

Full Length Research Paper

A 22-bp sequence of the core promoter from the *Indica* rice sucrose-phosphate synthase gene (*sps1*) is sufficient to confer basal transcription activity

Miguel Martínez-Trujillo¹, Gamaliel Valdivia-Rojas¹, Gloria Solís Guzmán¹, José Luis Cabrera-Ponce²

¹Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo. Morelia, Michoacán, México

²Departamento de Ingeniería Genética de Plantas, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Irapuato, Guanajuato, México.

Accepted 29 July 2008

The 5' region of the sucrose-phosphate synthase gene (*sps1*) of the *Indica* rice consists of an atypical promoter which lacks TATA box but has a putative initiator sequence overriding a second transcription initiation site. Analysis of the transient expression of truncated versions of the *sps1* promoter (from -148 to +21) fused to the *uidA* reporter gene was performed. The results showed that a stretch of 22 bp (GTGTCACCCGCCAGCCTCCCT), from -1 to +21, is sufficient to confer basal transcription activity. These data suggest that an initiator-like sequence (TCACCC) is the responsible of this basal activity.

Key words: Core promoter, *sps1* gene, expression analysis.

INTRODUCTION

Transcription of structural genes is directed by the corresponding promoters whose DNA sequence determines timing, strength, direction and position of transcriptional initiation. In eukaryotic cells, the promoter elements from protein-coding genes are classified into two categories: common core promoter elements, which are needed for basal transcription initiation, and gene-specific regulatory elements located upstream and necessary to regulate the strength and specificity of the expression (Roeder, 1996). The core promoter is located at or near the transcription initiation site and a TATA box is essential in most cases, with the TATAa/tAa/t consensus sequence located 25 to 30 bp upstream of the transcription initiation site (Nikolov and Burley, 1997). However, an increasing number of genes lacking a TATA box-like sequence at the expected positions have been described in *Drosophila* and human genomes (FitzGerald et al., 2006). The initiator element (INR) is a sequence that overlaps the transcription initiation site and may compensate the absence of a TATA box (Roeder, 1991); however, when the TATA box and the INR motif are both present, the second element de-

termines the position of the transcription initiation site (Zhu et al., 1995). The consensus sequence of INR in mammals is a motif pyrimidine-rich (PyPyAN(t/a)PyPy), where A is at +1 (Smale and Baltimore, 1989). A genome comparison of humans, mouse, rice and *Arabidopsis thaliana* showed that INR is more represented in the promoters of mammals than those from plants, and that the consensus sequence is restricted to the positions -1/+1 (the YR rule) with Y (C or T) at the -1 position and R (A or G) at +1 (Yamamoto et al., 2007). This INR consensus sequence suggests an origin in the early evolving eukaryotes, and has also been reported in protists (Liston and Johnson, 1999).

RNA polymerase II is assembled in the promoter after several general factors are recruited (Orphanides, 1996), being the first a complex of 8 - 12 proteins associated to the TATA binding protein (TBP), named collectively as TFIID (Burke and Kadonaga, 1996). The primary function, either of the TATA box or the INR is to locate TFIID in the core promoter resulting in formation of the TFIID/DNA complex; whereas the TATA box is recognized directly by the TBP (Hernández, 1993), the INR motif is recognized by factors associated to TBP (TAF_{II}150-TAF_{II}250) (Chalkey and Verrijzer, 1999). Some promoters lacking TATA box have substitutes known as downstream promoter elements (DPEs) around

*Corresponding author: E-mail: codigogenetico@gmail.com.
Tel-Fax: +52 (443) 3166973

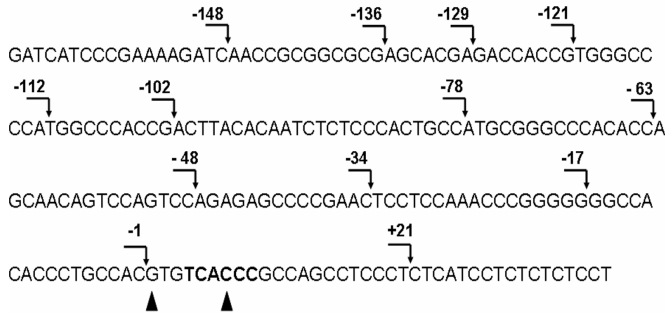


Figure 1. Deletions at the *Sps1* core promoter gene. The arrows indicate the 5' deletions generated in the promoter. The transcription initiation sites (*tis*) are indicated with arrowheads; the first *tis* is considered as +1 for convenience. A typical initiator motif is highlighted in bold nucleotides. The pSPS1 plasmid (Martinez-Trujillo et al., 2004) with the -2196 *sps1* promoter fused to the *uidA* gene (Jefferson et al., 1987) was used as a template to amplify DNA fragments by PCR and to generate the different truncated promoter versions. PCR fragments were cloned into the pGEM-T-easy vector (PROMEGA) and the constructs were verified by sequencing with the dideoxynucleotide chain termination method (Sanger et al., 1977) using fluorescent nucleotides (ABI) and an ABI PRISM™ DNA sequencer.

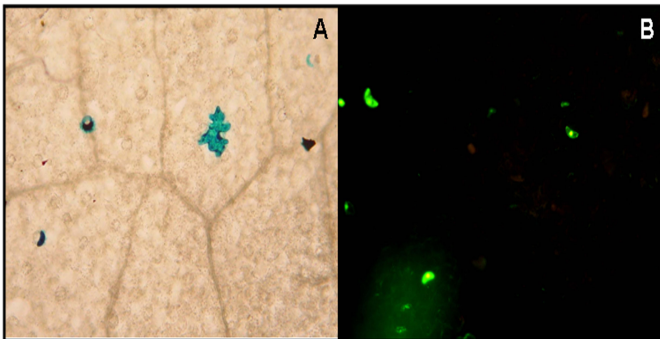


Figure 2. Expression of *uidA* and green fluorescent protein genes in tobacco bombarded leaves. Tobacco leaves in MS medium (Murashige and Skoog, 1962), were bombarded with a mix of the plasmids containing the *sps1* promoter versions fused to the *uidA* gene, and the plasmid pCKGFPS65C (Reichel et al., 1996) with the gene encoding the green fluorescent protein under the control of the CaMV 35S promoter. A BioRad particle bombardment device Helium-driven PDS-1000/He, at 900 psi, was used. (A) Expression of the *uidA* reporter gene was visualized as blue cells after staining the tobacco leaves with x-gluc, according to Stomp (1992). (B) Expression of the green fluorescent protein gene was visualized as green fluorescent cells by using a fluorescence microscope Leica DMRE with a filter I3 (blue light exciter BP 450 - 490 nm; beamsplitter RKP 510 nm; emitter LP 515 nm). Images were photographed at 400X.

+30 with the consensus sequence a/gGa/tCGTG. These elements were described first in *Drosophila* and are recognized by TAF_{II}40 (Burke and Kadonaga, 1997). This DPE motif has been reported in plants but at a lower frequency than in *Drosophila* and mammals (Yamamoto et al., 2007). In plants, the majority of photosynthesis

nuclear genes have TATA-less promoters suggesting that their core promoters are activated in a better way by light (Nakamura et al., 2002).

Sucrose is synthesized in plants by the action of sucrose-6-phosphate synthase (SPS; EC 2.3.1.14), an enzyme which transfers the glucosyl moiety from UDP-glucose to fructose-6-P to produce sucrose-6-P, which is in turn dephosphorylated by a sucrose-6-phosphate phosphatase (SPP; EC 3.1.3.00) yielding sucrose as the final product and making the reaction practically irreversible (Huber and Huber, 1996). The sucrose phosphate synthase gene (*sps1*) of the *Indica* rice has a TATA-less promoter with GC-rich regions and two transcription initiation sites (Valdez-Alarcón et al., 1996). This gene encodes for the major SPS isoform in leaves (Castleden et al., 2004). When the *sps1* promoter region from -2196 fused to the *uidA* reporter gene was used to transform rice plants, there was expression in the mesophyll of leaves, seed scutellum and pollen grains, and its expression in leaves was increased by light (Chávez-Barcenas et al., 2000). Afterwards, Martinez-Trujillo et al. (2004) generated 5' deletions in the *sps1* promoter fused to the *uidA* reporter gene and the constructs were used to bombard tobacco leaves; transient expression was determined in the -148 truncated promoter and it was similar to the expression conferred by the complete promoter. In this work, the purpose was to determine the importance of the sequences downstream -148 in the basal expression conferred by the *sps1* promoter and to define the minimal sequence that still confers this basal expression.

RESULTS AND DISCUSSION

To determine the essential sequences of the core promoter involved in expression of the *sps1* gene, twelve 5' deletions were generated from -148 to +21 (Figure 1). These deleted promoters were fused to the *uidA* gene and the expression level conferred by these regions was determined by a transient expression assay. We observed that the *sps1* promoter region directs a tissue-specific and light-regulated expression in leaves of transgenic tobacco plants; this finding was similar to that observed in rice plants (data not shown). We have previously used this system to determine the strength of the *sps1* promoter as compared with the CaMV 35S promoter (Martínez-Trujillo et al., 2004).

The bombarded tobacco leaves with the different *sps1* constructs were analyzed to determine the transient expression, visualized as blue cells due to the expression of the β -glucuronidase enzyme (Figure 2A). A construct with the 35S constitutive promoter fused to the green fluorescent protein gene, was also transferred to the tobacco leaves and the expression was visualized as green cells (Figure 2B). The ratio of blue/green fluorescent cells was determined to normalize the results. The transient expression obtained for the different *sps1* constructs are shown in Figure 3. The constructs from -148 to -1 showed

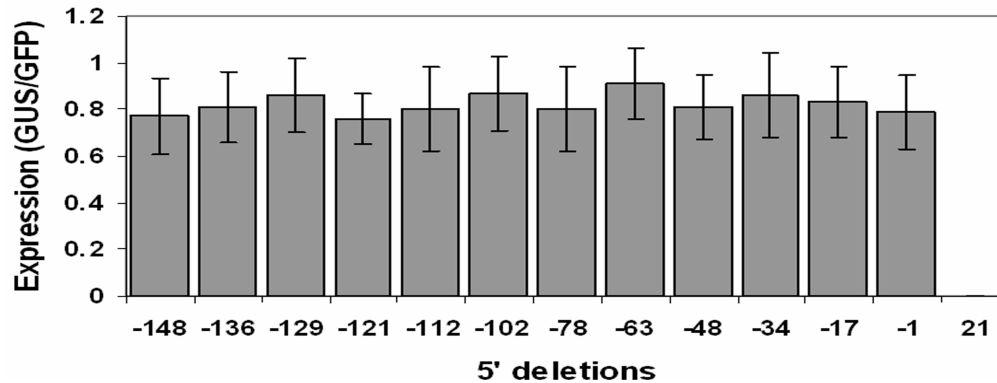


Figure 3. Level of expression conferred by the *sps1* promoter versions. Plasmids containing the *sps1* promoter deletions fused to the *uidA* gene, mixed with the plasmid containing the GFP gene fused to the 35S promoter, were delivered into tobacco leaves as described in the legend of figure 2. Ten leaf samples were bombarded with each plasmid mixture. The ratio of blue cells/fluorescent cells was determined for each *uidA* construct. The standard errors are shown as bars.

similar expression levels, but the +21 construct had no expression; two additional bombardment repetitions were made with similar results.

The elimination of the *sps1* GC-rich motifs (from -148 to -1), reported in mammals as sites for the Sp1 recognition factor, did not modify the level of expression conferred by the *sps1* promoter (Figure 3), in agreement with the finding that the Sp1 element is not associated with plant promoters (Yamamoto et al., 2007).

It has been previously demonstrated that elimination of the downstream region beyond +23 conserves the *sps1* promoter expression (Martínez-Trujillo et al., 2003), which combined with the present results, show that the minimal region of the *sps1* promoter capable to confer activity is delimited between -1 to +21 (GTGTCACCCGCCAGCCTCCCT). As two transcription initiation sites have been determined previously (Valdez-Alarcón et al., 1996) (Figure 1), the observed transient expression in the shorter construct (-1) can derive from the second site. Taking this second site as +1, the sequence would be -6 to +15 and the interactions of general transcription factors as well as further binding of the RNA polymerase II would be possible. The precise sequence or sequences responsible for the activity of the -1 to +21 *sps1* promoter construct might be considered as an initiator, since the second transcription initiation site is overlapped by pyrimidines and is a typical initiator motif (TCACCC) following the rule of a purine at +1 and a pyrimidine at -1. Another rice gene with a similar initiator (TCCAAG) is that reported for the phenylalanine ammonia-lyase enzyme (PAL) with a function in determining the precise transcription initiation site (Zhu et al., 1995).

The 22-bp sequence (-1 to +21 or -6 to +15) capable to confer basal transcription activity in the *sps1* promoter is not an isolated case, since a 15-bp sequence of the hepatitis B virus gene, without a TATA box, was reported to direct precise transcription initiation (Chen et al.,

1995). The possibility that a DPE may be involved in the basal activity of the *sps1* promoter is low, since the reported consensus sequence (g/aGt/aCGTG) (Burke and Kadonaga, 1996) is not present downstream of the second transcription initiation site.

Our results demonstrate that a 22-bp sequence is sufficient to confer basal expression by the *sps1* promoter, and that the consensus initiator sequence in this stretch possibly determines this basal activity.

ACKNOWLEDGEMENTS

This research was supported by grants from the Coordinación de Investigación Científica (UMSNH) and CONACYT. We are grateful to Dr. Carlos Cervantes for revision of the manuscript.

REFERENCES

- Burke TW, Kadonaga JT (1996). *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* 10: 711-724.
- Burke TW, Kadonaga JT (1997). The downstream promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* 11: 3020-3031.
- Castleden CK, Auki N, Gillespie VJ, McRae EA, Quick WP, Buchner P, Foyer CH, Furbank RT, Lunn J (2004). Evolution and function of the Sucrose-Phosphate Synthase gene families in wheat and other grasses. *Plant Physiol.* 135: 1753-1764.
- Chalkey G, Verrijzer CP (1999). DNA binding site selection by RNA polymerase II TAFs: a TAK₁250-TAF₁150 complex recognizes the Initiator. *EMBO J.* 18: 4835-4845.
- Chávez-Bárcenas AT, Valdez-Alarcón JJ, Martínez-Trujillo M, Chen L, Xoconostle-Cázares B, Lucas WJ, Herrera-Estrella L (2000). Tissue-specific and developmental pattern of expression of the rice *sps1* gene. *Plant Physiol.* 124: 641-653.
- Chen IH, Huang CJ, Ting LP (1995). Overlapping initiator and TATA box functions in the basal core promoter of Hepatitis B Virus. *J. Virol.* 69: 3647-3657.
- FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson Ch (2006). Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* 7: R53.

- Hernández N (1993) TBP, a universal eukaryotic transcription factor. *Genes Dev.* 7: 1291-1308.
- Huber SC, Huber JL (1996). Role and regulation of sucrose-phosphate synthase in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47: 431-444.
- Jefferson RA, Kavanagh TA, Bevan MW (1987). GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6: 3901-3907.
- Kriwacki RW, Schultz SC, Steitz TA, Caradonna JP (1992) Sequence-specific recognition of DDA by zinc-finger peptides derived from the transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* 89: 9759-9763.
- Liston DR, Johnson PJ (1999). Analysis of a ubiquitous promoter element in a primitive eukaryote: early evolution of the initiator element. *Mol. Cell Biol.* 19: 2380-2388.
- Martínez-Trujillo M, Chávez-Bárceñas T, Limones-Briones V, Simpson J, Herrera-Estrella L (2004). Functional analysis of the promoter of the rice sucrose phosphate synthase gene (*sps1*). *Plant Sci.* 166: 131-140.
- Martínez-Trujillo, M., Limones-Briones, V., Chávez-Bárceñas, T., Herrera-Estrella L. (2003). Functional analysis of the 5' untranslated region of the sucrose phosphate synthase rice gene (*sps1*). *Plant Sci.* 165: 9-20.
- Murashige T, Skoog K (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol Plant.* 15: 473-497.
- Nakamura M, Tsunoda T, Obokata J (2002). Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. *Plant J.* 29: 1-10.
- Nikolov DV, Burley SK (1997). RNA polymerase II transcription initiation: a structural view. *Proc. Natl. Acad. Sci. USA* 94: 15-22.
- Orphanides G, Lagrange T, Reinberg D (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* 10: 2657-2683.
- Reichel C, Mathur J, Eckes P, Langenkemper K, Koncz C, Schell J, Reiss B, Maas C (1996). Enhanced green fluorescence by the expression of an *Aquorea victoria* green fluorescent protein mutant in mono and dicotyledonous plant cells. *Proc. Natl. Acad. Sci. USA* 93: 5888-5893.
- Roeder RG (1991). The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem. Sci.* 16: 402-407.
- Roeder RG (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* 21: 327-335.
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- Smale ST, Baltimore D (1989). The "initiator" as a transcriptional control element. *Cell* 57: 103-113.
- Stomp AM (1992). Histochemical localization of β -glucuronidase. In Gallagher SR (ed), *GUS protocols: using the GUS gene as a reporter of gene expression*. Academic Press, San Diego CA, pp 103-113.
- Valdez-Alarcón JJ, Ferrando M, Salerno G, Jiménez-Morilla B, Herrera-Estrella L (1996). Characterization of a rice sucrose-phosphate synthase-encoding gene. *Gene* 170: 217-222.
- Yamamoto YY, Ichida H, Abe T, Susuki Y, Sugano S, Obokata J (2007) Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Ac. Res.* 35: 6219-6226.
- Zhu Q, Dabi T, Lamb C (1995). TATA box and Initiator functions in the accurate transcription of a plant minimal promoter in vitro. *Plant Cell* 7: 1681-1689.