S41-004

Core promoter architecture characteristic of photosynthesis nuclear genes

M. Nakamura[1], T. Tsunoda[2], T. Yoshitsugu[1], H. Hibino[1], J. Obokata[1]

[1]Center for Gene Research, Nagoya University, Nagoya 464-8602, JAPAN. *Fax: +81-52-789-3081 E-mail: obokata@gene.nagoya-u.ac.jp*

[2]*SNP Research Center, RIKEN, Tokyo 108-8639, JAPAN*

Keywords: photosynthesis genes, core promoter, initiator, TATA-less promoter

Introduction

Plant genomes encode tens of thousands of genes, and their expressions are basically controlled at the level of transcription. Because promoters play a critical role in transcriptional regulation of the genes, huge numbers of studies have been done on their structure and function. However, this description is not exactly true. Promoters of the eukaryotic protein-coding genes are composed of two parts: common core promoter elements needed for basal transcription initiation, and gene-specific regulatory elements located upstream (Roeder, 1996). Though huge numbers of studies have been done on the latter elements, we have very little knowledge about the former elements at least in the plant genome.

Most of the eukaryotic core promoters contain TATA box located around -25 relative to the transcription start site. TATA box is thought to be essential for the basal transcription initiation. However, a portion of eukaryotic genes is known to lack a TATA box (Azizkhan *et al.*, 1993). In some TATA-less promoters, Initiator (Inr) element is thought to compensate the lack of TATA box. Inr is rich in pyrimidine residues, and located over the transcription start site (Smale *et al.*, 1998). Combinations of TATA box and Inr possibly generate various core promoter subtypes (Roeder, 1996; Roeder, 1998).

In this study, we attempted to grasp the general characteristics of the photosynthesis gene promoters in the plant genome. We initiated this study by the detailed analysis of the promoter of a tobacco photosystem I gene, *psaDb*. This gene is transcriptionally activated by light stimuli (Yamamoto *et al.*, 1995; Yamamoto *et al.*, 1997), and canonical TATA box is not present in the core promoter region (Yamamoto *et al.*, 1993). Analysis of this gene revealed that the Inr motif is essential for this promoter activity, and it can not be replaced with TATA box. This finding raised a new question whether TATA-less promoter is unique to *psaDb* or common to certain groups of plant genes. In the later half of this manuscript, we address this subject from the viewpoint of plant genomics.

Materials and Methods

Plant material

Seeds of *Nitociana tabacum* cv. SR1 were surface-sterilized and sown on 0.8% agar-plate containing half-strength MS salts (Murashige and Skoog, 1962). 5 day-old etiolated seedlings were used for particle bombardment.

Construction of chimeric genes

5' deletion derivatives - The GUS-coding regions of the *psaDb* (-1722)::GUS and its 5' deletion derivatives (Yamamoto *et al*., 1997) were replaced with the firefly luciferase (LUC) coding region.
35S::RLUC - The GUS-coding region of pBI221 (Jefferson *et al*., 1987) was replaced with *Renilla reniformis* luciferase (RLUC) coding region.
Other constructs - 18 linker scanning constructs, 4 initiator mutation constructs, and 4 TATA-Inr swap constructs were made by PCR, using specifically designed primers.

Particle bombardment and Luciferase assay

Particle bombardment was performed under the dim green light as previously described (Miyamoto *et al.*, 2000). Bombarded seedlings were incubated in either continuous white light (70 μE m^{-2} s^{-1}) or continuous darkness for 24 h, followed by the determination of LUC and RLUC activities using the Dual Luciferase Reporter Assay System (Promega). LUC activities of the *psaDb* chimeric constructs were normalized by RLUC activities for correcting the efficiency of gene delivery in each bombardment.

Computer analysis of plant promoters

Nucleotide sequences of 232 plant promoters were compiled from the Eukaryotic Promoter Database (Cavin-Périer *et al*., 2000) and several other sources, and subjected to computer analysis. Presence or absence of TATA-boxes was predicted by use of an algorithm designed for identifying *cis*-regulatory elements (Tsunoda and Takagi, 1999).

Results

Effect of DNA methylation on transient expression of psaDb::LUC

In transient expression assays, methylation state of the introduced DNA changes the expression characteristics of the introduced genes (Terres *et al.*, 1993; Rogers and Rogers, 1995). We first examined whether the light-responsive transcription of *psaDb* is influenced by DNA methylation occurred in the bacterial host. Plasmid DNA harboring *psaDb*::LUC construct was isolated from two different bacterial strains, XL1-Blue (*dam*+ and *dcm*+) and GM2163 (*dam*- and *dcm*-). As shown in Figure 1, the plasmid prepared from XL1-Blue gave the higher expression level of *psaDb*::LUC, but expected light response did not occur at all (Fig. 1 left panel). In contrast, the plasmid DNA amplified in

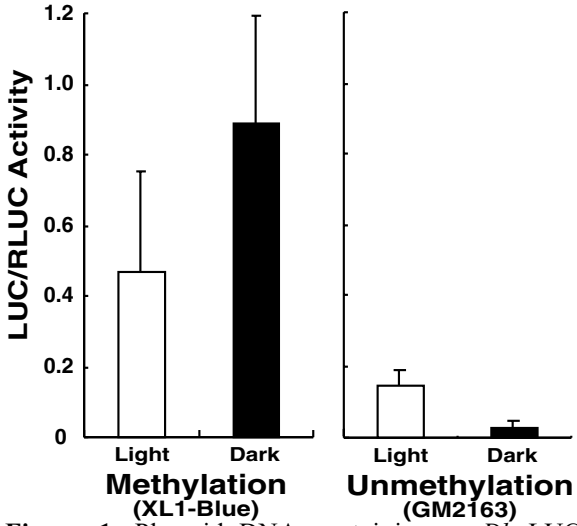

Figure 1. Plasmid DNA containing *psaDb*::LUC were prepared from two *E.coli* strains, XL1-Blue (left) and GM2163 (right). LUC/RLUC activities of the bombarded seedlings incubated in the light (open bar) or dark (solid bar) for 24 h.

GM2163 exhibited the lower expression level, but a considerable light-response occurred (Fig. 1 right panel). These results demonstrate that the light-responsive transcription of *psaDb* is greatly influenced by the methylation of DNA. Because of this reason, all the constructs used for particle bombardment after this were prepared from GM2163 strain.

5' deletion analysis and linker scanning analysis of psaDb promoter

The *psaDb* promoter was found to contain several characteristic motifs (Yamamoto *et al.*, 1997). To examine whether these *cis*-motifs play significant roles in light-responsive expression, we made a series of 5' deletion mutants of *psaDb*::LUC (Fig. 2A). When bombarded seedlings were exposed to light, the LUC activities of all constructs, except for *psaDb*(-170)::LUC, were significantly higher than those of the seedlings kept in darkness (Fig. 2B). The light-induced LUC activity of *psaDb*(-170)::LUC was so close to the background level that we could not estimate its light/dark difference (Fig. 2B). These results indicate that at least one light-responsive element (LRE) of *psaDb* is located between -233 and +24. For dissecting this region, we created 18 linker scanning constructs downstream of -249. Similarly to the wild-type construct, LUC activities of these constructs were generally much higher in light-exposed seedlings than in those kept in darkness. However, mutations introduced into 5 regions (-233 to -202, -185 to -170, -137 to -122, -105 to -41, and -11 to +12) significantly reduced the light-induced LUC activity (data not shown). This indicates that these regions should contain *cis*-elements needed for proper light responsive transcription of *psaDb*.

To our surprise, the mutations between -40 and -12 gave no effect on *psaDb* promoter activity, though TATA box is usually located within this region. Instead, the mutation between -11 and +12 significantly impaired the expression of this gene. Thus, the sequence element between -11 and +12 should be essential for this promoter activity. This region contains the major and minor transcription start sites of *psaDb*, at +1 and +9, respectively. We compared the sequence motifs around them with those of

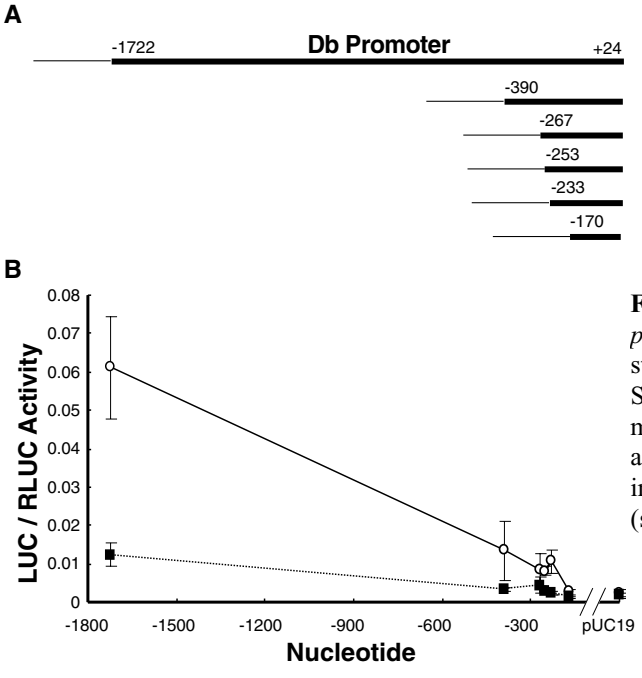

Figure 2. 5' deletion analysis of the *psaDb*::LUC constructs. The transcription start site of *psaDb* is denoted +1. (A) Schematic illustration of the 5' deletion mutants of *psaDb*::LUC. (B) LUC/RLUC activities of the bombarded seedlings incubated in the light (open circle) or dark (solid box) for 24 h.

the mammalian Initiators (Inrs) of TATA-/Inr+ type promoters, and found that the sequence motifs covering *psaDb* transcription initiation sites are highly homologous to the mammalian Inr motif, whose consensus sequence is CTCAYTYY, when Y is pyrimidine.

Initiator mutation

We further examined whether the Inr motifs are essential for the light-responsive transcription of *psaDb*. We created additional constructs that have mutations at the Inr motifs. We destroyed the Inr consensus motif at the minor transcription start site, major start site, and at both. Light-induced LUC activity was decreased in this order, and only faint activity was detected when both transcription start sites were mutated. This indicates that the Inr consensus motif is essential for the light-regulated transcription of *psaDb*. This is the first finding of the TATA-/Inr+ type core promoters in plants.

TATA-Inr swap

From the studies of animal genes, TATA-dependent and Inr-dependent transcription are thought to utilize a similar set of general transcription factors. If so, it might be possible that TATA box and Inr activate transcription in a synergistic manner (Roeder, 1996; Smale *et al.*, 1998). To examine whether TATA box synergistically activate *psaDb* transcription starting within the Inr motif, we introduced a TATA box into the *psaDb* promoter about 30 bp upstream of the transcription start sites. Constructs of the four combinations of TATA box and Inrs (TATA+/Inr+, TATA-/Inr+, TATA+/Inr-, and TATA-/Inr-) were made, and subjected to transient expression assay. Compared to the natural *psaDb* promoter (TATA-/Inr+), addition of TATA box (TATA+/Inr+) caused little effect on the light-responsive transcription. Moreover, light-induced LUC/RLUC activity of the TATA+/Inr- construct was less than half of wild-type construct, and the apparent activity of the TATA-/Inr- construct was very close to the background level. These results indicate that TATA box cannot compensate the Inr nor activate Inr-dependent transcription, in the context of *psaDb* promoter. In other words, the upstream regulatory elements of *psaDb* fulfill its function more efficiently when core promoters of suitable subtype are supplied.

General occurrence of TATA-less promoters in the plant genome

As far as we know, tobacco *psaDb* provides the first example of the TATA-/Inr+ promoters in plants. Whether TATA-less promoter architecture is unique to *psaDb* or more common to plant genes, we analyzed the 232 plant promoters using a computer algorithm designed for identifying *cis*-regulatory elements (Tsunoda and Takagi, 1999). The frequency of plant TATA-less promoters was estimated to be one sixth of the total genes. For further examination, we classified the 232 genes into seven groups according to their functions. Although more than 90% of non-photosynthesis genes have a TATA box at expected positions, its frequency in photosynthesis genes was less than one third. Thus, high abundance of TATA-less promoters appears to be a unique characteristic of the photosynthesis gene group.

Recent progress in the *Arabidopsis* genome project provides us the opportunity to systematically analyze the core promoter architectures in the plant genome. At the first on set, we carefully analyzed the core promoter architecture of thylakoid membrane protein genes in this weed. As is expected from the computer prediction

mentioned above, we found that the thylakoid membrane protein genes are driven by TATA-less promoters almost exclusively. Furthermore, the majority of the thylakoid membrane genes have neither TATA box nor Inr.

We summarized these results in Figure 3. The most of non-photosynthesis genes have TATA containing core promoters (Fig. 3 Types I and II), whereas the photosynthesis genes have generally TATA-less type promoter (Fig 3. Types III and IV); i.e. tobacco *psaDb* has a Type III promoter, and the majority of plant thylakoid membrane genes have Type IV.

Discussion

We found that the light response of the *psaDb* promoter is susceptible to the core promoter architecture. Light signal mediated by the light-responsive elements (LREs) can activate transcription in an Inr-dependent manner, irrespective of the presence or absence of TATA box. In the absence of Inr, basal transcription depending on TATA box was detected, but it was less competent in accepting the activation signals from the LREs. If core promoter lacks both Inr and TATA box, basal transcription occurs only to a very limited extent.

Though much attention has been paid to gene-specific activators in the distal promoter region, little studies have been done on the roles of core promoter modules in plant gene regulation. Core promoters have been thought to be a mere docking site of general transcription factors, and improbable to contribute to complexed gene regulation. However, here we demonstrated that the combinations of distal LREs and core promoter subtypes are critical for the proper regulation of photosynthesis gene.

Here we also demonstrate that the majority of photosynthesis genes have TATA-less promoters, similarly to *psaDb*. Moreover, most of the thylakoid membrane protein genes lack not only TATA box but also Inr. As far as we know, plant photosynthesis gene group is the first example where TATA-less promoters appear so frequently in the eukaryotic genomes. Since 14% of the total *Arabidopsis* genes are estimated to be photosynthesis genes (The Arabidopsis Genome Initiative, 2000), it

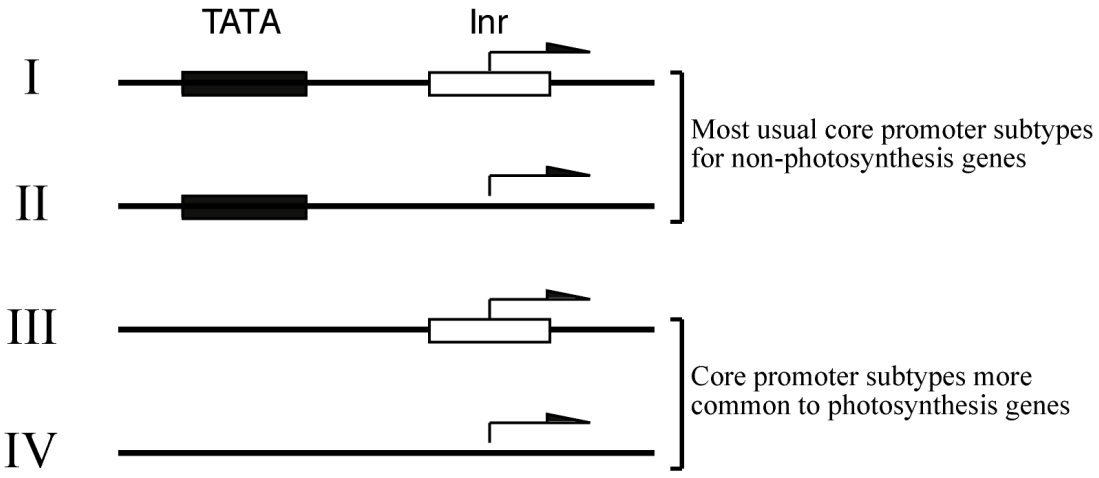

Figure 3. Core promoter subtypes in the plant genome. Core promoters are classified at least four subtypes. TATA box (solid box) containing type (I & II), and TATA-less promoter (III & IV). Type I and III has Inr motif (open box) at the transcription start site (arrow).

should be a significant subject in plant genomics, why photosynthesis nuclear genes mostly lack core promoter modules which are usually present in the other gene systems.

According to the endosymbiotic theory, nuclear genes for photosynthetic apparatus are derived from the prokaryotic endosymbiont, probably a kind of cyanobacteria. Thus, their eukaryotic promoter elements should have been acquired after their translocation to the nucleus, and subsequently integrated into the regulatory system of the host cells. Our finding described in this manuscript suggest that, somewhere in this process, there should have been strong pressure which biased the photosynthesis gene to use TATA-less type promoters.

Acknowledgements

We express our gratitude to Drs. M. Sugiura, Y.Y. Yamamoto, and T. Hirose for useful discussions and encouragement. A part of this study was supported by grants-in-aid from the Ministry of Education, Science, Sports and Culture, Japan.

References

Azizkhan JC, Jensen DE, Peirce AJ, Wade M (1993) *Crit Rev Eukaryot Gene Expr* **3**, 229-254.

Cavin-Périer R, Praz V, Junier T, Bonnard C, Bucher P (2000) *Nucleic Acids Res* **28**, 302-303.

Jefferson RA, Kavanagh TA, Bevan MW (1987) *EMBO J* **6**, 3901-3907.

Miyamoto T, Nakamura T, Nagao I, Obokata J (2000) *Plant Mol Biol Rep* **18**, 101-107.

Murashige T, Skoog F (1962) *Physiol Plant* **15**, 473-497.

Roeder RG (1996) *Trends Biochem Sci* **21**, 327-335.

Roeder RG (1998) *Cold Spring Harb Symp Quant Biol* **63**, 201-218.

Rogers JC, Rogers SW (1995) *Plant J* **7**, 221-233.

Smale ST, Jain A, Kaufmann J, Emami KH, Ko K, Garraway IP (1998) *Cold Spring Harb Symp Quant Biol* **63**, 21-31

Terrers JT, Block A, Hahlbrock K, Somssich IE (1993) *Plant J* **4**, 587-592.

The Arabidopsis Genome Initiative (2000) *Nature* **408**, 796-815.

Tsunoda T, Takagi T (1999) *Bioinformatics* **45**, 622-630.

Yamamoto YY, Kondo Y, Kato A, Tsuji H, Obokata J (1997) *Plant J* **12**, 255-265.

Yamamoto YY, Nakamura M, Kondo Y, Tsuji H, Obokata J (1995) *Plant Cell Physiol* **36**, 727-732.

Yamamoto Y, Tsuji H, Obokata J (1993) *Plant Mol Biol* **22**, 985-994.