Cloning and sequence analysis of germin-like protein gene 2 promoter from *Oryza sativa* L. ssp. indica

To cite this Article: , ‘Cloning and sequence analysis of germin-like protein gene 2 promoter from *Oryza sativa* L. ssp. indica’, DNA Sequence, 18:1, 26 - 32

http://dx.doi.org/10.1080/10425170600986688

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

© Taylor and Francis 2007
FULL LENGTH RESEARCH ARTICLE

Cloning and sequence analysis of germin-like protein gene 2 promoter from Oryza sativa L. ssp. indica

TARIQ MAHMOOD, M. ZEESHAN HYDER, & S. M. SAQLAN NAQVI

Department of Biochemistry, University of Arid Agriculture Rawalpindi, Rawalpindi 46300, Pakistan

(Received 22 March 2006)

Abstract
Germin and germin-like proteins (GLPs) are water soluble extracellular proteins reportedly expressed in response to some environmental and developmental signals. Some enzymatic activities have also been associated with germin/GLPs. However, their role in overall metabolism has not been fully understood. Significant insight into their function may also be gained by analysis of their promoter. During this study, about 1107 bp 5′ region of OsRGLP2 gene was amplified, cloned and sequenced. The sequence analysis by BLAST showed that this promoter sequence has five common regions (CR1–CR5) of different sizes, which are repeated at 3–6 other locations in 30 kb region in which this gene driven by its promoter is located. Interestingly, all the genes driven by promoter harboring these common regions are GLPs/putative germins. Analysis of these common regions located on OsRGLP2 indicated presence of many elements including those for light responsiveness, dehydration and dark induced senescence, stresses (pathogen and salt), plant growth regulators, pollen specific expression and elements related to seed storage proteins. Analysis of the 30 kb germin/GLP clustered region by GenScan detected each gene to have a putative 40 bp promoter which contains TATA box and Dof factor which turned out to be a part of CR2.

Keywords: GLP, promoter analysis, regulatory elements, Oryza sativa

Introduction
Germin was first isolated in association with wheat germination, from where it got its name “germin”. This protein is a water soluble glycoprotein with oxalate oxidase activity. It forms homopentamer that is highly resistant to proteases and to dissociation by various denaturing agents such as heat, SDS and extreme pH. Germins and germin-like proteins (GLPs) have also been reported to be involved in plant defense (Schaffleitner and Wilhelm 2002; Ramalingam et al. 2003). Their possible involvement in tolerance to osmotic stress is in agreement with the demonstration that germins and GLPs are part of a superfamily of proteins including seed storage globulins as well as a sucrose binding protein, potentially involved in cellular water homeostasis (Braun et al. 1996).

Four different enzymatic activities have been found to be associated with these proteins, i.e. oxalate oxidase associated with the true germins, SOD with some germins and GLPs, ADP-glucose pyrophosphate or phosphodiesterase (AGPase) with a barley GLP (Rodriguez-López et al. 2001) and finally a serine protease inhibitory activity in wheat apoplast (Segarra et al. 2003). There are many reports suggesting that germins and GLPs are synthesized in response to changes in plant water status and salt stress. In order to examine expression of germins and GLPs it is important to study their promoter regions. Pinus caribaea germin1 (PgGER1) promoter in tobacco bright yellow 2 transformed cells was found responsive to 2,4-D and BA and derived maximum expression of reporter gene in G1 growth phase (Mathieu et al. 2003).

Keeping in view the importance of germin and GLPs, the study was designed to clone the upstream promoter region of an Oryza sativa Root GLP 2 (OsRGLP2) already isolated, sequenced and submitted to Genbank.
Materials and methods

Plant material

Rice plants (Nonabokra) were grown from seeds in green house conditions.

DNA extraction and polymerase chain reaction (PCR)

DNA was isolated from leaf tissue by CTAB method as described by Richards (1997). A pair of primers was designed to amplify a fragment of ~1100 bp upstream of OsRGLP2.

RGLP2PF1: 5’ CCCGGGCTGGTCTACCTTG-GCATTTG 3’
RGLP2PR: 5’ CCCGGGCTTTCTGCTGTAAT-TATTTGCT 3’

These amplifications included pre-polymerase chain reaction (PCR) denaturation at 94°C for 5 min followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 1 min. Final cycle was same except extension for 20 min at 72°C. After PCR contents were held at 4°C till use. About 25 μl of the PCR mixture contained, 50 pmol of each primer, 2.5 μl of 10 × PCR buffer, 1.5 μl of 25 mM MgCl2, 1.5 μl of 2 mM dNTPs and 1.5 U of Taq polymerase (MBI Fermentas) were used.

Preparation of electro competent cells

A single colony of DH5α (a strain of Escherichia coli) was picked up and cultured in a test tube in 3 ml LB medium (tryptone 10 g/l, yeast extract 5 g/l, NaCl 10 g/l, pH 7.0) at 37°C with 250 rpm shaking for overnight. The starter culture was shifted to 1 l LB medium (tryptone 10 g/l, yeast extract 5 g/l, NaCl 10 g/l, pH 7.0) at 37°C and was again incubated at 37°C with 250 rpm shaking until the OD600 was 0.5. The medium, and was again incubated at 37°C overnight. The starter culture was shifted to 1 l LB medium (tryptone 10 g/l, yeast extract 5 g/l, NaCl 10 g/l, pH 7.0) at 37°C with 250 rpm shaking. Plasmids were isolated from the transformation mixture was then spread on LB X-gal/IPTG agar plates containing 50 mg/l ampicillin. Plates were incubated overnight at 37°C. Colonies were observed after 16–20 h incubation.

White colonies were picked up and grown in LB media containing Ampicillin (50 mg/ml) at 37°C with 250 rpm shaking. Plasmids were isolated from the cultures and subjected to PCR confirmation followed by restriction digestion.

Sequencing of promoter region

High quality engineered plasmid DNA was isolated by JETquick plasmid miniprep spin kit (GENOMID). Sequencing was carried out by using universal primers M13 forward or reverse present on either sides of the multiple cloning site of pTZ57R/T: Beckman Coulter sequencer (CEQ 8000) was used for sequencing.

M13F: 5’ TGTAAAACGACGGCCAGT 3’
M13R: 5’ CAGGAAACGCTATGACC 3’

Sequencing reaction was prepared using CEQ™ Dye Terminator Cycle sequencing kit according to the manufacturer’s instruction.

Sequence analysis

Sequence of cloned fragment was analyzed by using blast n to find out homology in rice genome. Blast returned a region from chromosome 8 of O. sativa with highest similarity (99.7%). Only two other significant similarities were found including one on BAC clone and another on PAC clone. There were several small regions of exact sequence matches common to same chromosome. These small regions were termed as common regions and their analysis was done to elucidate their importance.

Signal scan for regulatory element analysis

Assuming the importance of repeatedly distributed short sequence matches on a relatively small segment of chromosome 8; all common regions were subjected to regulatory elements analysis to perceive their significance. PLACE/Signal Scan (Higo et al. 1999) (www.dna.affrc.go.jp) was used to reveal the regulatory elements present in these common regions. All the regulatory elements found were searched in different databases to explore the functional importance already reported. Depending on the data obtained full length promoter region of OsRGLP2 was characterized for harboring different kinds of regulatory elements.
Each of the regulatory elements observed was analysed with the help of previously reported properties of that particular element.

**Results**

**Cloning and sequencing of OsRGLP2 promoter region**

Amplification of promoter region of OsRGLP2 was carried out by using RGLP2PF1 and RGLP2PR primers. A product of about 1100 bp was observed on 1% agarose in TAE. The product was ligated in pTZ57R/T vector and cloned in DH5a. Cloning was confirmed by PCR and restriction analysis. The cloned DNA fragment was sequenced by using universal primers (M13F, M13R) and fragment specific primers (RGLP2PF1, RGLP2PR). Resultant sequence has been submitted to Genbank (Accession No. DQ414400).

**Sequence analysis**

Sequencing revealed a fragment of 1107 bp. The sequence was found to contain five common regions (CR1–CR5) ranging from 28 to 71 bp in length on chromosome 8 of *O. sativa* (Accession No. AP008214), each repeated 3–6 times within a segment of 30 kb. These regions were separately analysed by PLACE/Signal Scan (Higo et al. 1999). It was observed that each common region has some regulatory elements including elements related to seed storage proteins, light responsiveness, dehydration and dark-induced senescence and elements specific to expression in pollen. Furthermore, different motifs responsive to environmental stresses (salt, pathogen attack) and plant growth regulators were also found in these common regions (Table I). Further, it was observed that each repeated common region has a member of GLP family downstream to it. This 30 kb region was found to have a number of associated ESTs belonging to GLP gene family (Table II) revealing clustering of members of GLP family.

The presence of these small stretches of sequences (CR1–CR5) upstream of different germin and GLP emphasize the importance of these regions which may be signified not only by their repetition in a relatively small part of the sequence but also by their conservation during the course of evolution. It is quite surprising to see such long stretches repeated so many times in a relatively small segment of a chromosome. A promoter of 40 bp was also identified by processing the 30 kb region with the help of GenScan (http://genes.mit.edu/GENSCAN.html), and was found to be present upstream to each member of GLP including OsRGLP2. Interestingly this 40 bp putative promoter turned out to be a part of CR2.

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Common region (CR)</th>
<th>Number of repetition within 30 kb region</th>
<th>Sequence length (base pairs)</th>
<th>CR sequence</th>
<th>Regulatory elements (copies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CR 1</td>
<td>5</td>
<td>28</td>
<td>GAGAAAACACGAAATAATTCACGAGAGAA</td>
<td>AGAAA element</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TCACAATTCACGCTGCCGCCGCCCT</td>
<td>Dof, TATA Box</td>
</tr>
<tr>
<td>2</td>
<td>CR 2</td>
<td>6</td>
<td>71</td>
<td>GCGGCCATATAGACGATTAATCCT</td>
<td>300 element, Dof</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CCACACCCACCATCACCC</td>
<td>ACGTATERD1 (ACGT), CAAT Box, GTGA motif</td>
</tr>
<tr>
<td>3</td>
<td>CR 3</td>
<td>6</td>
<td>25</td>
<td>TTTTAGCGAGGAGGTCACCTGCA</td>
<td>300element, ARR1 (2), Dof, GATA, IBOXCORE, CAAT (2), GT-1 (3)</td>
</tr>
<tr>
<td>4</td>
<td>CR 4</td>
<td>3</td>
<td>36</td>
<td>ATCTAAATGTTGAAAAACA</td>
<td>300element, ARR1 (2), Dof, GATA, IBOXCORE, CAAT (2), GT-1 (3)</td>
</tr>
<tr>
<td>5</td>
<td>CR 5</td>
<td>3</td>
<td>66</td>
<td>CAATTTGATTTGATCCCAAGATAAA</td>
<td>300element, ARR1 (2), Dof, GATA, IBOXCORE, CAAT (2), GT-1 (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TGGCGACAAATGAAAAAATGAAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GGTGAGAACATGTTTCTT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CR1</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>CR2</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>CR3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>CR4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>CR5</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

* GenBank Accession number.
Regulatory elements analysis

The part of the promoter region other than the common regions was also scanned for detection of regulatory elements. Interestingly such elements are clustered in four locations, rather than being distributed evenly on whole of the sequence (Figure 1).

Discussion

Light molecular weight and high molecular weight seed storage proteins regulatory elements

Genes encoding the most abundant storage protein have been isolated and characterized from several cereal species. While working on the analysis of OsRGLP2 promoter region, such endosperm-specific regulatory elements have also been found, which are well documented (Kreis et al. 1986; Colot et al. 1989). Three copies each of −300 elements and, HMW and LMW glutenin gene conserved regulatory sequence (TGAAAGTG/TGAAAAAG/TGAAAAAAG) are present in OsRGLP2 promoter region. This “−300 element” has been reported to be present upstream of the promoters for the β-hordein gene of barley and the α-gliadin, γ-gliadin and low molecular weight (LMW) glutenin genes of wheat (Kreis et al. 1986; Colot et al. 1989). The same element has been found in the corresponding position upstream of the coding sequences for the α-, β-, and γ-zeins (Forde et al. 1985; Kreis et al. 1986; Thompson and Larkins 1989). The −300 element has the consensus sequence TGACA/ATGACA and incorporates two smaller conserved motifs (underlined). One motif is identical to the consensus binding site (ATGAC/GTCAT) for the yeast GCN4 protein (Hill et al. 1986) and functionally equivalent transcription activators AP-1, Jun, and Fos (Pawson 1987; Struhl 1989; Vogt and Bos 1989). Another variant of −300 element core sequence TGCAAG has been reported by Thomas and Flavell (1990) in the important −148 to −186 interval of the high molecular weight (HMW) glutenin promoter. Similar regions with few variations have been found at positions −230 (TGAAAGTG), −461 (TGAAAAAG) and at −939 (TGAAAAAAG) in the upstream region of OsRGLP2. Colot et al. (1987) also reported similar motifs at about −300 bp from the transcription start of the LMW glutenin gene, in the HMW glutenin gene. It is also conserved among several storage protein genes in cereals (Kreis et al. 1985). These identical conserved sequences found in OsRGLP2 promoter indicate that this sequence may have some important roles for endosperm-specific expression.

GATA factors

GATA factors are reportedly involved in light regulated gene expression in higher plants (Teakle et al. 2002). The presence of seven copies of GATA elements in the upstream region of OsRGLP2 may be responsible for circadian oscillation as demonstrated in Sinapis alba (Heintzen et al. 1994).

Figure 1. Description of different regulatory elements present within five common regions and on entire sequence of OsRGLP2 promoter region. The elements in pink are present in CRs.
ARR1 and ARR2

These elements include putative binding sites for the arabidopsis response regulators (ARRs), specially the type-B ARR1, shown to be involved in early responses to cytokinins (Sakai et al. 2001; Oka et al. 2002). The type-B ARRs, along with a Myb–DNA binding domain functions as transcriptional activators of cytokinin-induced gene expression (Imamura et al. 1999; Sakai et al. 2001; Haberer and Kieber 2002). Presence of three copies of ARR1 elements in OsRGLP2 indicates that it may be involved in controlling expression of OsRGLP2 in response to cytokinins.

DNA binding with one finger transcriptional factors (Dof)

Dof factors are suggested to be involved in expression of photosynthetic genes, seed storage proteins genes and genes responsive to a plant hormone and stress signals (Yanagisawa 1995; Zhang et al. 1995; De Paolis et al. 1996; Vicente-Carbajosa et al. 1997; Kis u et al. 1998; Mena et al. 1998; Yanagisawa and Sheen 1998; Baumann et al. 1999). In OsRGLP2 promoter, ten copies of AAAG were found to be present with eight copies having sequence AAAAG. Other motif of Dof proteins including GAAAG, CAAAG and a variant GAAGAC are also found in the upstream region. It has been reported that at least one AAAG motif is important for DNA binding of Dof proteins (Yanagisawa and Schmidt 1999). These similarities in regulatory elements shows that GLPs may have the characteristics related to seed storage proteins (Braun et al. 1996), and signaling process in response to plant hormones and stress (De Los Reyes and McGrath 2003; Mathieu et al. 2003).

Pollen specific regulatory elements

The pathway of pollen maturation provides an accessible and well characterized system to study the tissue-specific and developmental control of gene expression. At maturity the pollen grain contains about 20,000 unique transcripts of which 2000 are estimated to be specifically or preferentially expressed in pollen (Willing et al. 1988; Schrauwen et al. 1990). Bate and Twell (1998) have reported two co-dependent pollen specific regulatory elements AGAAA and TCCACCATA during the 5’ promoter deletion analysis of tomato lat52 gene. Rogers et al. (2001) reported another conserved pollen specific GTGA motif of tomato lat56 and tobacco g10. While analyzing OsRGLP2 promoter region, it was observed that it contains five copies of AGAAA element, seven copies of GTGA motif and four copies of CCAC motif, while TCCACCATA is not present. Among these elements and motifs, a copy of GTGA at -294 and two copies of CCAC at -31 and -95 from transcriptional start site seems to be important for being present within 300bp in the core promoter region. Presence of five copies of CAAT box in OsRGLP2 promoter was also noted. These boxes are reported to be involved in tissue specific promoter activity of a pea legumin gene in tobacco (Shirsat et al. 1989).

GT elements

The GT-1 cis-element was first identified in pea as the Box II element (GTGTGTTAATATG) in the promoter of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) small subunit gene (Green et al. 1987). It is thought that the high degeneracy of the GT-1 cis-element partly explains its diverse functions as well as its light-specific regulatory functions. Root GT element-binding factor (RGTF) has a distinct sequence requirement for binding, which might help to determine the fine tuning of expression of individual genes in root cells. These reports suggest that upstream regions of genes having GT element rapidly induce their respective genes after encountering stress related stimuli like pathogen attack or NaCl stress (Snedden and Fromm 1998, 2001). Further, these elements are also involved in activation or repression of different plant photosynthetic genes upon light and dark condition, respectively. There are five copies of the core sequence of GT-1 element and one copy of GT-1 GMSCAM4 present in the promoter region of OsRGLP2, indicating that this promoter region may have some of the GT elements characteristics. This is consistent with the data suggesting the role of GLPs in plants under pathogen attack (Schafleitner and Wilhelm 2002; Ramalingam et al. 2003) or salt stress (Hurkman and Tanaka 1996). OsRGLP2 promoter also contains one copy of ACGT, a drought and a dark induced senescence responsive element described by Simpson et al. (2003).

On the basis of sequence analysis it has been found that OsRGLP2 promoter has regulatory elements related to seed storage proteins, light responsiveness, dehydration and dark-induced senescence, those specific to expression in pollen, to plant growth regulators and to environmental stresses (salt, pathogen attack).

It is quite understandable that in non-transcribed region of DNA; short stretches of DNA may appear simply by chance. In addition, their mere presence does not qualify them for function, as their location and in many cases presence of other co-dependent sequences at specific distances may also be important. Nevertheless such analysis is simply indicative and yields important clues for direction of experimental analysis. We are therefore, making transgenic plants containing GUS gene under the influence of different sizes of OsRGLP2 promoter region. These
plants will be analysed for the role of these putative promoter elements, during the developmental/environmental conditions in which their role is reported in literature.

References


De Paolis A, Sabatini S, de Pascalis L, Costantino P, Capone I. 1999. The GLP promoter region 31

Emoto E, Baroja-Fernández E, Zandueta-Criado A, Rodríguez-López M, León TJ, Bar controlo de una proteína que se une a interceptores de señales (GCN4) y sus reguladores. FEBS Lett 470:154–158.


