., 1985), y pBluescript (Stratagene cloning systems). Para evitar que durante la ligación, los extremos de los vectores se ligaran entre si mismos, resultando en un alto número de transformantes no recombinantes, generalmente, se trató el vector con fosfatasa alcalina para remover los grupos fosfato de los extremos 5'. De esta forma, después de una digestión con enzimas de restricción, se calentó la reacción a 65°C por 10 minutos para inactivar la enzima, después se ajustó el volumen del buffer de la fosfatasa alcalina (Tris-HCl 20 mM pH 8.0, MgCl₂ 1 mM para una concentración final 1X y se adicionaron 5 unidades de dicha enzima, incubándose a 37°C durante 1 hora. Después el ADN se extrajo 2 veces con fenol y se precipitó con etanol.

Las reacciones de ligación se llevaron a cabo siempre en un volumen de 10 μ l, utilizando una relación equimolar entre el vector y el inserto a una concentración de 5-10 μ g/ μ l en Tris-HCl 50 mM pH 7.6, MgCl₂ 10 mM, PEG 8000 5% (P/V), ATP 1mM, DTT 1mM, y 1 unidad de T4 ADN ligasa. La incubación se realizó durante toda la noche a 14°C en el caso de extremos no cohesivos, en el caso de extremos cohesivos, la temperatura de incubación fué de 25 °C durante 2-4 horas.

5.2 Transformación de células competentes de *Escherichia coli*.

Para preparar células competentes (Sambrook et al., 1989) se inocularon 100 ml de medio LB con 1 ml de un cultivo de toda la noche de *E. coli* DH5 α , y se incubaron durante 3-4 horas a 37°C hasta una densidad óptica de 0.5 a 550 nm. A esta D.O., las células poseen una densidad aproximada de 5×10^7 células por ml. Las células se separaron por centrifugación a 5,000 rpm durante 10 minutos a 4°C en un rotor Sorvall SS34, para resuspenderlas cuidadosamente en 33 ml de cloruro de calcio 0.1 M e incubarlas en hielo durante 1 hora. Después se centrifugaron nuevamente a las mismas condiciones y se resuspendieron en 2 ml de CaCl₂ 0.1 M, se tomaron alícuotas de 100 μ l y se depositaron en tubos eppendorf para incubarlos en hielo durante por lo menos 2 horas antes de la transformación. La eficiencia de transformación óptima para esta cepa (10^6 transformantes por μ g de ADN), se presentó cuando las células se incubaron toda la noche, sin embargo, también se pueden utilizar células de 2 días de incubación en hielo, pero su eficiencia decrece a 10^5 o 10^4 transformantes por μ g de ADN.

Para la transformación, una alícuota de 100 μ l de células competentes se mezcló con 25-50 ng de ADN religado o con 1 ng de vector no cortado, se incubó durante 30 minutos antes de someter a las células a un choque térmico de 2 minutos a 42°C. Inmediatamente después, las células se transfirieron al hielo y se enfriaron por 2 minutos más antes de adicionarles 0.9 ml de medio LB e incubarlas durante 1 hora a 37°C con agitación moderada. De este volumen se tomaron 100 μ l y se espatularon en placas de LB con ampicilina (100 μ g/ml), incubándose estas últimas durante toda la noche a 37°C.

5.3 Selección de clones recombinantes.

Las clonas recombinantes fueron identificadas por la inactivación del marcador de la β -galactosidasa (Sambrook et al., 1989). La región de policlonación tanto en el pUC 19 como en el pBluescrip se encuentra insertada en fase en el aminoácido 9 de la región N-terminal del gen que codifica el fragmento α de la β -galactosidasa, de 45 aminoácidos en total. Este gen que es expresado por el promotor del gen *lac*, puede ser inducido por el análogo de la galactosa IPTG, este compuesto impide que el represor se una al promotor *lac* produciéndose así el péptido α , el cual complementa a una β -galactosidasa a la cual le han sido deletados del aminoácido 11 al 41. De esta forma, la presencia de una enzima activa en una determinada clona es detectada por su habilidad de hidrolizar al compuesto X-gal, liberando bromocloro indol el cual se oxida a un compuesto color azul indigo. Las moléculas de vector que contienen ADN insertado en esta región no producen β -galactosidasa activa, ya que el inserto rompe el marco de lectura de la β -galactosidasa. Así, las colonias que poseen vector con inserto resultan incoloras.

6 Preparación de sonda radiactiva

6.1 Marcaje del ADN con oligos (Oligo-labelling)

Este método (Feinberg and Vogelstein, 1983) es muy eficiente para marcar fragmentos menores de 300 bases, sin embargo también se puede utilizar para fragmentos más grandes. La reacción se llevó a cabo en un volumen final de 25 μ l los cuales contienen 25-50 ng de ADN, el cual se debe de calentar durante 5 minutos a 90°C para separar las dos cadenas y después se enfría en hielo, la reacción también incluye 6 μ l de Buffer para extensión por mezcla de nucleótidos 5X (Dupont, Biotechnology Systems NEN Research Products), 6 μ l de mezcla de deoxinucleótidos trifosfato 5X (Dupont, Biotechnology Systems NEN Research Products), 20-50 μ Ci de [α -³²P] dCTP, y 2 unidades de enzima Klenow. La reacción se incubó a temperatura ambiente durante toda la noche.

7 Transferencia Southern

7.1 Transferencia al gel

Los geles de agarosa se prepararon y se corrieron de acuerdo al protocolo descrito anteriormente. La foto del gel se tomó con una regla de manera que se pudieran identificar las bandas de hibridación en la autoradiografía del gel. Para llevar a cabo la

transferencia de ADN de una sola cadena se siguió el método de Southern (Sambrook et al., 1989), en el cual el gel se debe tratar con una solución de HCl 0.15 M durante 15 minutos, esto para facilitar la transferencia mediante la depurinización parcial del ADN produciendo fragmentos más pequeños.

Posteriormente el gel se lavó dos veces con una solución de 6e NaOH 0.5 M, NaCl 1.5 M durante 30 minutos para desnaturalizar el ADN. Finalmente se neutralizó con dos lavados de 30 minutos en solución Tris-HCl 0.5 M pH 7.5, de NaCl 1.5 M.

Una vez tratado el gel, se colocó sobre un soporte sobre el que se encuentra colocado un papel Whatman humedecido en una solución SSC 10X. Este papel Whatman debe ser lo suficientemente largo para estar en contacto con un reservorio que contiene también solución 10 X SSC (500 ml aproximadamente). Después se colocó una pieza de membrana Hybond-NTM (Amersham) del mismo tamaño del gel, humedecida en solución SSC 2X y esta se cubrió con una pila de aproximadamente 8 cm de toallas absorbentes. Se colocaron plásticos de manera que se evitará el contacto entre el papel Whatman y las toallas absorbentes y finalmente se colocó un peso de 0.5 a 1.0 Kg para dejar la transferencia durante toda la noche.

7.2 Hibridación

Después de la transferencia el ADN se fijó a la membrana mediante luz ultravioleta en un aparato UV Stratalinker (Stratagene) y después se dejó secar a temperatura ambiente. Esta membrana se prehibridó con 50-70 ml de una solución que contenía SDS 0.2%, solución de Denhardt's 5X, SSC 2X, y 50 mg/ml de ADN de esperma de salmón (hervido previamente durante 5 minutos y enfriado en hielo), a una temperatura de 55°C durante 2 horas. Una vez terminada la prehibridación, se agregó la sonda (hervida previamente durante 5 minutos), en 20 ml de la solución anterior fresca para hibridar durante toda la noche a 55°C. A la membrana se le realizaron 4 lavados, los dos primeros con 250 ml de una solución SSC 4X y SDS al 1% y los siguientes lavados con 250 ml de solución SSC 2X y SDS al 1%. Todos estos lavados fueron de 15 minutos a 55°C. Para realizar la autoradiografía de la membrana se utilizó una película Kodak X-OMAT AR expuesta durante 1 noche a -70 °C.

8 Secuenciación del ADN

El método que se utilizó para obtener la secuencia de ADN fue el de la terminación de la cadena, el cual es un método derivado del protocolo original (Sanger et al., 1977).

8.1 Reacciones de secuencia

Las reacciones de secuencia se llevaron a cabo utilizando ADN obtenido ya sea por el método de lisis alcalina, o por el de purificación por gradiente de cloruro de cesio. Con este último método se obtuvieron mejores resultados. El método que se utilizó fue una modificación del método convencional, ya que se utilizó Dimetilsulfóxido (DMSO) como agente desnaturalizante (Seto, 1990). De esta manera, 3 µg de ADN se llevaron a 50 µl con buffer TE en un tubo eppendorf, se desnaturalizaron añadiendo 4 µl de NaOH 2 M y 4 µl de EDTA 2 mM, incubando a 90 °C durante 5 minutos. Después, se adicionaron 6 µl de acetato de sodio 3 M, 14 µl de agua y 160 µl de etanol, se precipitó a temperatura ambiente durante 10 minutos, y se recuperó centrifugando a 13000 rpm, a 18 °C durante 20 minutos. Posteriormente se lavó dos veces con etanol al 70% y se secó perfectamente. Una vez listo el ADN, se agregaron 2 µl de buffer de reacción 5X, 1 µl de DMSO y 1 µl de primero (2.9 pmol o 10 ng). Se calentó a 65 °C 5 minutos y se dejó enfriar lentamente hasta temperatura ambiente. Antes de comenzar la reacción de marcado, se prepararon 4 tubos eppendorf, etiquetados como G, A, T, y C, y en cada uno se añadió 2.5 µl de mezcla de terminación y 0.25 µl de DMSO. Estos tubos se conservaron en hielo. Para comenzar con el paso de marcado del ADN, se agregó 1 µl de DTT (Ditiotreitol) 0.1 M, 2 µl de mezcla diluida de marcaje, 0.5 µl de ³²S-ATP (10 mCi/ml), 0.5 µl de DMSO y 2 µl de la enzima Sequenase (United States Biochemical) diluida 1:8 con buffer de dilución para dicha enzima. Esta mezcla se incubó 5 minutos a temperatura ambiente. Los tubos con la mezcla de terminación se deben precalentar a 37 °C al menos durante 1 minuto. Una vez que terminó la reacción de marcado, se transfirieron 3.5 µl de esta reacción a cada uno de dichos tubos, se centrifugó para sincronizar y se incubaron a 37 °C durante 5 minutos. Después de dicho lapso, se agregó 4 µl de colorante (stop mix; 95 % de formamida, EDTA 20 mM, 0.05% de azul de bromofenol y 0.05% de azul de cianol), se centrifugó un pulso y las reacciones se almacenaron como máximo durante 1 semana a -20 °C. Para cargar en el gel, las muestras se desnaturalizaron previamente a 80 °C durante 3 minutos, para enfriarlas inmediatamente después en hielo.

8.2 Gels de acrilamida

Primeramente, una vez que se encontraban perfectamente limpios y libres de grasas, los dos vidrios para secuencia, se trataron con soluciones distintas. Al vidrio inferior, se le aplicó una solución de 5 ml de etanol, 150 μ l de ácido acético al 10% y 20 μ l de MPTS (γ -Metacril oxipropil trimetoxi silano). Esta mezcla permite que el gel de acrilamida se mantenga unido al vidrio. El vidrio superior, se trató con una solución de 500 μ l de dimetil dicloro silano (DDCS) en 5 ml de cloroformo. Estas soluciones se dejaron secar durante 20 minutos y se procedió a preparar el gel de acrilamida. De manera general, se preparó un gel de acrilamida al 6%, utilizando 24 gramos de urea disueltos en 14.5 ml de agua, posteriormente se agregaron 5 ml de buffer TBE 10X, y 7.5 ml de una solución de acrilamida/bisacrilamida (19:1) al 40%. Por último se llevó a un volumen final de 50 ml y esta mezcla se filtró usando un papel Whatman 1.

Para iniciar y catalizar la reacción de polimerización de este volumen de mezcla de acrilamida, se utilizaron 250 μ l de persulfato de amonio al 10% y 29 μ l de TEMED (N,N,N',N'-Tetrametil etilendiamina).

9 Análisis de secuencia por computadora

Para analizar y comparar tanto la secuencia de ADN como de la proteína deducida se utilizaron los programas Pileup, Compare, Translate, Fasta y Tfasta del paquete de software de la Universidad de Wisconsin (Devereux et al., 1984). Para obtener el dendograma se utilizó el programa CLUSTAL (Higgins & Sharp, 1978). Estos programas se corrieron en una computadora Macintosh VxII conectada como terminal a una computadora VAX.

RESULTADOS

El objetivo inicial de este trabajo, y una de las líneas principales del laboratorio, fue el aislamiento y caracterización de factores de transcripción involucrados en la regulación de la expresión de genes durante la sequía. Con este fin se escogió como modelo de estudio la planta de resurrección *Craterostigma plantagineum*. Las ventajas de este modelo de estudio ya se mencionaron anteriormente.

Con este fin, se realizó un rastreo de un banco genómico de esta planta con una sonda que contenía una región muy conservada de un factor de transcripción para poder identificar genes regulatorios relacionados (Iturriaga et al., en preparación). La sonda que se utilizó para rastrear el banco proviene del ADN complementario del gen Hv5 de cebada (Wissenbach et al., 1993). Este gen, posee dentro de su dominio de unión al ADN, dos repeticiones imperfectas de alrededor de 50 aminoácidos, con triptofanos altamente conservados en posiciones específicas, características típicas de un dominio de unión tipo Myb. La sonda, de una longitud de 640 pares de bases, abarca todo el dominio Myb y una parte más de la región codificante hacia el extremo 3'. Para obtener este fragmento, se utilizó una enzima del sitio de policlonación del vector, la enzima Bam HI y un sitio de restricción Apa I, el cual se encuentra dentro de la secuencia del gen.

El resultado del rastreo del banco genómico, conjuntamente con el mapeo de restricción e hibridación Southern, permitieron la identificación de 6 posibles clones (Iturriaga et al., en preparación).

De estas 6 clones, 2 de ellas presentaban un patrón de restricción muy similar, ya que al cortar con Bam HI, se liberaba un fragmento de 500 pares de bases, el cual hibridaba fuertemente con la sonda de cebada. Esto hacía suponer que la región homóloga al dominio Myb de estas dos clones, se encontraba precisamente en dicho fragmento.

Al subclonar y obtener la secuencia nucleotídica del fragmento Bam HI-BamHI de 500 bases de la clona 5, se pudo comprobar que efectivamente, contenía un dominio de unión homólogo al dominio Myb. Esto mismo, ya se había comprobado anteriormente para la clona 10 (Iturriaga et al., en preparación) denominada como CpM10 (*Craterostigma plantagineum* Myb 10).

Con el fin de obtener la secuencia completa del gen contenido en la clona 5, denominado CpM5, y conociendo de antemano la posible estructura del gen, se subclonaron dos fragmentos adyacentes, uno en cada extremo del fragmento de 500 bases (figura 11).

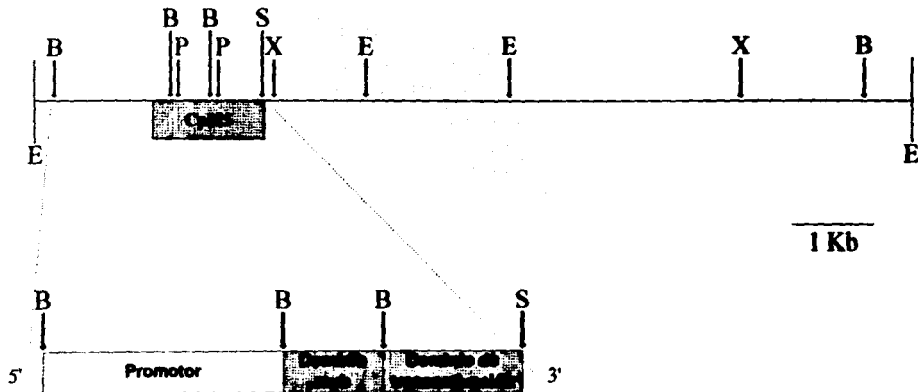


Figura 11. Mapa de restricción de la clona genómica 5. La longitud total de la clona es de 11.5 Kilobases (Kb). El significado de las letras es el siguiente: B= Bam HI; E= Eco RI; P= Pst I; S= Sal I; X= Xba I. El mapa está a escala, en la parte inferior se muestra una línea que equivale a una kilobase (Kb).

En la parte inferior se muestra un esquema del gen CpM5 mostrando los sitios de restricción que se usaron para subclonar el gen y poder secuenciarlo.

Ambos fragmentos se subclonaron en el vector pBluescript SK. El primero de ellos, de Bam HI a Sal I con una longitud de 700 pares de bases y ubicado en el extremo 3' de la clona CpM5, contiene lo que constituye el posible dominio de transactivación del gen. El otro fragmento que va de Bam HI a Bam HI, es de 1200 pares de bases de largo y se encuentra en el extremo 5' de la clona, contiene el resto del dominio homólogo a Myb y la región codificante hacia el extremo 5', el inicio de la transcripción, el líder, y el promotor (figura 11). De esta manera, la longitud total de la secuencia que se obtuvo fue de 1558 pares de bases, ya que no se secuenció el promotor.

La deducción de la estructura del gen se hizo en base al análisis de la secuencia, y a la comparación con la secuencia de otros genes con dominio Myb, tomando como base los resultados obtenidos del gen CpM10 y del ADN complementario del gen CpM7 (Iturriaga et al., en preparación).

La secuencia genómica del gen CpM5 presenta tres exones interrumpidos por dos intrones de longitud diferente (figura 12). Esta secuencia codifica para una proteína de 333 aminoácidos con un peso molecular aproximado de 36.6 KDa.

El primer exón codificaría una secuencia de 61 aminoácidos y esta secuencia contiene una parte del dominio tipo Myb del gen, que corresponde a las dos terceras partes de la primera repetición del dominio. El primer intrón de 167 pares de bases de longitud es precedido por la secuencia TG al final del primer exón y comienza con la secuencia GT. Esta secuencia no coincide exactamente con el consenso AG/TG, sin embargo, la timina es el segundo nucleótido más frecuente (24%) después de la adenina en el par de bases que constituyen el final del exón (Brown, 1986). Al final del intrón se encuentra la secuencia AG/G, la cual cumple con el consenso encontrado en la mayoría de los genes que codifican proteínas nucleares de plantas (Brown, 1986). Además, dicho intrón se encuentra en la misma posición que en otros genes con homología Myb de plantas, ya que inicia después de la primera base (guanina) del codón que codifica al aminoácido glicina (Paz-Ares et al., 1987; Oppenheimer et al., 1991; Noda et al., 1994). Esto se confirmó con la secuencia del ADNc del gen CpM7 (Iturriaga et al., en preparación).

El segundo exón posee una longitud de 130 pares de bases que codifican para 43 aminoácidos contiene la otra parte de la primera repetición y la mitad de la segunda del dominio Myb. El otro intrón, con una longitud de 75 bases, también posee las secuencias características en los extremos donador y receptor de los intrones.

CAGACATGAACCAACAGCAGGTTAAAGTTTCCAAGAACACAAGCAAGTTAATAACTGTG 60
D M N Q Q Q V K V S K N N K Q V N N C E

AAGACGATGATGATTCTTCGGACCTCAGAAGAGGACCTTGGACTGTTGATGAAGACTTCA 120
D D D D S S D L R R G P W T V D E D F T

CACTCATCAACTACATCGCTCACCATGGCGAAGGAAGATGGAACTCTCTTGCACGTTTCG 180
L I N Y I A H H G E G R W N S L A R F A

CTGGTAATAATAAAGACATATTTATATTACATGTACACACAGAGAGAGAGAGAAGA 240
G

GTAGTGCATGGATATGGATCCTATTTCTTTGTCCGTATTGAAAAAGAGGAAAAGGAAAA 300

AGAACACAAACTCAATTATAAAGTGATCAAGTGGCTTCTGAGAAAAAAGGGCTAAAGGA 360

AACAAGGGAAGATCTTATTTTTTTTGTTTTACTTTCTACTAACTTTTTGCATCGATCGCA 420

GGCCTGAAACGAACTGGAAAGAGCTGCAGATTGAGATGGTTAAACTACTTGAGACCCGAC 480
L K R T G K S C R L R W L N Y L R P D

GTTCTGCGAGGAAATATCACACTCGAAGAACAGCTTCTCATTCTTGAACTCCATTCTCGC 540
V R R G N I T L E E Q L L I L E L H S R

TGGGGCAATAGGTACACGTCATAACATACTTACTGTAATAACATATGATCAAAAGAACT 600
W G N R

AATATTTGCAAGAATTAATCCGTAGGTGGTCTCGAAGATTGCCCAACATTTGCCTGGAAGG 660
 W S K I A Q H L P G R

ACGGACAATGAGATAAAGAACTACTGGAGAACAAGGGTCCAAAAACATGCCAAACAGCTA 720
 T D N E I K N Y W R T R V Q K H A K Q L

AAGTGGCAGCTCAACAGCAAACAGTTCAAAGACACCATGAGATACCTTTGGATGCCAAGA 780
 K C D V N S K Q F K D T M R Y L W M P R

TTGGTCGAGAGGATCCAAGCTTCCGGCCACCACCGACGAGCGGAGCACCGCGGGCGGTGGTC 840
 L V E R I Q A S A T T D D G A P P A V V

TCCTCGCCGTCGTCCGCCATGAATACAGCCTGCTACACCACTGCCATGGTGCAGGGCAT 900
 S S P S S A M N T A C Y T T A M A A G D

CACCGTCGTCACCTTTATGCTGATGCCACAGTATTACGCGACGACGACGACGCCACAAT 960
 H R R Q F M L M P Q Y A T T T T T L N

AATTCGATGATCGCCCAAGAAAATTCGAGCACCGTGGCCTCGTCGGAATCTTTTGGCAGT 1020
 N S M I A Q E N S S T V A S S E S F G S

CTTTCTTCTGAACTAACGGCGGAGGCCAACTACGCTAACTATCACCGAGTAATCAACGGC 1080
 L S S E L T A E A N Y A N Y H R V I N G

GCCGATCACCAACAGATTGACTCCTCGACGACTTCTATGGTTGGCAGAACTGCCGAGTC 1140
 A D H Q Q I D S S T T S Y G W Q N C A V

GGAAATAATGGAAATTCGACCAATTAGGGATGGGTTTTGCGGATGATCGGCGGAGCAAT 1200
 G N N G N S D Q L G M G F A D D R R S N

GAACAGTGGATGATGATGACGGACGACGTCGTCGATAATGGTGGATCGTCGGATCAAGAC 1260
 E Q W M M M T D D V V D N G G S S D Q D

AACAACTTGTTGGAATGTGGATGACGTGTGGTTCTTACAACAGTTCAGCAGCTGCTTTTAG 1320
N N L W N V D D V W F L Q Q F S S C F .

 ATATAAGAAATAGAACCACCATTCAAGGACGTACGTTATGATTGAATATTTAATGATCGA 1380

 TTACTATACATGGTTTAATTATTTAATAAACTCAAAAACCCCTAGAAAGTGAATAGATGTAA 1440
Señal de
poliadenilación
 TATTCTTGAAAAAAAAAAAAAAAAATTTTGTAGCACGTGTTTTATTTTCTTTCTTCTT 1500

 ATTACTGGTATAATTTGATCTGAATAAGATATATACCTCACGTCGAC

Figura 12. Secuencia del gen CpM5 de *Craterostigma plantagineum*. La longitud total de la secuencia que se obtuvo fue de 1558 bases. La traducción propuesta, basándose en los distintos marcos de lectura se encuentra en la parte de abajo. El probable inicio de la transcripción (ATG ó M en el caso de aminoácidos) está indicado por letras remarcadas. En la figura también se indican los pares de bases que denotan el inicio y el final de los intrones (GT y AG respectivamente). Los aminoácidos que conforman el dominio homólogo a Myb se encuentran resaltados.

En lo que se refiere al último exón, el más largo de los tres, posee una secuencia que codifica para 230 aminoácidos, incluyendo los 27 aminoácidos restantes que conforman el dominio homólogo a Myb, y la otra parte de la región codificante, la que probablemente constituye la región de transactivación del gen CpM5. El final del gen está delimitado por un codón de terminación en fase (TAG), ubicado en la posición 1320 de la secuencia. Además, 80 bases después de este codón de terminación, se encontró una posible señal de poliadenilación (AATAAA).

La comparación a nivel de ADN entre los genes CpM10 y CpM5, muestra que poseen una enorme similitud a lo largo de toda la secuencia, presentando un 98% de identidad. Las principales diferencias se observan en las regiones que corresponden a los intrones (Figura 13). Esto era de esperarse, ya que ambos genes poseen un mapa de restricción similar pero no idéntico, en donde una de las principales diferencias, es la presencia de un sitio Xba I en la región del promotor del gen CpM10, el cual no se encuentra en el CpM5 (Iturriaga et al., en preparación).

Para deducir la proteína del gen CpM5, se tradujo la secuencia en los distintos marcos de lectura posibles (figura 14). La secuencia de las dos proteínas deducidas, es casi idéntica, a excepción de cuatro residuos que se encuentran en la región que corresponde al dominio de transactivación (figura 15). Es de hacerse notar la sustitución en CpM5 de un residuo de ácido aspártico por glicina, ya que esto implica la pérdida de un sitio potencial de fosforilación, de acuerdo al consenso (S/T)XX(E/D) (Pemsan & Kemp, 1991).

Cuando se compara la región amino-terminal de la proteína deducida del gen CpM5, con los dominios tipo *myb* de genes de animales, plantas e insectos (figura 16), se puede apreciar que dicho gen presenta las dos repeticiones imperfectas de 53 y 51 aminoácidos respectivamente. Las repeticiones 1 y 2 en genes de plantas equivalen a las repeticiones 2 y 3 de los genes *myb* de animales. Dentro de estas repeticiones, se encuentran tres triptofanos separados entre 18 y 19 aminoácidos. En el caso de la tercera repetición de la mayoría de los genes *myb* de plantas, una fenilalanina sustituye al primer triptofano conservado. Sin embargo, existen varios genes como el gen C1 de maíz (Paz-Ares et al., 1987), el Atmyb2 de *Arabidopsis* (Urao et al., 1993), y los genes Am 305 y Am 340 de *Antirrhinum* (Jackson et al., 1991), que en vez de triptofano o fenilalanina contienen una isoleucina, la cual, al igual que los aminoácidos anteriores, posee un carácter no polar. Este es el caso también del gen CpM5 y también del gen CpM10 (figura 16).

	1					50
Cpm10	ACCAACACGT	CCCAACTTAG	TGATGAAGTC	AAAAAATATC	AGTCACAAGG	
Cpm5	
Cpm7	
	51					100
Cpm10	GAATATAATA	TATATGTTTT	TATATAATGA	AAACTATTA	TATGATAAAA	
Cpm5	
Cpm7	
	101					150
Cpm10	AAGAAATGAA	GCTAGAGAGC	AGAGCATCCA	TTCAACACG	GTGAAGTAAG	
Cpm5	
Cpm7	
	151					200
Cpm10	AAAGTCAAAG	TGTCAAACCT	CAAGCGCCTC	ACTGGGTAAT	GAAGAAATA	
Cpm5	
Cpm7	
	201					250
Cpm10	ATAATAATAT	TGCTTCAAAG	CTTAAAGTGC	CGTGCCTGCT	TTATTAGTAC	
Cpm5	
Cpm7	
	251					300
Cpm10	ATAAAAAATT	TGCATATTAT	ATAGTTAACT	GAAACTGCTA	TTATTCTCAT	
Cpm5	
Cpm7	
	301					350
Cpm10	CTCTCTCTCT	CTCCACCATT	TTCAAACATC	TCTGTCTCTC	TCTCCATCTC	
Cpm5	
Cpm7	
	351					400
Cpm10	TCTATGCTCT	CCCTCTCTCC	ATTTCCCAT	CTATATAAAG	CCCCACACTC	
Cpm5	
Cpm7	
	401					450
Cpm10	ACAATTAATC	CTTCTCCTTA	ATAATCTATC	TCATCTCCCC	ATCCATGAAA	
Cpm5	
Cpm7	
	451					500
Cpm10	AGCTTCCAAC	CCTTATAAGG	CATCAGCTTT	TTCTTGTCTT	TTTTTATATC	
Cpm5	
Cpm7	TTTTTCATTG	TGCTCTCATC	AATCAATCAT	CAGTTTCCTT	TTTTATTTTC	
	501					550
Cpm10	TCTTAAATCT	TTATACTATT	ATTCAAGGAA	GAAGAAAATA	CAGACATGAA	
Cpm5	
Cpm7	TTACATATCT	TTATATTAGT	ATTCAAGGAA	GAAGAAAATA	CAGACATGAA	

	551				600
Cpm10	CCAACAGCAG	...GTTAAAG	TTTCCAAGAA	CAACAAGCAA	GTTAATAACT
Cpm5	CCAACAGCAG	...GTTAAAG	TTTCCAAGAA	CAACAAGCAA	GTTAATAACT
Cpm7	CCAACAGCAG	GTTGTTAAAG	TTTCCAAGAA	CAAGCA....	..AATTAACA
	601				650
Cpm10	GTGAAGACGA	TGATGATTCT	TGGGACCTCA	GAAGAGGACC	TTGGACTGTT
Cpm5	GTGAAGACGA	TGATGATTCT	TGGGACCTCA	GAAGAGGACC	TTGGACTGTT
Cpm7	GTGAAGACGA	TGATGATTCT	TGGGACCTCA	GAAGAGGACC	GTGGACTGTT
	651				700
Cpm10	GATGAAGACT	TCACACTCAT	CAACTACATC	GCTCACCATG	GCGAAGGAAG
Cpm5	GATGAAGACT	TCACACTCAT	CAACTACATC	GCTCACCATG	GCGAAGGAAG
Cpm7	GATGAAGACT	TCACACTCAT	CAACTACATC	GCTCACCATG	GCGAAGGAAG
	701				750
Cpm10	ATGGAACTCT	CTTGACGTT	TCGCTGGTAA	TAATAAT...	ACATATTTAT
Cpm5	ATGGAACTCT	CTTGACGTT	TCGCTGGTAA	TAATAATAAG	ACATATTTAT
Cpm7	ATGGAACTCT	CTTGACGTT	TCGCT.....	
	751				800
Cpm10	ATTACATGTA	TATACACAGA	GAGAGAGAGA	GAGAGAGAGA	GAGAGAGAGA
Cpm5	ATTACATGTA	CACAC.....AGAGA	GAGAGAGAGA
Cpm7
	801				850
Cpm10	AGAGTAGTGC	ATGGATATGG	ATCCTATTTC	CTTTGTCCGT	ATTGAAAAAG
Cpm5	AGAGTAGTGC	ATGGATATGG	ATCCTATTTC	CTTTGTCCGT	ATTGAAAAAG
Cpm7
	851				900
Cpm10	AGAAAAAGGA	AAAAGAACAC	AAACTCAATG	ATAAAAGTGA	TCAAGTGGCT
Cpm5	AGAAAAAGGA	AAAAGAACAC	AAACTCAATT	AT.AAAGTGA	TCAAGTGGCT
Cpm7
	901				950
Cpm10	TATGAGAAA.AA	AGGCTTAAAG	GAAAAATCTT	ATTTTTTTT
Cpm5	TCTGAGAAA	AAAGGGCTAA	AGGAAACAAG	GGAAGATCTT	ATTTTTTTT
Cpm7
	951				1000
Cpm10	TTTTTAATTT	CTACTAATCT	TTGATCGAT	GCGAGGCCTG	AAACGAACTG
Cpm5	TTTTAGTTTC	TACTAACTTT	TTGCATCGAT	GCGAGGCCTG	AAACGAACTG
Cpm7GGCCTG	AAACGAACTG
	1001				1050
Cpm10	GAAAGAGCTG	CAGATTGAGA	TGGTTGAACT	ACTTGAGACC	CGAGGTTCTG
Cpm5	GAAAGAGCTG	CAGATTGAGA	TGGTTAAACT	ACTTGAGACC	CGAGGTTCTG
Cpm7	GAAAGAGCTG	CAGATTGAGA	TGGTTAAACT	ACTTGAGACC	CGATGTTCTG
	1051				1100
Cpm10	CGAGGAAATA	TCACACTCGA	AGAACAGCTT	CTCATTCTTG	AACTCCATTC
Cpm5	CGAGGAAATA	TCACACTCGA	AGAACAGCTT	CTCATTCTTG	AACTCCATTC
Cpm7	CGAGGAAATA	TCACACTCGA	AGAACAGCTT	CTCATTCTTG	AACTCCATTC

	1101				1150
Cpm10	ACGTTGGGGC	AATAGGTAAT	TTATATACAC	ATACGCAGT	CATAACATAC
Cpm5	TCGCTGGGGC	AATAGGTA..CACGT	CATAACATAC
Cpm7	TCGTTGGGGC	AA.....
	1151				1200
Cpm10	TTACTGTAAT	AACATATGAT	CAAAAGAAA	TTAATCTTCA	CAACGAAAT
Cpm5	TTACTGTAAT	AACATATGAT	CAAAAGAAA	TAATATTTCG	AA...GAAT
Cpm7
	1201				1250
Cpm10	AAATCCATAG	GTGGTCGAAG	ATTGCCCAAC	ATTTGCCTGG	AAGGACGGAC
Cpm5	AAATCCGTAG	GTGGTCGAAG	ATTGCCCAAC	ATTTGCCTGG	AAGGACGGAC
Cpm7TAG	GTGGTCGAAG	ATTGCCCAAC	ATTTGCCTGG	AAGGACGACG
	1251				1300
Cpm10	AATGAGATAA	AGAACTACTG	GAGAACAAGG	GTCCAAAAAC	ATGCCAAACA
Cpm5	AATGAGATAA	AGAACTACTG	GAGAACAAGG	GTCCAAAAAC	ATGCCAAACA
Cpm7	AATGAGATAA	AGAACTACTG	GAGAACAAGG	GTCCAAAAAC	ATGCCAAACA
	1301				1350
Cpm10	GCTAAAGTGC	GACGTCAACA	GCAAACAGTT	CAAAGACACC	ATGAGATACC
Cpm5	GCTAAAGTGC	GACGTCAACA	GCAAACAGTT	CAAAGACACC	ATGAGATACC
Cpm7	GCTAAAGTGC	GACGTCAACA	GCAAACAGTT	CAAAGACACC	ATGAGATACC
	1351				1400
Cpm10	TTTGGATGCC	AAGATTGGTC	GAGAGGATCC	AAGCTTCCGC	CACCACCGAC
Cpm5	TTTGGATGCC	AAGATTGGTC	GAGAGGATCC	AAGCTTCCGC	CACCACCGAC
Cpm7	TTTGGATGCC	AAGATTGGTC	GAGAGGATCC	AAGCTTCCGC	CACCACCGAC
	1401				1450
Cpm10	GACGGAGCAC	CGC.....C	GGCGGTGGCC	TCCTCGCCGT	CGTCCGCCAT
Cpm5	GACGGAGCAC	CGC.....C	GGCGGTGGCC	TCCTCGCCGT	CGTCCGCCAT
Cpm7	GACGGGCAC	CCCCTCTGCG	GTCTCGCCGC	TCCTCGCCGT	CATCCGCCAT
	1451				1500
Cpm10	GAATACAGCC	TGCTACAGCG	CTGCCATGGC	TGCAGGCGAT	CACCCTCGTC
Cpm5	GAATACAGCC	TGCTACAGCA	CTGCCATGGC	TGCAGGCGAT	CACCCTCGTC
Cpm7	GAATACAGCC	TGCTACAGCA	CTGCCATGGC	TGCAGGCGAT	CACCCTCGTC
	1501				1550
Cpm10	AGTTTATGCT	GATGCCACAG	TATTACGCGA	CGACGACGAC	GACCCACAAT
Cpm5	AGTTTATGCT	GATGCCACAG	TATTACGCGA	CGACGACGAC	GACCCACAAT
Cpm7	AGTTTATGCT	GATGCCACAG	TATTATG... ..	CGATGAC	GACCCACAGT
	1551				1600
Cpm10	AATTCGATGA	TCGCCCAAGA	AAATTCGAGC	ACCGTGGCCT	CGTCGGAATC
Cpm5	AATTCGATGA	TCGCCCAAGA	AAATTCGAGC	ACCGTGGCCT	CGTCGGAATC
Cpm7	AATTTGATGA	TCGCCCAAGA	AAATTCGAGC	ACCGTAGCCT	CGTCGGAATC
	1601				1650
Cpm10	TTTGGCAGT	CTTCTTCTG	AACTAACGGC	GGAGGCCAAC	TACGCTAACT
Cpm5	TTTGGCAGT	CTTCTTCTG	AACTAACGGC	GGAGGCCAAC	TACGCTAACT
Cpm7	TTTGGCAGT	CTTCTTCTG	AACTCA...C	GGAGGCCAAC	TACGCTAACT
	1651				1700
Cpm10	ATCACCAGT	AATCAACGGC	GCCGATCACC	AACAGATTGA	CTCCT...CG
Cpm5	ATCACCAGT	AATCAACGGC	GCCGATCACC	AACAGATTGA	CTCCT...CG
Cpm7	ATCACCAAGT	AATCAACGGC	GCCGATCATC	AACAGATTGA	CTCCTCGACG

	1701				1750
Cpm10	ACGACTTCCT	ATGATTGGCA	GAAGTGC...G	CGGTCGGAAA
Cpm5	ACGACTTCCT	ATGATTGGCA	GAAGTGC...G	CAGTCGGAAA
Cpm7	ACGACTTCCT	ATGATTGGCA	GAAGTGC...G	CGGTCGGAAA
	1751				1800
Cpm10	TAATGGAAAT	TCCGACCAAT	TAGGGATGGG	TTTTGCGGAT	GATCGGCGGA
Cpm5	TAATGGAAAT	TCCGACCAAT	TAGGGATGGG	TTTTGCGGAT	GATCGGCGGA
Cpm7	TAATGGAAAT	TCCGACCAAT	TAGGGATGGG	CTCTGTAGAT	GATCGGCGGA
	1801				1850
Cpm10	GCAATGAACA	GTGGATGATG	ATGACGGACG	ACGTCGTCGA	TAATGGTGGG
Cpm5	GCAATGAACA	GTGGATGATG	ATGACGGACG	ACGTCGTCGA	TAATGGTGGG
Cpm7	GCAATGAACA	GTGGATGATG	ATG.....G	ACGACGTCGA	TAATGGTGGG
	1851				1900
Cpm10	TCGTCGGATC	AAGACAACAA	CT...TGTGG	AATGTGGATG	ACGTGTGGTT
Cpm5	TCGTCGGATC	AAGACAACAA	CT...TGTGG	AATGTGGATG	ACGTGTGGTT
Cpm7	TCGTCGGATA	AAGACAACAA	CTTCTGTGG	AATGTGGATG	ACGTGTGGTT
	1901				1950
Cpm10	CTTACAACAG	TTCAGCAGCT	GCTTTTAGAT	ATACGAAATA	GAACCCACAT
Cpm5	CTTACAACAG	TTCAGCAGCT	GCTTTTAGAT	ATAAGAAATA	GAACCCACAT
Cpm7	CTTACAACAG	TTCAGCAGCT	GCTTTTAGAT	GGAG.....	...CCACAT
	1951				2000
Cpm10	CAAGGGACGT	ACGTTATGAT	TGAATATTTA	ATGATCAATT	ACTATACATG
Cpm5	TCAAGGACGT	ACGTTATGAT	TGAATATTTA	ATGATCGATT	ACTATACATG
Cpm7	CAAGGGACGT	ACGTTATGAT	TGAA.ATTAA	ATGATCG...ATG
	2001				2050
Cpm10	GTTTAATTAT	TTAATAAATC	AAAAACCCTA	GAAGTGAAT	AGATGTAATA
Cpm5	GTTTAATTAT	TTAATAAATC	AAAAACCCTA	GAAGTGAAT	AGATGTAATA
Cpm7	GTTTAATTAT	TTAATAAATC	AAAAACCCTA	GAAGTGAAT	AGATGTAAT
	2051				2100
Cpm10	TTCTTGAAAA	AAAAAAAAAA	ACTTTTGTGA	AGCACGTGTT	TTTATTTTTC
Cpm5	TTCTTGAAAA	AAAAAAAAAA	A.TTTTGTGA	AGCACGTGTT	TTTATTTTTC
Cpm7	TCTT.....	...GAAAAA	ACTTTTGTGA	AGCACATGTT	TTTTATTTT
	2101				2150
Cpm10	TTTCTTCTTA	TTACTTGGTA	TAATTTGATA	GAATGAAAC	TGAATAAGAT
Cpm5	TTTCTTCTTA	TTACTTGGTA	TAATTTGATA	GAAAGGAAAC	TGAATAAG..
Cpm7	TCAAAAAAAAA	AAAAAAAAAA
	2151				2200
Cpm10	ATATATACCT	CACGTCGACG	GAGAAAATTA	TGGCGAATAG	TTTAGGGTT
Cpm5	ATATATACCT	CACGTCGAC.
Cpm7
	2201				2250
Cpm10	TATTTAGGTA	TAATAATATA	TGTTGGAGAT	GGAAGCGCAT	AATAATTGAT
Cpm5
Cpm7

	2251				2300
Cpm10	TCCACCAGCT	TTGTATAATA	GTCTTTGTTG	CTTGTTATTT	TAATATATAT
Cpm5
Cpm7
	2301				2350
Cpm10	TCGCTACTTT	GTGGAATTTG	GAGACATAAT	TGGGAATGA	ACCANTTAAA
Cpm5
Cpm7
	2351			2384	
Cpm10	GCTAAGCTTC	ACATCGTTTA	ATTCTAGATT	AGCG	
Cpm5	
Cpm7	

Figura 13. Comparación entre las secuencias obtenidas de los genes Cpm5, Cpm10 y Cpm7(ADNc). Para mejorar la comparación, se introdujeron espacios dentro de las secuencias. Como se puede observar, si existen diferencias entre los tres genes. La mayoría de estas diferencias, aunque no todas se presentan en los intrones. Obviamente la secuencia del gen Cpm7 no contiene intrones, por lo que presenta una gran cantidad de espacios introducidos. Así mismo, la secuencia del gen Cpm5 no contiene la región del líder, mientras que la de los otros dos genes sí.

Marco 3 D M N E Q Q V K V S K N N K Q V N
 Marco 2 R H E P T A G . S F Q E Q Q A S L
 Marco 1: Q T . T N S R L K F P R T T S K L
 CAG ACA TGA ACC AAC AGC AGG TTA AAG TTT CCA AGA ACA ACA AGC AAG TTA
 9 18 27 36 45

N C E D D D D S S D L R R G P W T V D
 . L . R R . . F F G P O K R R T L D C .
 I T V K T M M I L R T S E E D L G L L L
 ATA ACT GTG AAG ACG ATG ATG ATT CTT CGG ACC TCA GAA GAG GAC CTT GGA CTG TTG
 60 69 78 87 96 105

E D F T L I N Y I A H H G E G R M N S
 . R L H T H Q L H R S P W R R K M E L
 N K T S H S S T T S L T M A K E D G T
 ATG AAG ACT TCA CAC TCA TCA ACT ACA TCG CTC ACC ATG GCG AAG GAA GAT GGA ACT
 117 126 135 144 153 162

L A R F A G N N N K T Y L Y M Y T Q
 S C T F R M . . D I F I L H V H T
 L L H V S L V I I I R H I Y I T C T H
 CTC TTG CAC GTT TCG CTG GTA ATA ATA AGA CAT ATT TAT ATT ACA TGT ACA CAC
 174 183 192 201 210 219

R E R E K S S A W I W I L F P L S V L
 R E R E R E E . C M D M D P I S F V R I
 R E R E R R V V H G Y G S Y F L C P Y
 AGA GAG AGA GAG AGA AGA GTA GTG CAT GGA TAT GGA TCC TAT TTC CTT TGT CCG TAT
 231 240 249 258 267 276

K K R K R K K N T N S I I K . S S G F
 E K E E K E K E H K L N Y K V I K W L
 . K R G K G K R T Q T Q L . S D O V A
 TGA AAA AGA GGA AAA GGA AAA AGA ACA CAA ACT CAA TTA TAA AGT GAT CAA GGT GCT
 288 297 306 315 324 333

E K K G . R K Q G K I L F F L F . F
 L R K K G L K E T R E D L I F F V L V
 S E K K R A K G N K G R S Y F F C F S
 TCT GAG AAA AAA AGG GCT AAA GGA AAC AAG GGA AGA TCT TAT TTT TTT TGT TTT AGT
 345 354 363 372 381 390

L L T F C I D R R P E T N M K E L Q I
 S T N F L H R S Q A . N E L E R A A D
 F Y . L F A S I A C L K R T G K S C R
 TTC TAC TAA CTT TTT GCA TCG ATC GCA GGC CTG AAA CGA ACT GGA AAG AGC TGC AGA
 402 411 420 429 438 447

E M V K L L E T R R S S R K Y H T R R
 . D G . T T D P T F V E E I S H S K
 L R W L N Y L R P D V R R G N I T L E
 TTG AGA TGG TTA AAC TAC TTG AGA CCC GAC GTT CGT CGA GGA AAT ATC ACA CTC GAA
 459 468 477 486 495 504

T A S H S . T P F S L G Q . V H V I T
 N S F S F L N S I L A G A I G Y T R H N
 E Q L L I L E L H S R M G N R Y T S .
 GAA CAG CTT CTC ATT CTT GAA CTC CAT TCT CGC TGG GGC AAT AGG TAC ACG TCA TAA
 516 525 534 543 552 561

Y L L . . H M I K R N . Y L O E L N P
 I L T V I T Y D O K K L I F A O R I K S
 H T Y C N N I S K E T N I C K N . I
 CAT ACT TAC TGT AAT AAG ATA TGA TCA AAA GAA ACT AAT ATT TGC AAG AAT TAA ATC
 573 582 591 600 609 618

V V E D C P T F A W K D G Q . D K E
 V G G R R L P N I C L E G R T M R . R
 R R M S K I A Q H L P G R T D N E I K
 CGT AGG TGG TCG AAG ATT GCC CAA CAT TTG CCT GGA AGG ACG GAC AAT GAG ATA AAG
 630 639 648 657 666 675

L L E N K G P K T C Q T A K V R R Q O
 T T G E O G S K N M P N S . S A T S O T
 N Y W R T R V Q K H A K Q L K C D V N
 AAC TAC TGG AGA ACA AGG GTC CAA AAA CAT GCC AAA CAG CTA AAG TGC GAC GTC AAC
 687 696 705 714 723 732

Q T V Q R H H E I P L D A K I G R E D
 A N S S K T P . D T F G C Q D W S R G
 S K Q F K D T M R Y L W M P R L V E R
 AGC AAA CAG TTC AAA GAC ACC ATG AGA TAC CTT TGG ATG CCA AGA TTG GTC GAG AGG
 744 753 762 771 780 789

P S F R H R R R S T A G G G L L A V
 S K L P P P P T T E H R R R W S P R R
 I Q A S A T T D D G A P P A V V S S P
 ATC CAA GCT TCC GCC ACC ACC GAC GAC GGA GCA CCG CCG GCG GTG GTC TCC TCG CCG
 801 810 819 828 837 846

R V R H E Y S L L H H C H G C R R S P S
 R P P . I Q P A T P L P W L Q A I T V
 S S A M N T A C Y T T A M A A G D H R
 TCG TCC GCC ATG AAT ACA GCC TGC TAC ACC ACT GCC ATG GCT GCA GGC GAT CAC CGT
 858 867 876 885 894 903

S V Y A D A T V L R D D D D D P Q . F
 V S L C . C H S I Y T R R R R R P T I I
 R Q F M L M P Q Y A T T T T T P H N N
 CGT CAG TTT ATG CTG ATG CCA CAG TAT TAC GCG ACG ACG ACG ACG ACC CAC AAT AAT
 915 924 933 942 951 960

D D R P R K K F E H R G L V R G I F W Q S
 R . S P K I R A P W P R V N L L A V
 S M I A Q E N S S T V A S S E S F G S
 TCG ATG ATC GCC CAA GAA AAT TCG AGC ACC GTG GCC TCG TCG GAA TCT TTT GGC AGT
 972 981 990 999 1008 1017

F F . T N G G G Q L R . L S P S N Q R
F L L N . R R R P T T L T I T E . S T
L S S E L T A E A N Y A N Y H R V I N
C T T T C T G A A C T A A C G C G G A A C T A C G C T A A C T A C C G A G T A A T C A A C
1029 1038 1047 1056 1065 1074

R R S P T D . L L D D F L W L A E L R
A P I T N R L T P R R L P M V G R T A
G A D H Q Q I D S S T T S Y G W Q N C
G G C C C G A T C A C C A A C A G A T T G A C T C C T C G A C G A C T C C T A T G G T T G G C A G A A C T G C
1086 1095 1104 1113 1122 1131

S R K . W K F R P I R D G F C G . S A
Q S E I M E I P T N . G W V L R M I G
A V G G N N G G N S D O L G M G F A D D R
G C A G T C G G A A A T G G A A A T T C C G A C C A A T T A G G G A T G G T T T G C G A T G A T C G G
1143 1152 1161 1170 1179 1188

E O . T V D D D D G R R R R . W W I V
G A M N S G . . R T T S S I N G V D R
R S N E Q W M M M T D D V V D N G G S
C G G A G A A T G A A C A G T G G A T G A T G A C G A C G A C G T C G C G A T A A T G G T G G A T C G G
1200 1209 1218 1227 1236 1245

G S R Q Q L V E C G . R V V L T T V Q
R I K T T T C G M W M T C G S Y N S S
S D Q D N N L W N V D D V W F L Q Q F
T C G G A T C A A G A C A A C T T G T G G A A T G T G G A T G A C G T G G T T C T T A C A A C A G T T C
1257 1266 1275 1284 1293 1302

Q L L L D I R N R T T I Q G R T L . L
A A A F R Y K K . N H H S R T Y V M I
S S C F . I . E I E P F F K D V R Y D
A G C A G C T G C T T T A G A T A T A A G A A C C A C C A T T C A A G A C G T A C G T A T G A T
1314 1323 1332 1341 1350 1359

N I . . S I T I H G L I I . I K N P
E Y L M I D Y Y T W F N Y L I N O K P
. I F M D R L L Y M V . L F N K S K T
T G A A T T T A A T G A T C G A T T A C T A T G T T A A T T A T T T A A T A A A T C A A A A C C
1371 1380 1389 1398 1407 1416

R S E I D V I F L K K K K K F L . A R
. K . N R C N I L E K K K K I F V S T
L E V K . M . Y S . K K K K N F C K H
C T A G A A G T G A A A T A G T A A T A T T C T T G A A A A A A A A A A A T T T T G T A A C A C
1428 1437 1446 1455 1464 1473

V F I F L S S Y Y L V . P D R K E T E
C F Y F F F F L L L G I I . K G N L
V F L F F F L L I T M Y N L I E R K L
G T G T T T T A T T T C T T C T T A T T A T T A C T T G G T A T A A T T G A T A A G A G A A C T G
1485 1494 1503 1512 1521 1530

. D I Y L T S
 I R Y I P H V
 N K I Y T S R R
 AAT AG ATA TAT ACC TCA CGT CGA C
 1542 1551

Figura 14. En esta figura se presentan los diferentes marcos de lectura que puede tener la secuencia genómica del gen CpMS. El marco de lectura correcto se encuentra subrayado.

	1				50
Cpm7	MNQQQVVKVS	KN.KQINS.E	DDDDSSDLRR	GPWTVDEFT	LINYIAHGE
Cpm10	MNQQQVVKV.S	KNNKQVNNCE	DDDDSSDLRR	GPWTVDEFT	LINYIAHGE
Cpm5	MNQQQVVKV.S	KNNKQVNNCE	DDDDSSDLRR	GPWTVDEFT	LINYIAHGE
	51				100
Cpm7	GRWNSLARFA	GLKRTGKSCR	LRWLNYLRPD	VRRGNITLLE	QLLILELHSHR
Cpm10	GRWNSLARFA	GLKRTGKSCR	LRWLNYLRPD	VRRGNITLLE	QLLILELHSHR
Cpm5	GRWNSLARFA	GLKRTGKSCR	LRWLNYLRPD	VRRGNITLLE	QLLILELHSHR
	101				150
Cpm7	WGNRWSKIAQ	HLPGRTDNEI	KNYWRTRVQK	HAKQLKCDVN	SKQFKDTMRY
Cpm10	WGNRWSKIAQ	HLPGRTDNEI	KNYWRTRVQK	HAKQLKCDVN	SKQFKDTMRY
Cpm5	WGNRWSKIAQ	HLPGRTDNEI	KNYWRTRVQK	HAKQLKCDVN	SKQFKDTMRY
	151				200
Cpm7	LWMPRLVERI	QRAATDDGA	PPLASSASSP	SSAMNTACYS	TAMAAGDHRR
Cpm10	LWMPRLVERI	QASATDDGA	PP..AVASSP	SSAMNTACYS	AAMAAGDHRR
Cpm5	LWMPRLVERI	QASATDDGA	PP..AVVSSP	SSAMNTACT	TAMAAGDHRR
	201				250
Cpm7	QFMLMPQYYA	MTTHSN..LM	IAQENSSTVA	SSESEFGLSS	ELT.EANYAN
Cpm10	QFMLMPQYYA	TTTTTHNSM	IAQENSSTVA	SSESEFGLSS	ELTAEANYAN
Cpm5	QFMLMPQYYA	TTTTTHNSM	IAQENSSTVA	SSESEFGLSS	ELTAEANYAN
	251				300
Cpm7	YHQVINGADH	QQIDSSTTS	YGWQDFAIN	TVGNNGNSDQ	LGMGSVDDRR
Cpm10	YHRVINGADH	QQIDSSTT.S	YDWQN....C	AVGNNGNSDQ	LGMGFADRR
Cpm5	YHRVINGADH	QQIDSSTT.S	YGWQN....C	AVGNNGNSDQ	LGMGFADRR
	301				335
Cpm7	SNEQWVMTD	..VDNGGSSD	KDNNLLNVD	DVWFQQFSS	CF
Cpm10	SNEQWVMTD	DVVDNGGSSD	QDNN.LNVD	DVWFQQFSS	CF
Cpm5	SNEQWVMTD	DVVDNGGSSD	QDNN.LNVD	DVWFQQFSS	CF

Figura 15. Comparación entre las proteínas deducidas de los genes Cpm5, Cpm7 y Cpm10. Para mejorar el alineamiento, se introdujeron espacios dentro de las secuencias.

Repetición 2

Repetición 3

100 106 60

1 50
Yds1 LKKGKWTQEDEQLLKAYEZHGFHMSLISMDIPG..RTDQCCKARYIEVLGPG
MuHyb LIKGPWTKEEDQRVIELVQYKPKRMSVIAKHLK.GRIGQCRERWNNHLNFE
HuHyb LIKGPWTKEEDQRVIELVQYKPKRMSVIAKHLK.GRIGQCRERWNNHLNFE
ChiHyb LIKGPWTKEEDQRVIELVQYKPKRMSVIAKHLK.GRIGQCRERWNNHLNFE
MuHybA LIKGPWTKEEDQRVIELVQYKPKRMSVIAKHLK.GRIGQCRERWNNHLNFE
LVKGPWTKEEDQRVIELVQYKPKRMSVIAKHLK.GRIGQCRERWNNHLNFE
DmHyb LIKGPWTRDEDDMVILKVRNFKGPKMTLVIARYLN..GRIGQCRERWNNHLNPN
AtGm1 RVKGFWSKEEDDVLSVQLRGLGARNWSFIARSIP.GRSKSCRLRWCNQLNPN
LRRGPTSEEDQRLVSHITNGLSCWRAPKLAGLRCGKSCRLRWNYLRF
Am315 LKRGFWTEEDQKLTSTVYLNKIGQGWRIPLKAGLSRCGKSCRLRWNYLRF
Ph1 LKKGFWTPEEDQKLLAYIEEHGHSWRALPAKAGLQRCGKSCRLRWNYLRF
Ph2 LKKGFWTPEEDQILVSYIEKHGHWRALPKLAGLRCGKSCRLRWNYLRF
Zm1 LNRGSWTQEDMRLIAYIEKHGHTNWRALPQAGLLRCGKSCRLRWNYLRF
Am306 VKKGFWTPEEDIILVSYIQEHGPGNWRAPSNITGLLRCSKSCRLRWNYLRF
Am308 TNKGAWTKEEDDLVAVYI RAHGEGCWRS LPKAAGLSRCGKSCRLRWNYLRF
Am308 TNRGAWTKEEDDLVAVYI RAHGEGCWRS LPKAAGLSRCGKSCRLRWNYLRF
Am330 TNKGAWTKEEDQRLINVI RAHGEGCWRS LPKAAGLSRCGKSCRLRWNYLRF
Hv1 TNKGAWTKEEDDLRTAYI RAHGEGCWRS LPKAAGLSRCGKSCRLRWNYLRF
Hv5 TNKGAWTKEEDQRLIAYI RAHGEGCWRS LPKAAGLSRCGKSCRLRWNYLRF
Hv31 TNKGAWTKEEDDLRTAYI KAHGCVGWSVFLPAAALMRCGKSCRLRWNYLRF
RmP1 LKRGFWTAEEDQLLANYIAEHGEGSWRS LPKLAGLRCGKSCRLRWNYLRF
Snc1 LKRGFWTSKEDDLAAYIKAHGEGKWREVPQKAGLRCGKSCRLRWNYLRF
Ph3 LKKGFWTAEEDSIMEYVKKHGEGNVAVKRNSGLMRCGKSCRLRWANHLREN
AtGL1 YKGLWTVVEEDNIMDYVLNHTGQGNRIVRKTLGRKCGKSCRLRWNYLSPN
Am305 VKKGFWTEEDLLINLYIAHGEGCWRS LPAKAGLRTGKSCRLRWNYLRF
Am340 VRKGFWTEEDDLLINFLIASHGEGVWNTLARSAGLKRRTGKSCRLRWNYLRF
AtGm2 VRKGFWTEEDAILVNFVSIHGDAFWNTIARSSGLKRTGKSCRLRWNYLRF
Cm65 LKRGFWTVEEDFTLILNVIASHGEGWNSLARAFLKRTGKSCRLRWNYLRF
Cm610

SKGRLEWVTEEDLNLSKVKAYGTKWRKISSEMFEFRPLTCKNRWRRKIITMV
VKK..TSWTEEDDRIIYQAHKRLGNRWAETIAKLLPGRTDNAIKHWNSTMRKK
VKK..TSWTEEDDRIIYQAHKRLGNRWAETIAKLLPGRTDNAIKHWNSTMRKK
VKK..TSWTEEDDRIIYQAHKRLGNRWAETIAKLLPGRTDNAIKHWNSTMRKK
VKK..SSWTEEDDRIIYEAHKRLGNRWAETIAKLLPGRTDNSIKHWNSTMRKK
VKK..SCWTEEDDRIICEAHKVLGNRWAETIAKLLPGRTDNAIKHWNSTIRRK
VKK..TAMTEKDEIYQAHLELGNQWAKIAKRLPGRTDNAIKHWNSTMRKK
LTR..NSFTEVEDKAIIAAHAIHGNKMAVIAKLLPGRTDNAIKHWNSTALRRR
LKR..GIFSEAEENLIDLHATLGNWRSIAAQLPGRTDNEIKYWNTRLKKR
LKK..GFLTMEENQIIELEHAHLGNRWSKIAQLPGRTDNEIKYWNTRLKKK
IKR..GKFTLQEQDTIQLHALLGNRWSAIATHLPKRTDNEIKYWNTHLKKR
IKR..GNFTKEEEDTIQLHEMLGNRWSIAAARLPGRTDNEIKYWNTHLKKR
LKR..GNFTDEEEAIIRLHGLLGNKWSIAACLPGRTDNEIKYWNTHLKKK
IKR..GDFTEHEEKMIHLQALLGNRWSAIASYLPHRTDNDIRKYWNTHLKKK
LKR..GNFTEEDELIIKLHSLGNKWSIAGRLPGRTDNEIKYWNTHIRRK
LKR..GNFTAEDDILLVLEHLGNKWSIAGRLPGRTDNEIKYWNTHVRRK
LKR..GNFTEEEDIIKLHSLGNKWSIAGRLPGRTDNEIKYWNTHIRRK
LKR..GNFSHEEDEIILKLSLLGNKWSIAGRLPGRTDNEIKYWNTHIRRK
LKR..GNFTDEDELIIRLHSLGNKWSIAGQLPGRTDNEIKYWNTHIRRK
LKR..GCFSQQEEDHIVALHQLGNRWSQIASHLPGRTDNEIKYWNTHIRRK
VKK..GNISKEEEDIILKHLATLGNRWSIASHLPGRTDNEIKYWNTHLSRQ
VRR..GNISYDEEDLIRLHSLGNRWSIAAGRLPGRTDNEIKYWNTRIQKH
VRR..GNITPEEQQLIMELHAKGNRWSIAKHLPGRTDNEIKYWNTRIQKH
LKK..GAFTEVEERIIELEHAHLGNKWSIAAQLPGRTDNEIKYWNTRLKKR
VKK..GNFTEQEDDLIRLHLKLLGNRWSIAKRVGRTRDQVQYWNTHLSKK
VRR..GNITPEEQQLIMELHAKGNRWSIAKTLPGRTDNEIKYWNTRIQKH
VRR..GNITPEEQQLIMELHAKGNRWSIAKHLPGRTDNEIKYWNTRIQKH
VRR..GNITLEEQQLIMELHAKGNRWSIAQYLPGRTDNEIKYWNTRIQKH
VRR..GNITLEEQLLILELHAKGNRWSIAQHLPGRTDNEIKYWNTRVQKH
VRR..GNITLEEQLLILELHAKGNRWSIAQHLPGRTDNEIKYWNTRVQKH

..+G#W₁.eED..L#.#8..h.G.G.#.#+aGL.RcgKSCRLR#NylZrp.

#+r..G.#c..E..#Yi.Lh..LGN+S.IA..lPGRTDNeIKYWNt#..++

Figura 16. En esta figura aparecen alineados los dominios tipo Myb de todos los genes descritos hasta ahora, en plantas (Am305, Am340, AtMyb2, Am315, Pp2, Ph1, Ph2, Zm1, Am308, Zm30, Hv1, Hv5, Hv33, Zm P1, ZmC1, Ph3, y AtGH), en pollo (ChiHyb), en humano (HuHyb, HuHybA, HuHybB), ratón (MuHyb), *Drosophila* (DmHyb) y levadura (Yds1). En la parte de abajo aparece el consenso (Avila et al., 1993) el cual significa: Letra mayúscula = Aminoácido completamente conservado (en algunos casos únicamente en plantas; Minúscula = Aminoácido conservado en un 80% o más; + = Aminoácido básico; - = Aminoácido ácido; # = Aminoácido hidrofóbico. Los triptofanos altamente conservados están subrayados.

Repetición 2

1 50

Ybae1 LKKGKWTQEEDQLLKAYEEHGP⁺HWLSI⁺SM⁺DPG. . RTEDQCAKRYIEVL⁺GGP
 HuMyb LIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLK. GRIGKQCRERW⁺HNHLNPE
 HuMyb LIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLK. GRIGKQCRERW⁺HNHLNPE
 ChiMyb LIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLK. GRIGKQCRERW⁺HNHLNPE
 HuMybA LIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLK. GRIGKQCRERW⁺HNHLNPE
 HuMybB LVKGPWTKEEDQRVIELVKKYGT⁺KQMTVIAKHLK. GRIGKQCRERW⁺HNHLNPE
 DrosyB LIKGPWTKEEDDMVILKLVNFKP⁺KKWTLIARYLN. GRIGKQCRERW⁺HNHLNPN
 AtMyB1 PVKGPWSKEEDVLS⁺ELVQRLLGAP⁺WNSFIARSIP. GRSGKSCRL⁺WNCQLNPN
 Pp2 LPPGPWTSEEDQMLVSHITN⁺NGS⁺WRAP⁺IKLAGLLRCGKSCRL⁺RWTNYLRPD
 Am315 LKRPWTSEEDQMLT⁺SYVLKNGIQ⁺GWRI⁺PKLAGLSRCGKSCRL⁺RWNYLRPD
 Ph1 LKKGWPTFEEDQKLLAYI: EEHGGCW⁺RALPAKAGLRCGKSCRL⁺PWTNYLRPD
 Ph2 LKKGWPTFEEDQILVSYIEK⁺NGH⁺WRALPKAGLLRCGKSCRL⁺RWTNYLRPD
 Zm1 LNP⁺GWPTFEEDMLTAYIK⁺HGHT⁺NRWALPKOAGLLRCGKSCRL⁺RWNYLRPD
 Am306 VKKGPWTFEEDIILVSYIQ⁺ENG⁺PC⁺WRAP⁺IPSTGLLRCSKSCRL⁺RWTNYLRPG
 Am308 TNKGAWTKEEDRLVAYIRAH⁺GEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 Zm38 TNKGAWTKEEDERLVA⁺YIRAHGEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 Am330 TNKGAWTKEEDQPLIN⁺YIRAHGEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 Hv1 TNKGAWTKEEDRLTAYIRAH⁺GEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 Hv5 TNKGAWTKEEDQRLIAYIRAH⁺GEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 Hv33 TNKGAWTKEEDRLTAYIRAH⁺GEGC⁺WRSLFKAAGLSRCGKSCRL⁺RWNYLRPD
 ZmP1 LKRGWPTAEEDQLLANYIAEH⁺GEGC⁺WRSLFKAAGLLRCGKSCRL⁺RWNYLRAD
 ZmC1 VKRGAWTSEEDDALAAYVKAH⁺GEGK⁺WRV⁺EPQKAGLRGKSCRL⁺RWNYLRPN
 Ph3 LKKGWPTAEESILMEYVKKH⁺GEGN⁺WNAVKRNSGLMRCGKSCRL⁺RWNYLRPN
 AtGL1 YKKGWLTVEEENILMDYV⁺LNHG⁺TGGWNRIVRKTGLKRCGKSCRL⁺RWNYLSPN
 Am305 VPKGPWTMEEDLILINIYI⁺ANHGEG⁺VWNSLARSAGLKR⁺TGKSCRL⁺RWNYLRPD
 Am304 VRKGPWTMEEDLILINFI⁺SNHGEG⁺VWNTIARSAGLKR⁺TGKSCRL⁺RWNYLRPD
 AtMyB2 LVKGPWTSEEDALLVNFI⁺HGDARWNTIARSAGLKR⁺TGKSCRL⁺RWNYLRPD
 Cpl5 LRRGPWTVD⁺EDFTL⁺IN⁺YIAHGG⁺EGRWNSLARFAGLKR⁺TGKSCRL⁺RWNYLRPD
 Cpl10 GLKRTGKSCRL⁺RWNYLRPD

Repetición 3

100 106 60

SKGRLEWTL⁺LEEDLNLSKVKAYGT⁺KWRKISSEMEF⁺PSLTCRN⁺RWRKIITMV
 VKK. . TSWTEEDRIIYQAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 VKK. . TSWTEEDRIIYQAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 VKK. . TSWTEEDRIIYQAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 VKK. . SSWTEEDRIIYEAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 VKK. . SCWTEEDRIICEAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 IKK. . TAWTEKEDEIYQAHKRLGN⁺RWAEIAKLLPGR⁺TDNAIKHNW⁺STMRK
 LIR. . NSFTVEEDKAI⁺IAAHAIHGK⁺WAVIAKLLPGR⁺TDNAIKHNW⁺SALRRR
 LKR. . GIFSEAEENILIDLHATLGN⁺RSRIAQ⁺LPGR⁺TDNEIKY⁺WNTRLKRR
 LKK. . GPLTEMEENQI⁺IELHAHLGN⁺RSKIA⁺LHPGR⁺TDNEIKY⁺WNTHLKRR
 IKR. . GKFTLOEEQT⁺IQLHALLGN⁺RSIAI⁺ATL⁺PKR⁺TDNEIKY⁺WNTHLKRR
 IKR. . GNFTKEEEDT⁺IQLHEMLGN⁺RSIAA⁺LRPGR⁺TDNEIKY⁺WNTHLKRR
 LKR. . GNFTDEEEEA⁺IYHLGLLGN⁺RSKIA⁺ACL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 IKR. . GDFTEHEEK⁺IMHQLGALLGN⁺RSIA⁺SYLPH⁺RTDNDIKY⁺WNTHLKRR
 LKR. . GNFTDEEDEL⁺IILKLSLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 LKR. . GNFTAEDDDL⁺IVKLSLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 LKR. . GNFTDEEDEI⁺IILKLSLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 LKR. . GNFTSHEEDEI⁺IILKLSLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 LKR. . GNFTDDEDEI⁺ILRLSLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTHLKRR
 VKR. . GCFSQOEDH⁺IVALHQILGN⁺RSQ⁺IASHL⁺LPGR⁺TDNEIKY⁺WNTSLKRR
 LKR. . GNI⁺SKEEEDI⁺IKLHATLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTSLKRR
 IRR. . GNI⁺SYDEEDI⁺ILRLHLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTSLGRR
 VNK. . GAFTVEEERI⁺IIEHLKALGN⁺RSIA⁺Q⁺LPGR⁺TDNEIKY⁺WNTLKRR
 LKK. . GNFTQEEDLITLRLKLLGN⁺RSI⁺AKR⁺VPGR⁺TDNDQVQRYWNTSLKRR
 VRR. . GNITPEEQLLIMELHAK⁺WNR⁺SKIA⁺KT⁺LPGR⁺TDNEIKY⁺WNTRIQKH
 VRR. . GNITPEEQLLIMELHAK⁺WNR⁺SKIA⁺H⁺LPGR⁺TDNEIKY⁺WNTRIQKH
 VRR. . GNITLEEQ⁺MLLKLHLLGN⁺RSI⁺AGL⁺LPGR⁺TDNEIKY⁺WNTRIQKH
 VRR. . GNITLEEQLLILELH⁺SRW⁺NSKIA⁺Q⁺HLPGR⁺TDNEIKY⁺WNTRVQKH
 VRR. . GNITLEEQLLILELH⁺SRW⁺NSKIA⁺Q⁺HLPGR⁺TDNEIKY⁺WNTRVQKH

..+G#W.t.eED..L#.Y#.h.g.G.W.##+aGL.RcgKSCRLRW#NyLrp.

##+.G.@t...E-.#Yi.Lh..lGN+W.S.lA..lPgrTDNEIKY#Wnt+##++

Figura 16. En esta figura aparecen alineados los dominios tipo Myb de todos los genes descritos hasta ahora, en plantas (Am305, Am340, AtMyb2, Am315, Pp2, Ph1, Ph2, Zm1, Am305, Am330, Am308, Zm30, Hv1, Hv5, Hv33, Zm P1, ZmC1, Ph3, y AtG1), en pollo (ChiMyb), en humano (HuMyb, HuMybA, HuMybB), ratón (MoMyb), *Drosophila* (DmMyb) y leavadura (Ybae1). En la parte de abajo aparece el consenso (Avila et al., 1993) el cual significa: Letra mayúscula = Aminoácido completamente conservado (en algunos casos únicamente en plantas; Minúscula = Aminoácido conservado en un 80% o más; + = Aminoácido básico; - = Aminoácido ácido; # = Aminoácido hidrofóbico. Los triptofanos altamente conservados están subrayados.

Al comparar la secuencia del gen CpM5 con diferentes bancos de datos, se encontró que el gen más relacionado, además del gen CpM10, es el gen Atmyb2 (figura 17), obteniéndose un 38 % de similitud a nivel de aminoácidos. El gen Atmyb2, se induce por el estrés hídrico, salino y en presencia de ABA, en la planta *Arabidopsis thaliana* (Urao et al., 1993). De esta manera, se realizó una comparación entre las proteínas codificadas por los dos genes (Figura 18), encontrándose homología incluso fuera del dominio tipo Myb, en una secuencia de 31 aminoácidos, adyacente a este dominio, 22 son idénticos (el 71%).

La región carboxilo-terminal tanto de la proteína deducida del gen CpM5, como del gen CpM10, con una longitud de 203 aminoácidos, son prácticamente iguales (figura 15). Esta región, que presumiblemente constituye el dominio de transactivación de la proteína, no presenta una serie de secuencias conservadas encontradas en algunos otros genes tipo myb de plantas (Jackson et al., 1991; Avila et al., 1993). Sin embargo, la proteína del gen CpM5, posee un alto contenido de serinas (24, es decir el 12%) y de treoninas (16, el 8%), estos aminoácidos pueden tener un carácter ácido, por lo que estas características son compatibles con un tipo específico de dominio de transactivación rico en dichos aminoácidos (Ptashne, 1988; Seipel et al., 1992).

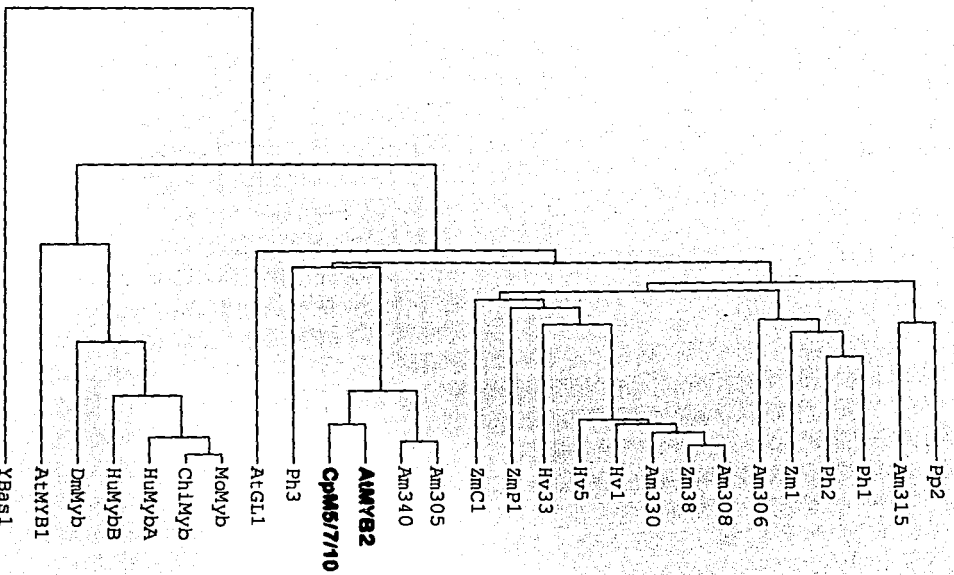


Figura 17. Dendrograma que denota la relación entre el gen CpmS y otros genes con dominio tipo Myb en plantas. La distancia a través del eje vertical es proporcional a la diferencia entre las secuencias. La distancia en el eje horizontal no tiene ningún significado.

	1					50
Atmyb2MED	YERINSNSPT	HEEDSDVRKG	PWTEEDAIL	VNFVSIHGDA	
Cpm5	MNQQVKVSK	NNKQVNNCED	DDSSDLRRG	PWTVDEDFTL	INYIAHHGGG	
	51					100
Atmyb2	RWNHIARSSG	LKRTGKSCRL	RWLNYLRPDV	RRGNITLLEEQ	FMILKLHSLW	
Cpm5	RWNSLARFAG	LKRTGKSCRL	RWLNYLRPDV	RRGNITLLEEQ	LLILELHSRW	
	101					150
Atmyb2	GNRWSKIAQY	LPGRTDNEIK	NYWRTRVQKQ	AKHLRCDVNS	NLFKETMRNV	
Cpm5	GNRWSKIAQH	LPGRTDNEIK	NYWRTRVQKH	AKQLKCDVNS	KQFKDTMRYL	
	151					200
Atmyb2	WMPRLVERIN	AQSLPTTCEQ	VESMITDPDQ	PVNEPS....	
Cpm5	WMPRLVERIQ	A.SATDDGA	PPAVVSSPSS	AMNTACYTTA	MAAGDHRRQF	
	201					250
Atmyb2	PVEPGFVQFS	QNHHQFVPA	TELSAT..SS	NSPAETFSDV	RGG.....	
Cpm5	MLMPQYYATT	TTTHNSMIA	QENSSTVASS	ESFGSLSEL	TAEANYANYH	
	251					300
Atmyb2	.VVNGSGYDF	SGQTGFGEFN	DW.....	
Cpm5	RVINGADHQQ	IDSSTTSYGW	QNCVAVGNNGN	SDQLGMGFAD	DRRSNEQMM	
	301					335
Atmyb2	...GCVGGDN	MWTDEESFWF	LQDQFCPDTT	SYSYN		
Cpm5	MTDDVVDNGG	SSPQDNNLWN	VDDVWFQQF	SSCF.		

Figura 18. Comparación entre las proteínas de los genes Cpm5 y AtMyb2 en donde se puede apreciar que existe homología entre estos genes inclusive fuera del dominio Myb (a partir del residuo 131).

DISCUSIÓN

En este trabajo, se obtuvo la secuencia y se caracterizó un gen con homología a oncogenes celulares de tipo Myb, denominado CpM5, que fue aislado de la planta de resurrección africana *Craterostigma plantagineum*. En base a esto se puede afirmar que se cumplieron los objetivos establecidos en un principio.

Si bien la secuencia que se obtuvo fué de una clona genómica, basándose en la secuencia del ADN complementario, se puede afirmar que se tiene la secuencia completa de la región codificante del gen, incluyendo los intrones y el extremo 3' no traducido. La estructura propuesta es correcta, ya que si se realiza una comparación entre la secuencia del gen CpM5 y otros genes tipo Myb de plantas, se puede observar que dicha estructura conforma el patrón típico de estos genes. Existen una excepción a este hecho, ya que el gen recientemente reportado, el MybSt1 de papa (Baranowski et al., 1994), posee un dominio Myb funcional que consta de una sola repetición.

Por lo tanto, afirmar que el gen CpM5 posee una serie de características que permiten clasificarlo como un factor de transcripción.

En primer lugar, la proteína deducida presenta en su región amino terminal un dominio homólogo a Myb característico de los genes de este tipo en plantas, en donde las dos repeticiones imperfectas poseen una gran cantidad de posiciones conservadas (Figura 16) en especial 3 triptofanos en cada repetición. Cada una de estas repeticiones conforma 3 α -hélices las cuales están separadas por vueltas parecidas a las encontradas en el homeodominio (Frampton et al., 1991). Ya se ha demostrado que tanto el dominio Myb como genes relacionados en diferentes organismos, se unen al ADN de una manera específica (Biedenkapp et al., 1988; Stober-Grässer, 1992; Urao et al., 1993; Sablowski et al., 1994). Esta unión específica es una de las principales características, de un factor de transcripción.

El hecho de que la región carboxilo de la proteína deducida del gen CpM5 no muestre una distribución de aminoácidos ácidos como el ácido aspártico o glutámico, como en el caso de la proteína del gen C1 por ejemplo (Paz-Ares et al., 1987), no implica que el CpM5 no posea un probable dominio de transactivación, ya que dicha proteína, tiene una región rica en treoninas y serinas (20%). Esto sugiere que el gen CpM5 pertenece a un grupo de factores de transcripción que presentan un dominio de transactivación rico en esos aminoácidos. Se ha demostrado que esta clase de dominio puede activar la transcripción de un gen reportero tanto a nivel proximal como a nivel remoto (Seipel et al., 1992).

Como se puede observar en la figura 15, las proteínas deducidas de los genes CpM5, y CpM10 son casi idénticas, pero el hecho de que existan varias sustituciones de aminoácidos en lo que corresponde al dominio de transactivación, permite clasificarlos como genes diferentes que se encuentran altamente relacionados.

En cuanto a la posible razón de la existencia de los dos genes no se puede afirmar nada, ya que aún no se conoce con certeza la función de ninguno de los dos. Sin embargo existe la posibilidad de que estos genes puedan actuar sobre distintos blancos a través de modificaciones distintas en su dominio de transactivación. De esta manera, la proteína deducida de CpM5, posee dos sitios potenciales de fosforilación, en contraste con 3 sitios en CpM10. Los aminoácidos fosforilados pueden regular la transactivación de distintos factores de transcripción (Hunter and Karin, '92).

Otra posibilidad sería que los pequeños cambios que presenta la proteína del gen CpM5 en su región carboxilo, provocaran que adoptara una conformación diferente lo cual le permitiera tener una función distinta, pero relacionada a la del gen CpM10, desafortunadamente, es muy difícil predecir la estructura de una proteína basándose únicamente en la secuencia de aminoácidos.

Una opción más, sería que ambos genes pudieran estar regulados diferencialmente a través de una región promotora distinta, por esta razón resultaría importante obtener la secuencia de la región promotora del gen CpM5, buscar posibles secuencias regulatorias y compararla esta secuencia con la de del gen CpM10.

La similitud del gen CpM5 con algún otro gen con homología Myb también podría proporcionar una posible idea de su función. Obviamente el gen más relacionado es el CpM10, sin embargo no se conoce exactamente la función de este gen tampoco. Se sabe sin embargo, que la expresión de este gen se induce por la aplicación de la hormona ABA, pero no por estrés hídrico (Iturriaga et al., en preparación). Esto es muy probable que suceda también con el gen CpM5 aunque todavía no se ha demostrado.

El siguiente gen más relacionado con el CpM5 es el Atmyb2 (Urao et al., 1993), el cual se expresa bajo estrés hídrico en la planta *Arabidopsis thaliana*. La similitud con la proteína de este gen no se limita únicamente al dominio homólogo a Myb, sino que al comparar ambas proteínas (figura 17) se observa una homología significativa en el dominio de transactivación. Esto podría sugerir en un principio que tanto el gen Atmyb2 como el gen CpM5 poseen una función similar, de manera que se encuentre involucrado en la regulación de la respuesta al estrés hídrico por parte de la planta *Craterostigma plantagineum*. Obviamente, todas estas aseveraciones deben de ser comprobadas o descartadas según sea el caso, y con este fin se plantean los siguientes experimentos.

En primer lugar, se requiere caracterizar a CpM5 como una proteína que se une al ADN. Para esto, es necesario expresar al ADNc en *E. coli* y por medio de SouthWestern comprobar de manera definitiva que se trata de un factor de transcripción, realizando ensayos de unión al ADN con el sitio blanco consenso de los genes *myb*.

Otro experimento importante, sería la obtención de la secuencia del promotor. Esto es necesario para verificar el inicio de la transcripción, para identificar la longitud y ubicación del líder, la posición de la caja TATA y otras secuencias regulatorias. El análisis del promotor permite identificar los posibles elementos a los cuales pueden unirse otros factores de transcripción para regular el gen. En el caso específico del gen CpM5, sería muy trascendente la identificación de elementos específicos para la hormona ABA (secuencias ABRE), ya que estos elementos se han identificado en diferentes genes que son inducidos durante el estrés hídrico. Por lo anterior se plantea para un futuro, realizar fusiones del promotor con el gen reportero GUS (Jefferson et al., 1987), para identificar en plantas transgénicas la expresión tejida específica del gen CpM5. El conocer en que condiciones se expresa el gen, e incluso en que tejidos específicamente, ayudaría enormemente a elucidar la posible función de CpM5 dentro de la planta.

Por último, podrían hacerse experimentos de genética reversa, en los que se exprese el ADNc correspondiente a CpM5 en antisentido en *Cratogeomys plantagineum*. Esto permitiría obtener mutantes de dicha planta cuyo fenotipo daría una idea más clara de la función del gen CpM5.

BIBLIOGRAFIA

1. Atchison, M. (1988) Enhancers: Mechanisms of action and cell specificity. *Ann. Rev. Cell Biol.* **4**: 127-153.
2. Avila, J., Nieto, C., Cañas, L. (1993) *Petunia hybrida* genes related to the maize regulatory C1 gene an to animal myb proto-oncogenes. *The Plant J.* **3**(4) 553-562.
3. Baranowskij, N., Froberg, C., Prat, S. (1994) A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. *EMBO J.* **13**, 5383-5392.
4. Bartels, D., Singh, M., Salamini, F. (1988) Onset of desiccation tolerance during development of barley embryo. *Planta* **175**, 485-492.
5. Bartels, D., Schneider, K., Terstappen, G. (1990). Molecular cloning of abscisic acid-modulated genes wich are induced during desiccation of the resurrection plant *Craterostigma plantagineum*. *Planta* **181**, 27-34.
6. Bartels, D., Schneider, K., Piakowski, D., Iturriaga, G. (1991) Molecular analysis of desiccation tolerance in the resurrection plant *Craterostigma plantagineum*. *Plant Mol. Biol.* **14**, 663-671.
7. Beardsley, T. (1992) Smart genes. *Sci Am.* **265**(8) 72-81
8. Biedenkapp, H., Borgmeyer, U., Sippel, A.E. (1988) Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**, 835-837.
9. Bray, E.A. (1993) Molecular responses to water deficit. *Plant Physiol.* **103**, 1035-1040.
10. Brown, J. W. (1986) A catalogue of splice junction and putative point sequences from plant introns. *Nucleic Acids Research* **14**(24) 9549-9559.

11. Brunelle, A.N., Chua, N. (1993) Transcription regulatory proteins in higher plants. *Curr. Biol.* **3**, 254-258.
12. Darnell, J. *Molecular cell biology*. Scientific American Books. New York, chapter 11, 391-448, 1990.
13. Deng, X., Matsui, M., Wey, N. (1992) COP1, an Arabidopsis regulatory gene encodes a protein with both a zinc-binding motif and a Gb homologous domain. *Cell* **71**, 791-801.
14. Devereux, J. Haerberli, P. (1984) A comprehensive set of sequence analysis programmes for the VAX. *Nucl. Acids Res.* **12**, 387-395.
15. Feinberg, A., Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13.
16. Frampton, J., Gibson, T.J., Graf, T. (1991) Proposed structure for the DNA-binding domain of the Myb oncoprotein based on model building and mutational analysis. *Protein Engineering* **4**(8) 891-901.
17. Gaff, D.F. (1971) Desiccation-tolerant flowering plants in South Africa. *Science* **174**, 1033-1034.
18. Gonda, T.J., Gough, N.M., Dunn, A.R. (1985) Nucleotide sequence of cDNA clones of murine myb proto-oncogene. *EMBO J.* **4**, 2003-2008.
19. Grotenwold, E., Athma, P., Peterson, T. (1991) Alternative spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of MYB-like transcription factors. *Proc. Natl. Acad. Sci. USA* **88**, 4587-4591.
20. Guiltinan, M.J. Marcotte, W.R., Quatrano, R.S. (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* **250**, 267-271.
21. Harrinson, S.C. (1991) Structural taxonomy of DNA-binding domains. *Nature*, **253**, 715-719.

22. Hartings, H., Maddaloni, M., Lazzaroni, N. (1989) The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J.* **8**(10) 2795-2801.
23. Hattori, T., Vasil, V., Rosenkrans, L. (1992) The Viviparous-1 gene and abscisic acid activate the C1 regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Devel.* **6**, 609-618.
24. Higgins, D.E., Sharp, P.M. (1978) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Genes* **73**: 237-244.
25. Hunter, T., Karin, M. (1992) The regulation of transcription by phosphorylation. *Cell* **70**, 375-387.
26. Iturriaga, G. (1989) The study of protein transport to the secretory pathway in plants. Ph.D. Thesis. University of Cambridge, England.
27. Iturriaga, G., Schneider, K., Salamini, F., Bartels, D. (1992) Expression of desiccation-related proteins from the resurrection plant *Craterostigma plantagineum* in transgenic tobacco. *Plant Mol. Biol.* **20**, 555-558.
28. Jackson, D., Cullanez-Macia, F., Prescott, A. (1991) Expression patterns of *myb* genes from *Antirrhinum* flowers. *The Plant Cell* **3**, 115-125.
29. Jefferson, R.A., Kavanagh, T.A., Bevan, M.W. (1987) Gus fusions: b-glucuronidase as a sensitive versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3001-3907.
30. Johnson, P.F., Mcknight, S.L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**, 799-839.
31. Ju, Q., Morrow, B.E., Wagner, J.R. (1990) REB1, a yeast DNA-binding protein with many targets, is essential for cell growth and bears some resemblance to the oncogene myb. *Mol. Cell. Biol.* **10**, 5226-5234.
32. Katagiri, F., Chua, N. (1992) Plant transcription factors: present knowledge and future challenges. *TIG* **8**(1) 22-27.

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

33. Katzen, A.L., Kornberg, T.B., Bishop, J.M. (1985) Isolation of the proto-oncogene *c-myc* from *D. melanogaster*. *Cell* **41**, 449-456.
34. Klempnauer, K.H., Gonda, T.J., Bishop, J.M. (1982) Nucleotide sequence of the retroviral leukemia gene *v-myc* and its cellular progenitor *c-myc*: The Architecture of a transduced Oncogene. *Cell* **31**, 453-463.
35. Klempnauer, K., Sippel, A. (1987) The highly conserved amino-terminal region of the protein encoded by the *v-myc* oncogene functions as a DNA-binding domain. *EMBO J.*, **6**(9), 2719-2725.
36. Latchman, D. (1990) Eukaryotic transcription factors. *Biochem. J.* **270**, 271-289.
37. Latchman, D. Eukaryotic transcription factors. Academic Press. London, Chapters 1, 2, 7 and 8, 1992.
38. Leech, M.J., Kammerer, W., Core, D.J., Martin, C., Wang, T.L. (1993) Expression of *myb*-related genes in the moss *Phycomitrella patens*. *Plant J.* **3**, 51-61.
39. Lewin, B. *Genes*. Oxford University Press, London, 845-1110, 1994.
40. Lüscher, B., Christensen, E., Litchfield, D.W., Krebs, E.G. (1990) *Myb* binding inhibited by phosphorylation at a site deleted during oncogenic activation. *Nature* **334**, 517-522.
41. Lüscher, B., Eisenman, R.N. (1990) New light on *Myc* and *Myb*. Part II. *Myb*. *Genes Devel.* **4**, 2235-2241.
42. McKnight, S.L. (1991) Molecular zippers in gene regulation. *Sci. Am.* **264**(4) 32-39.
43. Mitchell, J.M., Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding domains. *Science* **245**, 371-378.

44. Marocco, A., Wissenbach, M., Becker, D., Paz-Ares, J. (1989) Multiple genes are transcribed in *Hordeum vulgare* and *Zea mays* that carry the DNA binding domain of the myb oncoproteins. *Mol. Gen. Genet.* **216**, 183-187.
45. Nomura, N., Takahashi, M., Matusni, M., Ishii, S., Date, T., Sasamoto, S., Ishizaki, R. (1988) Isolation of human cDNA clones of myb-related genes. A-myb and B-myb. *Nucl. Acids Res.* **16**, 11075-11089.
46. Noda, K., Glover, B., Linstead, P., Martin, C. (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Science.* **369**(23) 661-664.
47. Ogata, K., Hojo, H., Aimoto, S. (1992) Solution structure of a DNA-binding domain unit of Myb: A helix-turn-helix related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. USA* **89**, 6428-6432.
48. Oppenheimer, D.G., Herman, P. (1991) A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* **67**, 483-493.
49. Pabo, C.O., Sauer, R.T. (1992) Transcription factors: Structural families and principles of DNA recognition. *Ann. Rev. Biochem.* **61**, 1053-1095.
50. Paz-Ares, J., Ghosal, D., Wienand, U. (1987) The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**(12) 3553-3558.
51. Pearson, R.B., Kemp, B.E. (1991) Protein kinase phosphorylation site sequences and consensus specificity motifs tabulations. *Methods Enzymol* **200**, 62-81.
52. Piatkowski, D., Schneider, K., Salamini, F., Bartels, D. (1990) Characterization of five abscisic acid-responsive cDNA clones isolated from the desiccation tolerant plant *Craterostigma plantagineum* and their relationship to other water-stress genes. *Plant Physiol.* **94**, 1682-1688.
53. Ptashne, M. (1988) How eukaryotic transcriptional activators work. *Nature* **335**, 683-689.

54. Ptashne, M., Gann, A. (1990) Activators and targets. *Nature* **346**, 329-331.
55. Rhodes, D., Klug, A. (1993) Zinc finger. *Sci. Am.* **266**(2) 32-39.
56. Sablowski, R., Moyano, E., Cullanez-Macia, A. (1994) A flower-specific *Myb* protein activates transcription of phenylpropanoid biosynthetic genes. *EMBO J.* **13**(1) 128-137.
57. Saikumar, P., Gabriel, J.L., Reddy, E.P. (1994) The *Myb* oncogene product induces DNA-bending. *Oncogene* **9**, 1279-1287.
58. Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl. Acad. Sci. USA* **74**, 5463-5467.
59. Schleif, R. (1988) DNA binding by proteins. *Science* **241**, 1182-1187.
60. Sheng-Ong, G.L. (1990) The *myb* oncogene. *Biochim Biophys Acta* **1032**, 39-52.
61. Seipel, K., Georgiev, O., Schaffner, W. (1992) Different activation domains stimulate transcription from remote ("enhancer") and proximal ("promoter") positions. *EMBO J.* **11**(13), 4961-4968.
62. Seto, D., (1990) An improved method for sequencing double stranded plamid DNA from minipreps using DMSO and modified template preparation. *Nucleic Acids Research* **18**(19) 5905-5906.
63. Shinozaki, K., Yamaguchi-Schinozaki, K., Urao, T. (1992) Nucleotide sequence of a gene from *Arabidopsis thaliana* encoding a *myb* homologue. *Plant Mol. Biol.* **19**, 493-499.
64. Stober-Grässer, U., Brydolf, B., Bine, X. (1992) The *Myb* DNA-binding domain is highly conserved in *Dictyostelium discoideum*. *Oncogene* **7**, 589-596.

65. Tice-Baldwin, K., Fink, G.R., Arndt, K.T. (1989) BAS 1 has a Myb motif and activates HIS 4 transcription only in combination with BAS 2. *Science* **246**, 931-935.

66. Travers, A. (1993) DNA-protein interactions. Chapman and Hall. London, 52-55, 131-157. 1993.

67. Urao, T., Yamaguchi-Schinozaki, K. (1993) An *Arabidopsis* myb homolog is induced by dehydration stress and its gene product binds to the conserved myb recognition sequence. *The Plant Cell* **5**, 1529-1539.

68. Vollbrecht, E., Veit, B., Scinha, N., Hake, S. (1991) The developmental gene Knotted-1 is a member of a meize homeobox gene family. *Nature* **350**, 241-243.

69. Wissenbach, M., Überlacker, B., Vogt, F. (1993) Myb genes from *Hordeum vulgare*: tissue-specific expression of chimeric Myb promoter/Gus genes in transgenic tobacco. *The Plant J.* **4**(3) 411-422.

70. Yancey, P.H., Clark, M.E. Hand, S.C., Somero, G.N. (1982) Living with water stress: Evolution of osmolyte systems. *Science* **217**, 1214-1222.

71. Yanisch-Perron, C., Viera, J., Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC 19 vectors. *Gene* **33**, 103-119.