



# Identificación y análisis de proteínas de *Arabidopsis thaliana* resultantes de la interacción con el ARN de la región líder del gen sacarosa fosfato sintasa de arroz

Leslie Toledo Jacobo, María del Rosario Gómez González, Gloria Solís Guzmán, Ernesto Vázquez Chimalhua, Miguel Martínez Trujillo

Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo. Edificio B4, Ciudad Universitaria. Morelia, Michoacán, México. Tel. +52(443)3223500 ext. 4224. Email: codigogenetico@gmail.com.

## PALABRAS CLAVE

proteínas;  
interacción;  
*sps1*

## RESUMEN

Se utilizaron treinta y tres clones de *Saccharomyces cerevisiae* en las que previamente se determinó la interacción del ARN de la región líder del gen Sacarosa Fosfato Sintasa de arroz (*sps1*) con proteínas de *Arabidopsis thaliana*. El ADN de las levaduras fue extraído y utilizado para transformar *Escherichia coli*, de la cual se recuperaron los plásmidos portadores de los ADNc que codificaban para las proteínas de *A. thaliana*. Las secuencias obtenidas permitieron identificar tres ADNc que se encontraban en la posición correcta y en fase de traducción con relación al codón de inicio. Entre las proteínas para las cuales codifican estos ADNc se encuentra una ribosomal y una proteína posible de tipo Spen, esta última con dominios RRM y SPOC, que permiten la unión al ARN y la interacción con proteínas, respectivamente. Se sugiere la participación de la proteína Spen en la regulación postranscripcional del gen *sps1*.

## ABSTRACT

Thirty three clones of *Saccharomyces cerevisiae* in which previously the interaction of the RNA of the Sucrose Phosphate Synthase gene with proteins of *Arabidopsis thaliana* was demonstrated, were used. The DNA of the yeasts was extracted and used to transform *Escherichia coli* to rescue the plasmids harboring the cDNA encoding the proteins of *A. thaliana*. The obtained sequences allowed identifying three cDNA that were in the correct position and in translation phase with relation to the start codon. Among the proteins encoded by the cDNAs there were a ribosomal one and a putative protein from the Spen type, this last with domains RRM and SPOC, that allow the union to RNA and the interaction with proteins, respectively. The participation of the Spen protein in the regulation of the *sps1* gene is suggested.

## KEYWORDS

proteins;  
interaction;  
*sps1*

## INTRODUCCIÓN

La regulación de la transcripción se da por la unión de proteínas a secuencias en el ADN, lo que permite la unión de la ARN polimerasa específica para que proceda la transcripción. No obstante que la regulación transcripcional es el principal mecanismo utilizado tanto en procariontes como en eucariontes, la regulación post-transcripcional permite controlar aspectos finos para que finalmente

se tenga una actividad en una proteína determinada. En la regulación postranscripcional puede darse la interacción de proteínas con el ARN mensajero, afectando ya sea la estabilidad o bien la eficiencia de traducción.

El gen de sacarosa fosfato sintasa de arroz (*sps1*) fue aislado de la subespecie Indica cv IR36 (Valdez-Alarcón *et al.* 1996) y tiene una similitud de 93% con el gen *OsSPS1* reportado en la base de datos

de la subespecie *Japonica* (<http://www.tigr.org/tdb/e2k1/osa1/>) y similitudes de 45%, 46%, 50% y 49% con los genes *OsSPS2*, *OsSPS6*, *OsSPS8* y *OsSPS11*, respectivamente, por lo que se considera que el gen *sps1* de la subespecie *Indica* estudiado en este trabajo es el ortólogo de *OsSPS1* de la subespecie *Japonica*.

La actividad de la sacarosa 6-fosfato sintasa se encuentra altamente regulada tanto al nivel transcripcional como el traduccional, siendo la modificación covalente de la enzima por fosfato en maíz y arroz dependiente de las condiciones de luz y oscuridad (Stitt *et al.* 1988, Huber y Huber 1996). La regulación al nivel transcripcional se ha reportado durante la maduración de frutos de plátano y kiwi (Olivieira *et al.* 1997, Langerkämper *et al.* 1998), en arroz durante el desarrollo de la hoja siguiendo la transición de consumidor a fuente y durante la germinación de la semilla, donde la actividad del promotor se encuentra restringida al escutelo (Chávez-Bárceñas *et al.* 2000, Martínez-Trujillo *et al.* 2004).

Las regiones líder de los genes consisten en la secuencia que se encuentra entre el sitio de inicio de transcripción y el de la traducción. Aunque la presencia de regiones líder de más de 100 bases es común en virus, es poco usual en genes celulares y generalmente están asociadas a la regulación de la expresión genética (Fütterer y Hohn 1996). Las secuencias líder pueden actuar ya sea pasivamente al no presentar barreras para el proceso de barrido de la subunidad ribosomal 40S, o bien por un mecanismo activo a través de un factor regulador que actúa en *trans*. Este último mecanismo se ha estudiado en el Virus del Mosaico del Tabaco, donde el líder omega interacciona con una proteína de 102 kDa (Tanguay y Gallie 1996) que ha sido identificada como la proteína celular HSP101 inducible por choque térmico. HSP101 funciona como un potenciador de la traducción y su actividad es regulada negativamente por limitación de nutrientes, sin que se afecte su función de tolerancia a altas temperaturas (Wells *et al.* 1998). El gen *sps1*, tiene un líder inusualmente largo de 368 nucleótidos, con una estructura secundaria compleja y tiene un efecto en la potenciación de la traducción tanto *in vivo* en plantas de *A. thaliana* como *in vitro* en un sistema de extractos de germen de trigo (Martínez-Trujillo *et al.* 2003). Sin embargo, se desconoce la forma en la cual

la región líder de este gen participa en la potenciación de la traducción.

Las interacciones de ARN y proteínas son críticas para una amplia variedad de procesos biológicos desde la proliferación de ciertos virus hasta las decisiones del desarrollo en organismos pluricelulares. Por esta razón, se han desarrollado varios métodos para analizar las interacciones, siendo el más reciente el de tres híbridos, que permite determinar las interacciones *in vivo*, ya sea confirmando interacciones previas *in vitro* o revelando algunas nuevas (Bernstein *et al.* 2002, Jaeger *et al.* 2003, Hook *et al.* 2005, Riley *et al.* 2006). El sistema de tres híbridos además de la ventaja de determinar las interacciones *in vivo* permite identificar las proteínas de interacción mediante la secuenciación de los genes que las codifican y tiene como fundamento la interacción de tres moléculas híbridas (dos proteínas y un ARN) que permiten el reclutamiento de la ARN polimerasa II y la expresión de un gen reportero (SenGupta *et al.* 1996) (Figura 1).

Utilizando una biblioteca de expresión de ADNc de *A. thaliana* se realizó un escrutinio mediante el sistema de tres híbridos para buscar proteínas que interaccionaran con la región líder del gen *sps1* y se obtuvieron 33 colonias que crecieron en un medio selectivo con alta astringencia (Martínez-Trujillo *et al.* en prensa). En este trabajo se recuperaron los plásmidos de las levaduras con los ADNc clonados y fueron secuenciados para determinar la identidad de las proteínas codificadas y su posible función.

## **MATERIALES Y MÉTODOS**

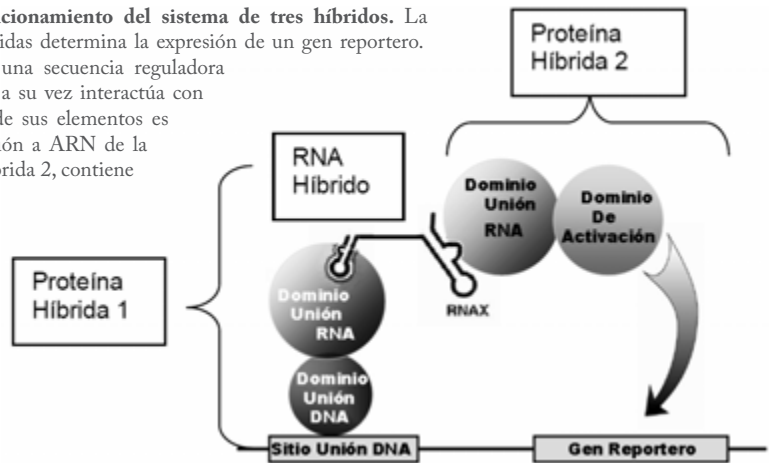
### **Aislamiento de ácidos nucleicos**

El ADN total de levadura fue aislado de acuerdo a Amberg *et al.* (2005). El ADN de plásmidos fue aislado de *Escherichia coli* mediante las técnicas de Mini-prep y Maxi-prep de acuerdo a Sambrook y Russell (2001).

### **Transformación de *Escherichia coli***

Se transformaron células competentes de *E. coli*

**FIGURA 1. Fundamento del funcionamiento del sistema de tres híbridos.** La interacción de tres moléculas híbridas determina la expresión de un gen reportero. La proteína híbrida 1 se une a una secuencia reguladora río arriba del gen reportero. Ésta a su vez interactúa con el ARN híbrido, en el que uno de sus elementos es reconocido por el dominio de unión a ARN de la proteína híbrida 2, contiene un dominio de activación que completa la naturaleza modular de este factor de transcripción sólo si el dominio de unión de ésta interactúa con el ARN del ARN híbrido.



DH5 $\alpha$  siguiendo el protocolo descrito a continuación: El DNA extraído de la levadura se dializó y se adicionaron 2  $\mu$ l de éste a 50  $\mu$ l de células competentes. Se almacenó en hielo por 1 minuto y posteriormente se pasaron a celdas de electroporación Eppendorf de 100  $\mu$ l. Se dio un choque eléctrico de 1800 voltios en un electroporador Eppendorf modelo 2510 y posteriormente las células se resuspendieron en un ml de medio YENB (Sambrook y Russell 2001). Las células se incubaron a 37 °C durante una hora sin agitación y otra hora con agitación. Se sembraron las células en medio LB con carbenicilina a una concentración de 100  $\mu$ g/ml y se incubaron a 37 °C durante 24h para que ocurriera el crecimiento de colonias.

### Corte y electroforesis del ADN plasmídico

Para la digestión de los plásmidos aislados por la técnica de Mini-prep se utilizaron las siguientes condiciones: 7  $\mu$ l de ADN plasmídico aislado de *E. coli*, 1  $\mu$ l de agua, 1  $\mu$ l de Buffer 10X y 1  $\mu$ l de enzima *EcoR*1. Las reacciones se incubaron a 37 °C por 4 hrs. Para la digestión de los plásmidos aislados por la técnica de Maxi-prep, se utilizaron las siguientes condiciones: 1  $\mu$ l de ADN plasmídico aislado de *E. coli*, 7  $\mu$ l de agua, 1  $\mu$ l de Buffer 10X y 1  $\mu$ l de la enzima *EcoR*1. Posteriormente se corrieron 10  $\mu$ l de la reacción de digestión con 3  $\mu$ l de la mezcla de azul de

bromofenol y ARNasa en un gel de agarosa al 1%. Los gels se colocaron en una cámara de luz ultravioleta y se fotografiaron.

### Secuenciación

Se utilizó el método de terminación de cadenas, usando dideoxinucleótidos fluorescentes (ABI) y un secuenciador de ADN ABI PRISM374. Se solicitaron los servicios del CINVESTAV-IPN, Unidad Irapuato.

### Análisis de secuencias

Las secuencias de los ADNc de los plásmidos se utilizaron para alinearse con secuencias de la base de datos del servidor National Center of Biotechnology Information, mediante el uso del programa nBLAST.

### Análisis de dominios de proteínas

Los dominios conservados de proteínas fueron obtenidos en la base de datos de Pfam del Instituto Sanger (Bateman *et al.* 2002) y del National Center of Biotechnology Information.

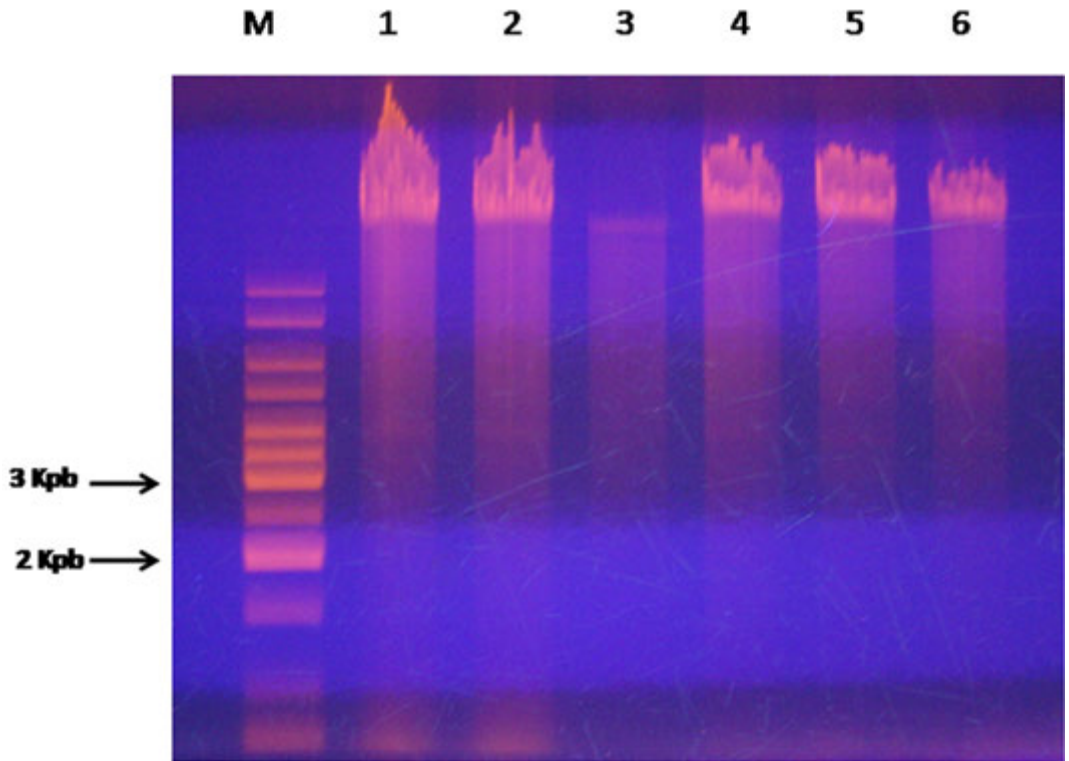


FIGURA 2. ADN total aislado de 6 clonas diferentes de la levadura L40-*ura3*. El ADN fue separado en un gel de agarosa al 1% y teñido con bromuro de etidio.

## RESULTADOS Y DISCUSIÓN

Se utilizaron 33 colonias de la cepa L40-*ura3* de *Saccharomyces cerevisiae*, las cuales fueron transformadas con los plásmidos pHybLex/Zeo-MS2 y pSPS-MS2 y posteriormente con una biblioteca de ADNc de *Arabidopsis thaliana* clonados en el vector pGAD10; las clonas fueron seleccionadas en medio con zeocina, carente de uracilo e histidina, por lo que potencialmente manifestaron una interacción del ARN de la región líder del gen *sps1* con algunas proteínas codificadas en el ADNc (Martínez-Trujillo *et al.*, en prensa). Las 33 colonias de levadura fueron utilizadas en este trabajo para extraer el ADN total y recuperar los plásmidos con los ADNc. En la figura 2 se presenta la extracción de ADN total de 6 colonias diferentes de levadura, en la que se observa el ADN cromosómico en la parte superior y bandas en la parte central que corresponden a los plásmidos con los que

estaban transformadas las levaduras.

### Transformación de *E. coli* y aislamiento de plásmidos

Se transformó la cepa DH5 $\alpha$  de *E. coli* con el ADN total extraído de cada clona de levadura. Los plásmidos de la biblioteca de ADNc derivados de pGAD10 confieren resistencia a carbenicilina, así como el plásmido pMS2-SPS, mientras que pHybLex/Zeo-MS2 carece de esta capacidad. Las bacterias transformadas fueron seleccionadas en medio LB con carbenicilina; entre las colonias de *E. coli* con crecimiento en medio con el antibiótico se encontraron las que contenían los plásmidos de la biblioteca de ADNc. En la figura 3 se presenta la extracción de plásmidos de 10 colonias de *E. coli* resultantes de la transformación con el ADN total de la clona 79 de levadura y su posterior selección en

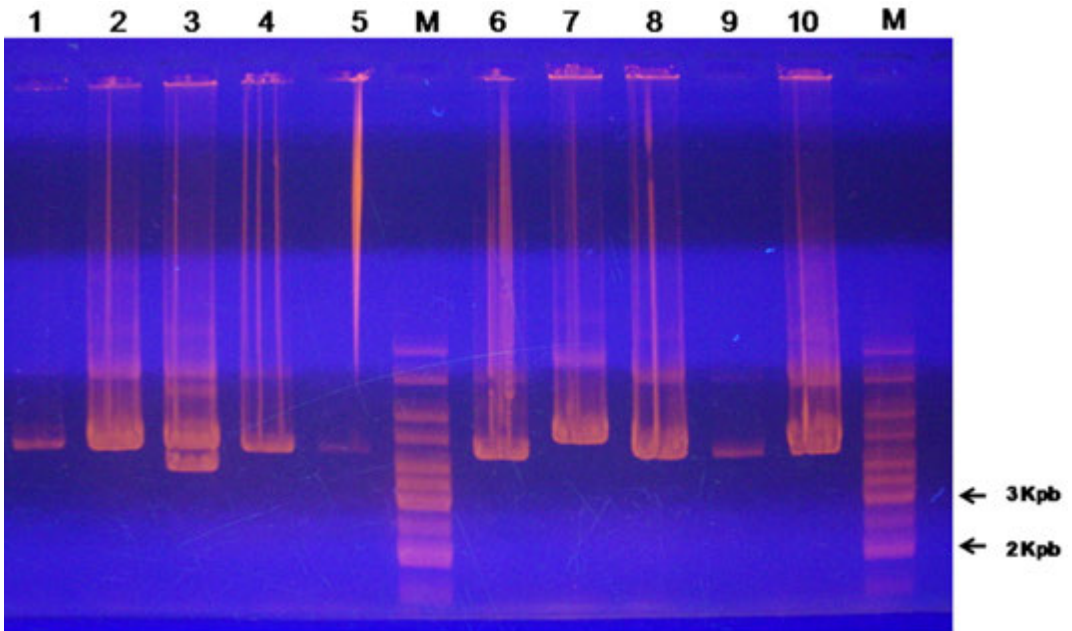


FIGURA 3. ADN de plásmidos aislados de *E. coli* transformada con ADN total de la clona 79 de levadura. Los plásmidos fueron aislados por la técnica de Mini-prep y separados en geles de agarosa al 1%.

medio con carbenicilina. Se observa variación en la migración del ADN, por lo que se asumió que había diferentes plásmidos.

#### Corte de plásmidos con *EcoRI* y análisis de fragmentos por electroforesis

Los plásmidos aislados de *E. coli* fueron cortados con *EcoRI* de acuerdo a materiales y métodos. Las mezclas de ADN cortado fueron separadas por electroforesis para determinar la presencia de los vectores cortados y la liberación de fragmentos. El vector pSPS-MS2 libera un fragmento de 7907 pb y otro de 400 pb, mientras que los plásmidos con los ADNc liberan un fragmento de 6650 (vector pGAD10) y fragmentos que pueden ser variables de acuerdo al tamaño del ADNc. Con esta consideración fue posible distinguir los plásmidos de cada uno de los dos tipos, principalmente con base en la presencia de fragmentos superiores a 400 pb. En la figura 4 se presenta la separación de los fragmentos de ADN de los plásmidos cortados con *EcoRI*, provenientes de la transformación

con el ADN de la clona 79 de levadura. El candidato 7 libera además del fragmento correspondiente al vector, fragmentos de aproximadamente 2000 pb y 700 pb, por lo que se consideró que este caso era adecuado para portar un ADNc. Este último plásmido se utilizó para transformar *E. coli* y posteriormente aislar ADN por Maxi-prep para secuenciar el ADNc. El mismo criterio se utilizó para recuperar los plásmidos con ADNc provenientes de otras clonas de levaduras.

#### Secuenciación, análisis de la orientación y análisis del marco de lectura

Los plásmidos con los ADNc recuperados fueron aislados por Maxi-prep y secuenciados de acuerdo a materiales y métodos. Las secuencias se utilizaron para ser alineadas con secuencias conocidas en la base de datos de *Arabidopsis thaliana* del servidor del National Center of Biotechnology Information. Se utilizó el programa nBLAST proporcionado en el servidor mencionado. Además, se determinó si el ADNc estaba en la orientación correcta, ya que podría



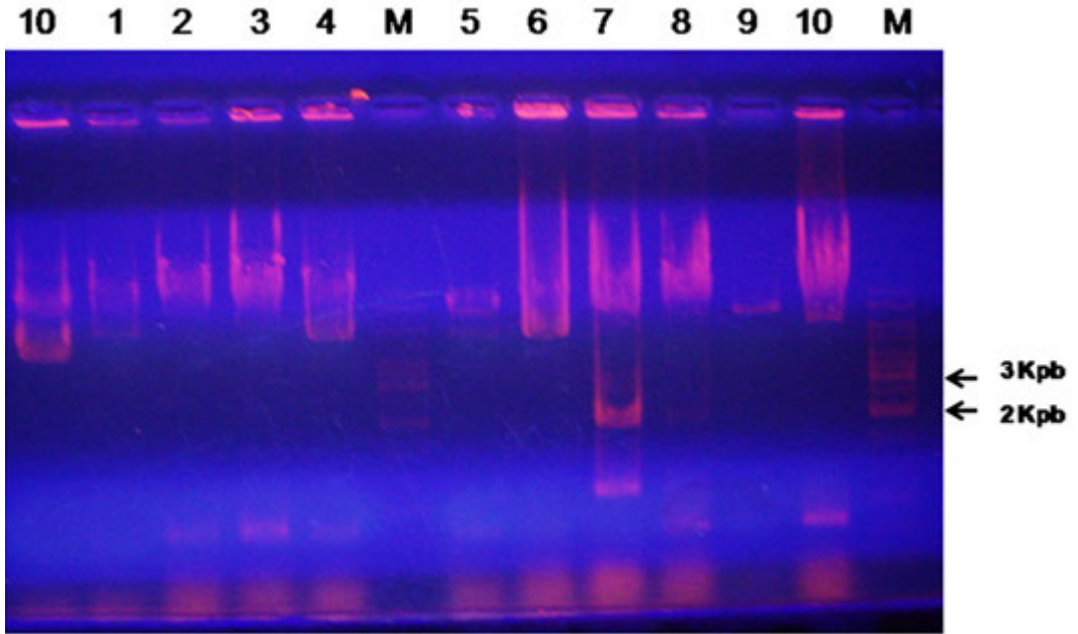


FIGURA 4. Digestión de los plásmidos con *EcoR1* aislados de colonias de *E. coli*, transformada con el ADN de la clona 79 de *S. cerevisiae*. Los plásmidos fueron aislados por la técnica de Mini-prep y separados en geles de agarosa al 1%.

estar invertido con relación al promotor a partir del cual se genera la expresión en el vector pGAD10. En los ADNc que tuvieron la orientación correcta se analizó el marco de lectura, ya que sólo una de tres posiciones es la correcta para que la traducción del ARN mensajero codificado por el ADNc se dé de manera adecuada. Una unión desfasada del ADNc con el marco de lectura proveniente del codón de inicio ocasionó una lectura incorrecta y una proteína diferente. En el cuadro 1 se presenta la relación de proteínas codificadas por los ADNc de los plásmidos aislados. De las nueve proteínas identificadas, 5 ADNc se encontraron en la orientación inversa. De las 4 proteínas con los ADNc orientados correctamente, tres se encontraron en el marco de lectura adecuado, NM\_125810, NM\_117319 y NM\_117334.

### Análisis de las proteínas

#### *Proteína codificada por el locus NM\_125810*

Corresponde a la proteína S28E, que es un

componente de la subunidad ribosomal 30S. Es probable que la S28E controle al precursor del empalme del ARN y el reciclaje en el ARNm durante el proceso de maduración, sin embargo, su función en el ribosoma es desconocida. Un análisis de dominios conservados permitió identificar el dominio de unión al ARN parecido a S1. Este dominio se encuentra en una amplia variedad de proteínas asociadas a ARN. El gen RPS28 está conservado en humanos, perros, vacas, ratas, ratones, pollo, el pez zebra, mosquito, arroz y en *Arabidopsis thaliana*. (Chang *et al.* 2005, Carrol *et al.* 2008).

La secuencia de la proteína es la siguiente:

mdsqikhavvkvvmgrtgrsgqvvtqrvkftdsdrfimirnkvpgpvre-gdvltilserearrlr

#### *Proteína codificada por el locus NM\_117319*

Corresponde a un inhibidor de proteasa, que participa en almacenamiento en semilla. La secuencia encontrada es la siguiente:

masknsaslalfalnilfflttagtncrcnpskprplpnpkvpskvp-tpsvpsvpyvptpsvpspsvptpsvpspsvpspnptvipprtgssg-

CUADRO 1. Secuencias Identificadas en la secuenciación de plásmidos

Locus	ARN codificado	Observaciones de la clonación	Observaciones de la proteína
NM_117188	ARNm lineal de 737 pb	Orientada de manera incorrecta en el vector pGAD10	Factor de iniciación IF-1
NM_125810	ARNm lineal de 480 pb	Orientada correctamente en el vector pGAD10 y en el marco de lectura adecuado	Proteína ribosomal 28S; componente estructural del ribosoma
NM_117319	ARNm lineal de 786 pb	Orientada correctamente en el vector pGAD10 y en el marco de lectura adecuado	Inhibidor de proteasa/ almacenamiento de semillas/ transferencia de lípidos. De la familia LTP
NM_122009	ARNm lineal de 1049 pb	Orientada correctamente en el vector pGAD10. El marco de lectura es incorrecto para la traducción	Proteína nuclear relacionada con Ras
NM_128268	ARNm lineal de 2501 pb	Orientada incorrectamente en el vector pGAD10	Factor de transcripción SE
NM_120116	ARNm lineal de 1513 pb	Orientada incorrectamente en el vector pGAD10	De la familia F-box
NM_129704	ARNm lineal de 1157 pb	Orientada incorrectamente en el vector pGAD10	Probable proteína
NM_120952	ARNm lineal de 2300 pb	Orientada incorrectamente en el vector pGAD10	Probable proteína LepA de enlace-GTP
NM_117334	ARNm lineal de 2802 pb	Orientada correctamente en el vector pGAD10 y en el marco de lectura adecuado	Probable proteína

ncpidalrlgvcavnlgllnvqlgqpsppccsliqglvlddaavclctar  
 lanvlglnlvpislsvlncnrllpsnfqca

Esta secuencia es similar a la secuencias Os03g0103300 de *Oryza sativa* subespecie *Japonica* y OsI\_009422 de *Oryza sativa* subespecie *Indica*. Se relaciona con la familia HPS y la superfamilia AAI\_LTSS. La familia HPS está compuesta por proteínas similares a HPS, la cual es una proteína pequeña e hidrofóbica de función desconocida relacionada con el inhibidor de alfa-amilasa y proteínas transportadoras de lípidos. Además de las HPS, se incluye en esta subfamilia a las siguientes proteínas: a) Proteína híbrida rica en prolina (HyPRP) proveniente del maíz, b) Proteína inducible en oscuridad (LeDI-2) proveniente de *Lithospermum erythrorhizon*, c) Proteína del maíz ZRP3 y, d) Proteína del arroz RcC3. HyPRP es una proteína específica del embrión que contiene un dominio N-terminal rico en prolina y un dominio similar a HPS y un extremo terminal rico en cisteína. Se ha sugerido que HyPRP podría estar involucrado

en la estabilidad y como defensa del embrión en desarrollo. LeDI-2 es una proteína específica de la raíz que está probablemente involucrada en la regulación de la biosíntesis de los derivados de *L. erythrorhizon*. ZRP3 del maíz y el RcC3 del arroz son proteínas específicas de la raíz cuya función aun no se ha determinado. La superfamilia AAI\_LTSS comprende proteínas únicas de plantas avanzadas, que incluye los inhibidores de alfa-amilasa del tipo cereal, proteínas de transferencia de lípidos, proteínas de almacenamiento de semilla y proteínas similares. Las proteínas de esta familia son conocidas por participar en importantes papeles en la defensa de insectos y patógenos en plantas, el transporte de lípidos entre las membranas intracelulares y el almacenamiento de nutrientes. Varias proteínas de esta superfamilia han sido identificadas como causantes de alergias para los humanos. (Baud *et al.* 1993, Yazaki *et al.* 2001, José-Estanyol *et al.* 1992).

**Proteína codificada por el locus NM\_117334**

Es una proteína hipotética de función desconocida de 823 aminoácidos (Figura 5). Considerando que el ARN líder *sps1* proviene del arroz subespecie *Indica* se buscaron proteínas similares en arroz y se encontró que existe una identidad del 51% con la proteína hipotética OsI\_029748 en la subespecie *Indica*. Esta proteína tiene en teoría 1012 aminoácidos y también se desconoce su función (Figura 5). En ambas proteínas se encontraron dos dominios RRM (ARN recognition

motif) hacia el extremo amino, y un dominio SPOC (Spen paralogue and orthologue C-terminal) en la región central de la proteína (Figura 5).

El dominio RRM también conocido como dominio de unión a ARN (RBD) o dominio ribonucleoproteico (RNP), es muy abundante en eucariontes. Este dominio se conforma de 90 aminoácidos y consiste de un paquete de cuatro láminas beta y dos hélices alfa. RRM regularmente interactúa con los ARN de cadena sencilla (ssRNA), pero además interacciona con ADN

```

1 mssrgrrerm mkedgrgrnp psrhlwvgnl phgilerela drflrfgele slafqgrsy
61 afvnfnhded afaaieslqg fpalsqnlri efakaeksst gsrtddiyrh deqrsaargs
121 sfvqrdsrmr yespdtyks kmndrnaeps evlyiqfpas lkvdallrn vfssfgeitk
181 vtvfqgrsya fvqfrnlmaa ckakeslqgk lfgnprvhic faksepsssg sgrgpgsrsl
241 sppyrsvdrl gssegylqdr nygsisrips vrephyiedr dledsegif nrkrdsssdg
301 gpaygrsrst hrpqdmhey hgspgemgts frdnphrfqt rsseyepwd lpeddyyyqe
361 ikrlktrssq perqlpghql sgieqerrpf srasadfspk dafernyeag qlrynqtveq
421 plnlairngd ksslrephee lmggyplpsv vperkrytpe ldrpslkdw wegtiakqgn
481 picrakcfpv qkvmddmlpe fldctartql dmlakhyys skawvfvfp gsdadivfyd
541 efmhyleekq raavsklddt ttflfvpsd fsekvkvpq klsisgvil lecggsgsgg
601 pvqqqgerkd tdlttyget syseasgafp dvgnprpgp tafllrsagr nqsasmdpyv
661 edkhdqlshr ysgsdwpprd tnprsspid htvqkhsgrv pgkqgnadls ryhdtetpvp
721 agfqpeqlth lasslprqqq qvqntpnqpe ryapegrasf shlqhaqtps ipqlvtpnqg
781 nvqiqssnsq qqeeteanpq krlqatlqla aallqqiqqa kps

```

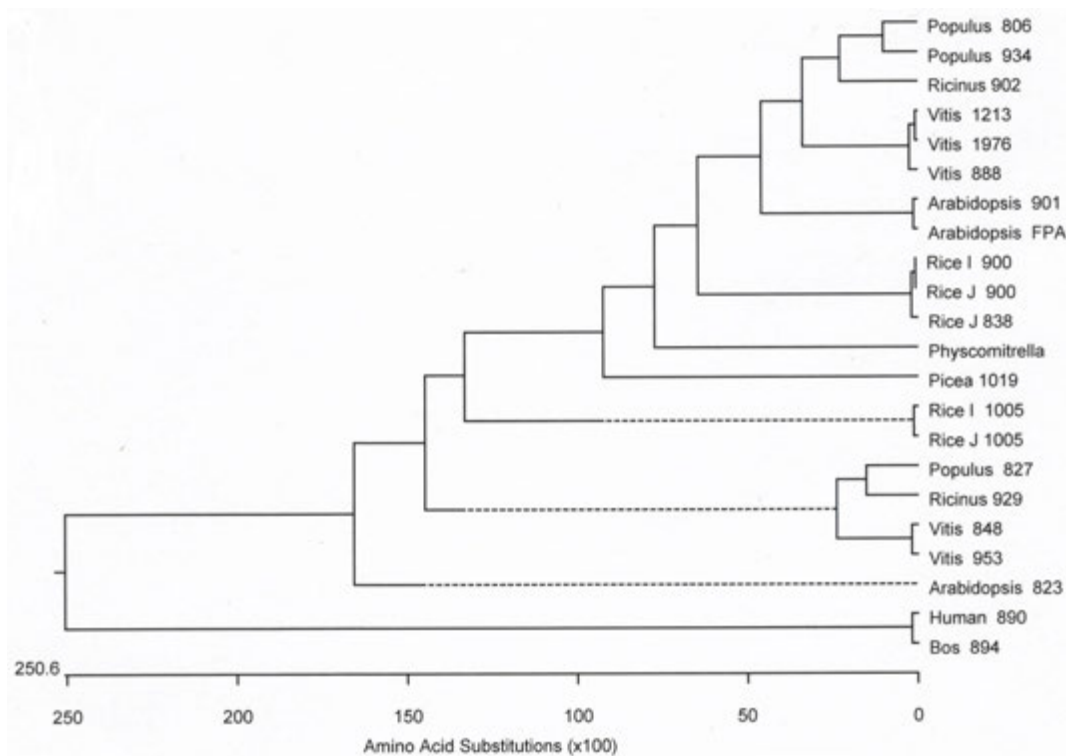
```

1 meyrDIMGR prgrgggggg grgrfggggg srfsaarddp pprsssgwg vappsrhlwv
61 qslspgvaaa dlselflrcq dveqisrdpq rsfafvtfar eedavaavre lqqihlrgap
121 iriefskgdk gssssmddry sqhadqrrft ergrnqqssp ekstdkskrs rpaepsevlw
181 igfpvqlkvd eatlweafsp fgevvkittf pgrtyafvgy ttiaaacrak etlqgnlfnn
241 prvsicfrs dsvsaefgkg sldapsyphl nssvrpifre qdfedfprar pfdspprdmv
301 mpsphygpk lsrdhdvfg srdnylrygp gvepdprsnf epfrirglgp errmsedpye
361 qhrrspagda pwhnipfers qgalpledsr yaredpypfs kklrtgeahd selpeypfse
421 fdrqkvgsay prprfygvpd ddihprgyql apmhgrnhvd plrnptplvd rhphghaqds
481 fsrhvevers tpeyhpllk eewkwdgtia kggtpicrar cfpvqkvlfn mlpefldcta
541 rtslemiskh yyqaasswv ffvpendadm aaynefmyl gdkqraavck lgerssflv
601 ppsdfseql rvpgkvisg vilkfeqsdp evssptrkne tfvshlnhd rahedldalr
661 rinpdirpl pqgsdylgls pgsynpasah lvppykfgna psyleelah qkhpdshre
721 iahdkyqqhp dvlpsrwsdn inynspgsgn lnylaesaip htstdrtpea ysfapqgvpk
781 vstsgyapva deasnmsypp mqqpasqqvvr pqqppslpls lppeqlaqla tllaqqnqpg
841 kepvdslnke sgfirnpgh ssmmphssgs ipvqnsllpa pssasqlqv appvqgsvpp
901 npsimhtpna pmpshntlpl ppmhpsgnpa hssmplrsfv pplpegpspl rqtssalqa
961 qpalsgpgqt sqqsagedh hgdpqkrlqa tlqlaatllq qiqqskpgg qk

```

FIGURA 5. Secuencias de aminoácidos de las proteínas Spen hipotéticas de *Arabidopsis thaliana* y arroz. Arriba. Proteína Spen de *A. thaliana*. Las secuencias de 25 a 91 y 152 a 219 (subrayadas) representan dominios RRM. La secuencia de 471 a 543 (subrayada) representa un dominio SPOC. Abajo. Proteína Spen de *Oryza sativa*. Las secuencias de 58 a 124 y 189 a 245 (subrayadas) representan dominios RRM. La secuencia de 505 a 577 (subrayada) representa un dominio SPOC.





**FIGURA 6. Árbol filogenético de las proteínas Spen en plantas.** La proteína codificada por el locus NM\_117334 de *A. thaliana* fue utilizada para buscar proteínas similares en la base de datos del National Center of Biotechnology Information y la presencia de dos dominios RRM y un dominio SPOC fue determinada por predicción en la base de datos Pfam. Las proteínas fueron alineadas con el programa Align de Lasergene para elaborar el árbol filogenético con la opción Clustal W. Los nombres se refieren al género o nombre común de la planta y número de aminoácidos de la proteína.

de cadena sencilla (ssDNA) así como con proteínas. Aunque la estructura de este dominio presenta cuatro cadenas beta y dos alfa, existen variaciones que lo hacen altamente versátil para unirse con diferente afinidad y especificidad al ARN, lo que permite desempeñar una multitud de funciones biológicas frecuentemente asociadas con la regulación post-transcripcional de los genes (Maris *et al.* 2005). Entre las funciones que se han encontrado para proteínas con este dominio se encuentra el splicing y la regulación traduccional por PABP (Maris *et al.* 2005).

El dominio SPOC consiste en 7 láminas beta formando un barril central y dos hélices pequeñas y se reporta como un dominio de interacción proteína-proteína, implicado en apoptosis y cáncer en animales, aunque los mecanismos moleculares no son claros

(Sánchez-Pulido *et al.* 2004). Las proteínas con combinación de dominios RRM y SPOC son conocidas como Spen y se han encontrado en gusanos, moscas y vertebrados. En ratones, la proteína MINT (MSx2 interactor nuclear) está involucrada en el desarrollo neuronal y esquelético, mediando la represión por el represor transcripcional con homeodominio Msx2. Se ha reportado que la estructura del dominio SPOC de la proteína SHARP es capaz de interactuar con los correpresores SMRT/NCOR (silencing mediator for retinoid and thyroid receptors); estos correpresores reprimen el nivel basal de transcripción mediando el ensamblaje de un gran complejo multiproteico que contiene histonas desacetilasas. (Sánchez-Pulido *et al.* 2004; Ariyoshi y Schwabe, 2003, Maris *et al.* 2005, Kuang *et al.* 2000).

Con el incremento de secuencias de genes y predicción de proteínas en las bases de datos es posible determinar la presencia de proteínas Spen en plantas y sus similitudes con la proteína de *Arabidopsis thaliana* encontrada. La figura 6 muestra que especies de *Oryza sativa*, *Vitis vinifera*, *Ricinus communis*, *Populus trichocarpa* y *Picea sp* tienen proteínas Spen con dos dominios RRM y un dominio SPOC con similitud entre ellas. Como controles externos se incluyeron las proteínas de humano y de toro.

## CONCLUSIONES

Se identificaron nueve proteínas, de las que sólo las codificadas por los loci NM\_125810, NM\_117319 y NM\_117334 se encontraron en la dirección adecuada en el vector y en el marco de lectura correcto.

La proteína ribosomal S28 tiene un dominio para unirse a ARN por lo que es lógico que haya interactuado con el ARN de la región líder *sp1*.

La interacción de la proteína que se relaciona con el inhibidor de alfa-amilasa y proteínas transportadoras de lípidos resultó un poco extraña y podría tratarse de una interacción que carece de significado biológico.

La interacción de la proteína Spen resultó interesante considerando que posee tanto dominios para unión a ARN (RRM) como un dominio de interacción con proteínas (SPOC). Además, la presencia de proteínas similares en arroz de donde es nativa la región líder *sp1*, permite predecir que también la proteína de este cereal mostrará interacción con la región líder.

La búsqueda de más proteínas Spen mostró que se encuentran presentes en diferentes géneros de angiospermas y en gimnospermas, por lo que deben estar asociadas a funciones comunes en las plantas, entre las que no se puede descartar la regulación genética y particularmente la regulación del gen *sp1*.

## REFERENCIAS

Amberg, D., Burke, D. J. y J. N. Strathern (2005). *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005

Edition. Cold Spring Harbor Laboratory. New York. 230 pp.

Ariyoshi, M. y J. W. R. Schwabe (2003). A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes and Development* 17: 1909-1920.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy S. R., Griffiths-Jones, S., Howe, K.L., Marshal, M. y E. L. Sonnhammer (2002). The Pfam protein families database. *Nucleic Acids Research* 30: 276-280.

Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. y M. S. (1993). Crystal structure of hydrophobic protein from soybean; a member of a new cysteine-rich family. *Journal of Molecular Biology* 231(3):877-887.

Bernstein, D. S., Buter, N., Stumpf A. y M. Wickens (2002). Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods* 26: 123-141.

Carroll, A. J., Heazlewood, J. L., Ito, J. y A. H. Millar (2008). Analysis of the Arabidopsis cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification. *Molecular Cell Proteomics* 7(2):347-69.

Chang, I. F., Szick-Miranda, K., Pan, S. y J. Bailey-Serres (2005). Proteomic characterization of evolutionarily conserved and variable proteins of Arabidopsis cytosolic ribosomes. *Plant Physiology* 137(3):848-62.

Chávez-Bárceñas, A. T., Valdez-Alarcón, J. J., Martínez-Trujillo, M., Chen, L., Xoconostle-Cázares, B., Lucas, W. J. y L. Herrera-Estrella (2000). Tissue-specific and developmental pattern of expresión of the rice *sp1* gene. *Plant Physiology* 124: 641-653.

Fütterer, J. y T. Hohn (1996) Translation in plants-rules and exceptions. *Plant Molecular Biology* 32: 158-189.

Hook, B., Bernstein, D., Zhang, B. y M. Wickens (2005). RNA-protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. *RNA* 11: 227-233.

- Huber, S. C. y J.L. Huber (1996). Role and regulation of sucrose-phosphate synthase in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 47: 431-444.
- Jaeger, S., Eriani, G. y F. Martin (2003). Results and prospects of the yeast three-hybrid system. *FEBS Letters* 556: 7-12.
- Josè-Estanyol, M., Ruiz-Avila, L. y P. Puigdomènech (1992). A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. *Plant Cell* 4(4):413-423.
- Kuang, B., Wu, S. C., Shin, Y., Luo, L., y Kolodziej, P. (2000). Split ends encodes large nuclear proteins that regulate neuronal cell fate and axon extension in the Drosophila embryo. *Development*, 127(7), 1517-1529.
- Langerkämper, G., McHale, R., Gardner, R. C. y E. MacRae (1998). Sucrose-phosphate synthase steady-state mRNA increases in ripening
- Maris, Ch., Domínguez, C. y H. T. A. Frédéric (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* 272: 2118-2131.
- Martínez-Trujillo, M., Chávez-Barcenas, T., Limones-Briones, V., Simpson, J. y L. Herrera-Estrella (2004) Functional analysis of the promoter of the rice sucrose phosphate synthase gene (*sps1*). *Plant Science* 166: 131-140.
- Martínez-Trujillo, M., Chávez-Barcenas, T., Limones-Briones, V. y L. Herrera-Estrella (2003). Functional analysis of the 5' untranslated region of the sucrose phosphate synthase rice gene (*sps1*). *Plant Science* 165: 9-20.
- Oliveira do Nascimento, J. R., Cordenunsi, B.R., Lajolo, F.M. y M.J.C. Alcocer, (1997). Banana sucrose phosphate synthase gene expression during fruit ripening. *Planta* 203: 283-288.
- Riley, K. J. L., Cassiday, L. A. y A. Kumar (2006). Recognition of RNA by the p53 tumor suppressor protein in the yeast three-hybrid system. *RNA* 12: 620-630.
- Sambrook, J. y D. Russell (2001). *Molecular Cloning: A Laboratory Manual*. Col Spring Harbor Laboratory Press. New York.
- Sánchez-Pulido, L., Rojas, A.M., van Wely, K.H., Martínez, A.C. y A. Valencia (2004). SPOC: A widely distributed domain associated with cancer, apoptosis and transcription. *BMC Bioinformatics* 5:91. Doi: 10.1186/1471-2105-5-91.
- SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P. y S. Fields (1996). A three hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci* 93: 8496-8501.
- Stitt, M., Wilke, I., Feil, R. y W. Heldt (1988). Coarse control of sucrose phosphate synthase in leaves: Alterations of the kinetic properties in response to the rate of photosynthesis and the accumulation of sucrose. *Planta* 174: 217-230.
- Tanguay, R. y D. Gallie (1996) Isolation and characterization of the 102-Kilodalton RNA-binding protein that binds to the 5' and 3' translational enhancers of Tobacco Mosaic Virus RNA. *Journal of Biological Chemistry* 271: 14316-14332.
- Valdez-Alarcón, J. J., Ferrando, M., Salerno, G., Jiménez-Moraila, B. y L. Herrera-Estrella (1996). Characterization of a rice sucrose-phosphate synthase-encoding gene. *Gene* 170: 217-222.
- Wells, D. R., Tanguay, R. L., Le, H. y D. R. Gallie (1998). HSP101 functions as a specific translational regulatory protein whose activity is regulated by nutrient status. *Genes and Development* 12: 3236-3251.
- Yazaki, K., Matsuoka, H., Shimomura, K., Bechthold, A. y F. Sato (2001). A novel dark-inducible protein, LeDI-2, and its involvement in root-specific secondary metabolism in *Lithospermum erythrorhizon*. *Plant Physiology* 125(4):1831-1841.