TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2015) 39: 389-400
© TÜBİTAK
doi:10.3906/bot-1405-48

Identification and analysis of regulatory elements in the germin and germin-like proteins family promoters in rice

Iftikhar ALI*, Tariq MAHMOOD
Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan

Received: 19.05.2014 • **Accepted:** 28.11.2014 • **Published Online:** 04.05.2015 • **Printed:** 29.05.2015

Abstract: Germins and germin-like proteins constitute a ubiquitous family of plant proteins and have a role in the defense against pathogen attacks. In the present work the *Oryza sativa* germin-like protein 1 (*OsGLP1*) and putative germin A (*OsGerA*) promoter regions of five Pakistani rice varieties were analyzed for variations in expression regulation of these promoters and similarity with others germin and germin-like protein promoters. Phylogenetic analyses showed that the studied *OsGLP1* promoters were tightly clustered around the ancestral group while putative *OsGerA* clustered around the descendant group. HADDOCK was used for docking selected transcription factors (TFs) with its corresponding regulatory elements, for determining the most stable interaction with the highest probability of hydrogen bonding between them. TATA box binding protein (TBP) elements mainly existed on the 3′ regions very near the gene, while far regions of promoters mostly lacked such elements. This study showed that TBP not only binds to (-30) TATA box element, but also binds to a wide range of elements at different rates, positioned on about a thousand base pairs upstream region of germin and germin-like genes. Arginine was found to be the most reactive residue in TFs, while adenine was the most sensitive in regulatory elements.

Key words: Germin and germin-like proteins, *OsGLP1*, DNA-protein interaction, TATA box binding protein (TBP)

1. Introduction

Germins and germin-like proteins (*GLPs*) belong to the cupin superfamily of proteins, identified in several plants, including monocot, dicot, gymnosperms, *Physarum polycephalum*, and myxomycete (slime mold), after initial discovery of a protein marker germin in the germination of wheat seeds. Germin relatives have also been recognized in animals, prokaryotes, and fern spores (Lu et al., 2010).

Germins and *GLPs* present major problems in classification due to their high sequence conservation in multiple plant species (Dunwell and Gane, 1998). A strong classification system is established by structural characteristics of these proteins due to the clustering of proteins with conserved functions (Agarwal et al., 2009). The proteins that have oxalate oxidase (OXO) enzymes activity, exclusively found within cereal plant species, are placed in a well-conserved homogeneous group called true germins (Carrillo et al., 2009; Davidson et al., 2009), while the germin-motif comprising proteins with an average of 50% amino acid sequence identity that either do not have OXO action or have not yet been allotted an enzymatic function are placed in a heterogeneous group called germin-like protein (Dunwell et al., 2008; Breen and Bellgard, 2010).

To date, major research on germins and *GLPs* has been conducted on cereal plants, especially wheat, maize, barley, and rice. On the basis of various enzymatic activities, six germin subfamilies (GER1-6) have been described. GER1 (true germin proteins) has been shown to possess OXO activity, while GER2 has been shown to possess superoxide dismutase (SOD) action (Banerjee and Maiti, 2010). It has also been shown that the GER1 subfamily is important for early plant development and germination in plants (Federico et al., 2006). The GLPs with mostly unknown function in plant genomes have been classified into subfamilies (Carter and Thornburg, 1999, 2000). In contrast to a true germin subfamily, both GLP1 and GLP2 subfamilies are limited to proteins of SOD action, while GLP3 is involved in phosphodiesterase activities. However, recently more subdivisions were suggested (Breen and Bellgard, 2010). A key feature of the germin and GLP subdivisions was the conservation of a motif derived from that of the cupin superfamily (Carter and Thornburg, 1999, 2000).

* Correspondence: iftikharphd@gmail.com

GLPs have been studied in a wide variety of systems and have been revealed to be connected with plant cell defense and diseases, and to be highly resistant to heat, extreme pH, sodium dodecyl sulfate (SDS), and proteases (Membré et al., 2000). Significant GLPs and/or OXO expression have been found in environmental conditions such as drought stress (Ke et al., 2009), aluminum stress (Houde and Diallo, 2008), and salt stress (Cramer et al., 2007). Expression of these genes also occurred when attacked by viruses and bacteria (Park et al., 2003), and fungal pathogens (Manosalva et al., 2008). Their expression occurred in a wide range of tissues and can act against pests like insects (Lou and Baldwin, 2006), and nematodes (Knecht et al., 2010). Moreover, *OsGLP1* downregulated transgenic plants in rice (a member of the GER2 subfamily) were shown to cause dramatic increases in sheath blight, discharge fungal diseases, change cell morphology, and induce dwarfism (Banerjee and Maiti, 2010; Breen and Bellgard, 2010).

GLPs contain the germin-motif that gives rise to a predicted b-barrel core involved in metal binding (Requena and Bornemann, 1999). Most of them share biochemical features such as sucrose-binding, globulins, and seed storage proteins, though they differ in their enzyme activities and tissue specificities (Kim et al., 2004). GLPs have been found in seeds, flowers, embryos, cotyledons, roots, stems, and leaves and were shown to be involved in several essential processes including apoptosis, defense, photoperiodic oscillation, development, and osmotic regulation (Mahmut, 2000; Breen and Bellgard, 2010). TATA box-binding protein (TBP) is a general eukaryotic transcription factor that is required by all three eukaryotic RNA polymerases for correct initiation of transcription of messenger, transfer, ribosomal, and small nuclear RNAs. Since the first gene encoding a TBP was cloned, it has been the object of considerable biochemical and genetic research. Substantial progress has also been made on structural studies, including three-dimensional structures of TBP and TBP-TATA box complex via protein–protein and protein–DNA interactions (Burley, 1996).

In contrast to the advanced knowledge of the expression features, cell biology, and structure of wheat and barley germins and *GLPs*, less is known about rice germin and *GLPs* promoters. In the present work, five *OSGLP1* and Putative Germin A promoter regions of five Pakistani rice varieties were analyzed to find expression variation and similarities with other germins and *GLPs* promoters. Phylogenetic analyses of these promoters were performed to describe the variations in the rice gene family control.

2. Materials and methods

2.1. Plant materials

The seeds of five different rice varieties (C-622-Basmati, Pakhal-Basmati, Kashmir-Basmati, JP-5-Basmati, and Rachna-Basmati), collected from the National Agriculture Research Centre, Islamabad, were sown in jiffy under controlled conditions of light, humidity, and temperature in a growth room during 2013–14. The total genomic DNA of all samples was extracted from young leaves of 2–3-week-old seedlings using the CTAB method (Richards, 1997). The presence of extracted DNA and its quality was checked by running DNA samples on 1.5% agarose gel, and visualized under ultraviolet (UV) light in Dolphin Doc^{plus} gel documentation system (Wealtech). Two pairs of oligonucleotide PCR primers (*OsGLP1* primers were designated from Accession No. AP004586, while *OsGerA* primers were designated from Accession No. AP005531, by Primer3 tools) were used to amplify the upstream promoter regions of *GerA* and *GLP1* genes. The sequences of the primers are given below:

*OsGLP1*F: 5′-CTTAGCTCCTACTTGCAAACAAC-3′

*OsGLP1*R: 5′-TGGCCATGGCTACCACACTA-3′

*OsGerA*F: 5′-GATCATTCGCAAAAGTGTTGG-3′

*OsGerA*R: 5′-GAAGGAAGACGAAGCCATTC-3′

PCR amplifications of *OsGerA* and *OsGLP1* promoter regions were carried out separately in a final volume of 25 µL reaction mixture containing 1 µL of genomic DNA, 1 µL of 50 pM/µL of each primer, 1.5 µL of each 2 mM dNTP, 2.5 µL of 10X PCR buffer, 1.5 µL of 50 µL $MgCl_2$, and 0.3 µL of *Taq* DNA polymerase in gradient Multi Gene Thermal Cycler (Labnet). The PCR cycling parameters consisted of an initial denaturation of 94 °C for 3 min, followed by 35 cycles of 94 °C for 1 min, 55 °C for 1 min, 71 °C for 2 min, and a final extension at 71 °C for 10 min. The PCR amplified product was purified by jet quick PCR product purification spin kit (Genomed), visualized with UV light by using 1.5% (w/v) agarose gel, and were sequenced commercially. Sequences were submitted to GenBank (Accession no. KJ807466–KJ807475).

2.2. Computational analysis of the sequenced data

Sequence analysis was managed with the help of bioinformatics tools used for DNA, RNA, or protein sequences to recognize and compare with already sequenced samples through sequence alignment and different sequence databases (Durbin et al., 1998) based on computational technology. Maximum Parsimony method in MEGA 5 was used to construct a phylogenetic

tree (Figure 1) for all studied samples along with already reported germin and *GLP* family promoter sequences taken from the NCBI database (http://www.ncbi.nlm.nih.gov/).

2.3. Identification of regulatory elements

The sequence data were analyzed for the identification of *cis*-acting elements and their corresponding transcription factors (TFs). An online available bioinformatics tool 'Consite' (http://consite.genereg.net/) was used to reveal different regulatory elements present in the upstream promoter regions of *OsGLP1* and *OsGerA* genes (Sandelin et al., 2004).

2.4. Modelling of regulatory elements and corresponding transcription factors

For modelling *cis*-acting elements, an online available tool named "3D-DART" (available at http://haddock.science.uu.nl/services/3DDART/) was utilized that provides DNA 3D structure in PDB format. Similarly, structural data of TFs were collected through the Protein Data Bank (PDB) accessible at http://www.rcsb.org/pdb/ (Berman et al., 2000). Following filtration of data, the best PDBs were selected on the basis of missing residues, amino acids length, mutation, resolution, and structure type. The best PDBs were explored for further analysis of protein-DNA docking.

2.5. Protein–DNA docking

Protein–DNA docking was done by the HADDOCK web-server, freely available at http://haddock.chem.uu.nl/. The web interface requires active residues (interacting residues) and structures of both bio-molecules. It also serves as a platform for protein-protein and protein-ligand docking (De Vries et al., 2010). Active residues of TFs were identified with the help of another tool named DISPLAR (http://pipe.scs.fsu.edu/displar.html). Active residues of DNA were provided along with the structure files produced by the 3D-DART web server. The HADDOCK webserver evaluates models on the basis of the HADDOCK score. The score is a combination of buried surface area, desolvation, electrostatic, Van der Waals, and restraint violation energies. Cluster size demonstrates the number of best structures; the structure with the lowest energy is favored because lowest energy structures are considered to be good models (De Vries et al., 2010).

2.6. Visualization of molecular dynamics

In order to visualize the results generated by the HADDOCK webserver, a molecular visualization program PyMOL (http://www.pymol.org/) was used. PyMOL is a user-sponsored molecular visualization system on an open-source foundation that facilitates the visualization of interactions between bio-molecules.

3. Results and discussion

3.1. Phylogenetic analysis

To fit the observed sequence data with the smallest number of nucleotide changes, the maximum parsimony (MP) method was adapted to make a true phylogenetic tree (Fitch, 1971). The details of the sequences used in the tree are given in Table 1. An evolutionary tree of the studied germin and *GLP1* promoters, with the germins and *GLPs* promoters family in rice and other relatives was constructed (Figure 1) to reveal relationships among different chromosomes in different rice varieties. The studied *OsGLP1* sequences clustered around the ancestor group of the tree, while the studied putative germin A sequences were limited to the descendant group, which showed them as the two highest dissimilar groups of germins and *GLPs* family promoters. The tree topology showed a parallel pattern of evolution throughout the germin family and a distinct level of diversity. Comparative analysis revealed that the studied samples (marked with *) segregate from the tree, making a genetic lineage with a very complex history and possibly the divergence of each sequence from the same ancestral line. Os08g0460000 (similar to germin-like protein 1 precursor; Table 1) revealed an outgroup with more genetic divergence than all other promoters, making it the ancestor of all varieties in the tree. Os08g0460000 exhibited the highest sequence identity with AP004707 and the studied *OsGLP1* promoters KJ807466, KJ807467, KJ807468, KJ807469, and KJ807470. This group emerged into a series of germins and *GLPs* promoters and finally ended at its most advanced evolutionary descendant (AP005505). AP005505 exhibited the highest sequence identity with Os08g0190100 (AP005531), AF141878, and the studied putative *GerA* promoters: KJ807471, KJ807472, KJ807473, KJ807474, and KJ807475. Notably, the studied sequences were most closely related with the promoter sequences located in the same group, sharing the same regulatory elements. This group might be thought of as the most advanced group evolved from the ancestral group through a series of evolutionary lineages among rice germins and *GLPs* family (Figures 2 and 3).

Furthermore, the multiple sequences alignment analysis of these promoters yielded consistent results supported by high sequence identity values. Os08g0189900 (KF673351 OsGLP2) exhibited the highest sequence identity with Os04g0617900. DQ324800 exhibited the highest sequence identity with DQ324801. Os08g0231400 exhibited the

Table 1. Germin and GLP gene promoters used in the present study.

S/No.	Annotated name[a]/Accession No.	Chromosome	Abbreviation	Promoter group
1	Os01g0952100	1	*OsGLP2-4*	Germin-like protein subfamily 2 member 4 precursor
2	Os02g0491600	2	*OsGLP8,12*	Germin-like protein (Germin-like 8) (Germin-like 12)
3	Os02g0491800	2	*OsGLP1-15*	Germin-like protein subfamily 1 member 15 precursor
4	Os03g0651800	3	*OsGLP2-1*	Germin-like protein subfamily 2 member 1 precursor
5	Os03g0804500	3	*OsGLPT-1*	Germin-like protein subfamily T member 1 precursor
6	Os03g0804700	3	*OsGLPT-1*	Germin-like protein subfamily T member 1 precursor
7	Os04g0617900	4	*OsGLP*	Hypothetical protein (Germin-like protein precursor)
8	Os05g0277500	5	*OsGLP2-4*	Germin-like protein subfamily 2 member 4 precursor
9	Os08g0188900	8	*OsGLP*	Hypothetical protein (Germin-like protein precursor)
10	Os08g0189100	8	*OsGLP*	Hypothetical protein (Germin-like protein precursor)
11	Os08g0189400	8	*OsGLP*	Hypothetical protein (Germin-like protein precursor)
12	Os08g018950	8	*OsGLP8-6*	Oryza sativa germin-like protein 8-6
13	Os08g018960	8	*OsGLP8-7*	Oryza sativa germin-like protein 8-7
14	Os08g0189700	8	*OsGLP8,12*	Germin-like protein (Germin-like 8) (Germin-like 12)
15	Os08g0189850	8	*OsGLP8,12*	Germin-like protein (Germin-like 8) (Germin-like 12)
16	Os08g0189900 (KF673351)	8	*OsGLP8,12*	Germin-like protein (Germin-like 8) (Germin-like 12)
17	Os08g0190100 (AP005531)	8	*OsGLP8-7*	Germin-like protein 8-7
18	Os08g0231400	8	*OsGLP8-12*	Oryza sativa germin-like protein 8-12
19	Os08g0460000 (AP004586)	8	*OsGLP1*	Germin-like protein 1 precursor
20	Os12g0154900	12	*OsGLP*	Hypothetical protein (Germin-like protein precursor)
21	Os12g0155000	12	*OsGLP1-8*	Germin-like protein subfamily 1 member 8
22	EU742684	8	*OsGLP1*	Germin-like protein 1
23	AP004707	8	*OsGLP1*	Germin-like protein 1
24	AF141878	8	*OsGLP8,12*	Germin-like protein (Germin-like 8) (Germin-like 12)
25	AP005505	8	*OsGerA*	Putative germin A
26	DQ414400	8	*OsGLP2*	Germin-like protein 2
27	DD057515	1	*OsGer4*	Oryza Germin protein 4
28	DQ324800	7	*HvGerF*	Hordeum vulgare Germin F
29	DQ324801	7	*HvGerB*	Hordeum vulgare Germin B
30	AY394010	-	*ZmGLP1*	Zea mays germin-like protein 1
31	AY077704	-	*PcGLP1*	Pinus caribaea germin-like protein 1
32	DQ058010	-	*LmGLP1*	Larix × marschlinsii germin-like protein 1
33	AY864922	-	*TaGLP3*	Triticum aestivum germin-like protein 3
34	D89055	1	*AtGLP1*	Arabidopsis thaliana AtGLP1
35	KJ807466*	8	*OsGLP1*	Germin-like protein 1 precursor
36	KJ807467*	8	*OsGLP1*	Germin-like protein 1 precursor
37	KJ807468*	8	*OsGLP1*	Germin-like protein 1 precursor
38	KJ807469*	8	*OsGLP1*	Germin-like protein 1 precursor
39	KJ807470*	8	*OsGLP1*	Germin-like protein 1 precursor
40	KJ807471*	8	*OsGerA*	Putative germin A
41	KJ807472*	8	*OsGerA*	Putative germin A
42	KJ807473*	8	*OsGerA*	Putative germin A
43	KJ807474*	8	*OsGerA*	Putative germin A
44	KJ807475*	8	*OsGerA*	Putative germin A

[a]Annotated names in respective genome annotation databases i.e. plantpromoterdb (http://133.66.216.33/ppdb/cgi-bin/index.cgi).

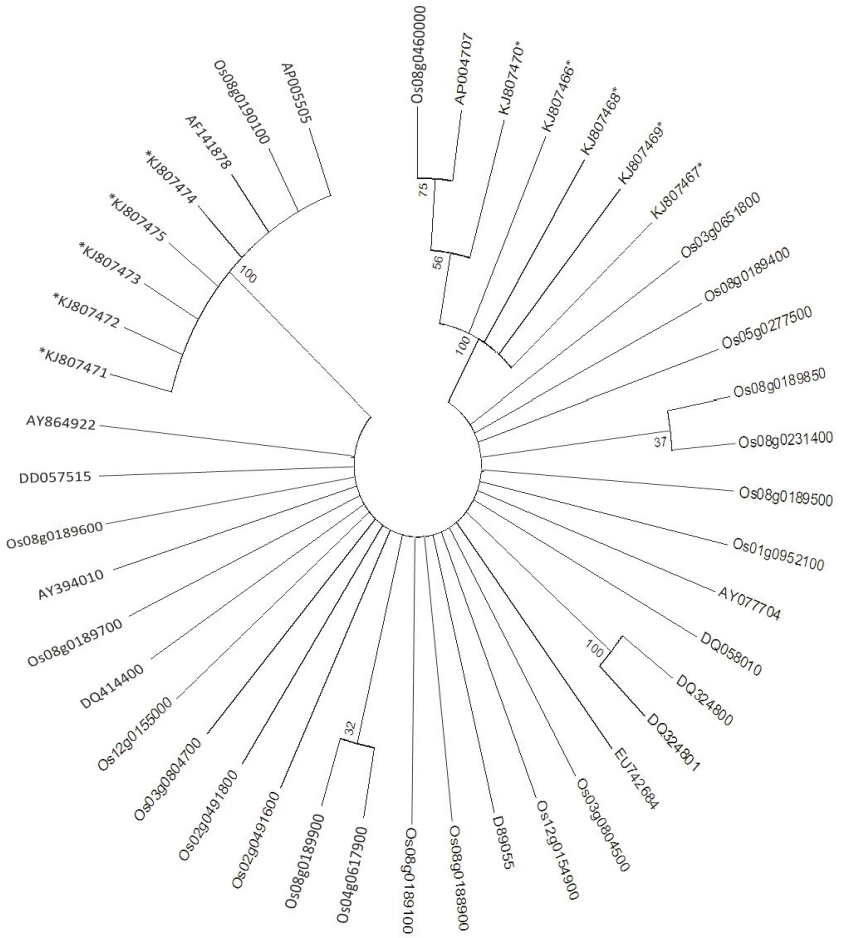

Figure 1. A phylogenetic tree constructed with the maximum parsimony method (500 bootstraps) of putative germin and germin-like protein 1 promoter from studied Pakistani rice varieties with other rice germin and germin-like protein family promoters. The accession number of the studied promoters are labeled with the sign '*' at the end.

highest sequence identity with Os08g0189850 (Figure 1). Sequences with lower distances were located nearer to each other (side by side), while sequences with larger diversity were positioned further away in the tree. Phylogenetic analyses showed that the studied promoter sequences were relatively tightly clustered together.

3.2. Analysis of regulatory elements

With the help of Consite different regulatory elements were explored, present in about 1 Kb promoter region of *OsGLP1* and putative *GerA* genes. The positions of selected important elements on both the plus and minus strands of these promoter sequences were manually mapped by DOG 1.0 User Interface (http://dog.biocuckoo.org/) (Figures 2–4). The main focus was on the location variation of the regulatory elements for the TATA box binding protein (TBP) (Figures 2–4, upper parts). Eleven regulatory elements of TBP were found for *OsGLP1* (Figure 2, upper portion) in which the TBP rich region was positioned from –118 to –141, the TBP medium region was located from position –178 to –425, while 5′ region was very poor in such elements. The highest numbers of TBP regulatory elements were found in putative *GerA* (Acc; AF141878), with the highest number located from position –250 to –347 (Figure 3). It can be said that in these sequences TBP most probably binds with several elements in the 1kb promoter regions rather than with a single one. The numbers, arrangement, and positions of TBP regulatory elements found in *OsGLP1* promoter (Ac; EU742684) were totally different (Figure 4) from those described above (i.e. Figures 2 and 3). In this case, there were only five main regions where TBP had the probability

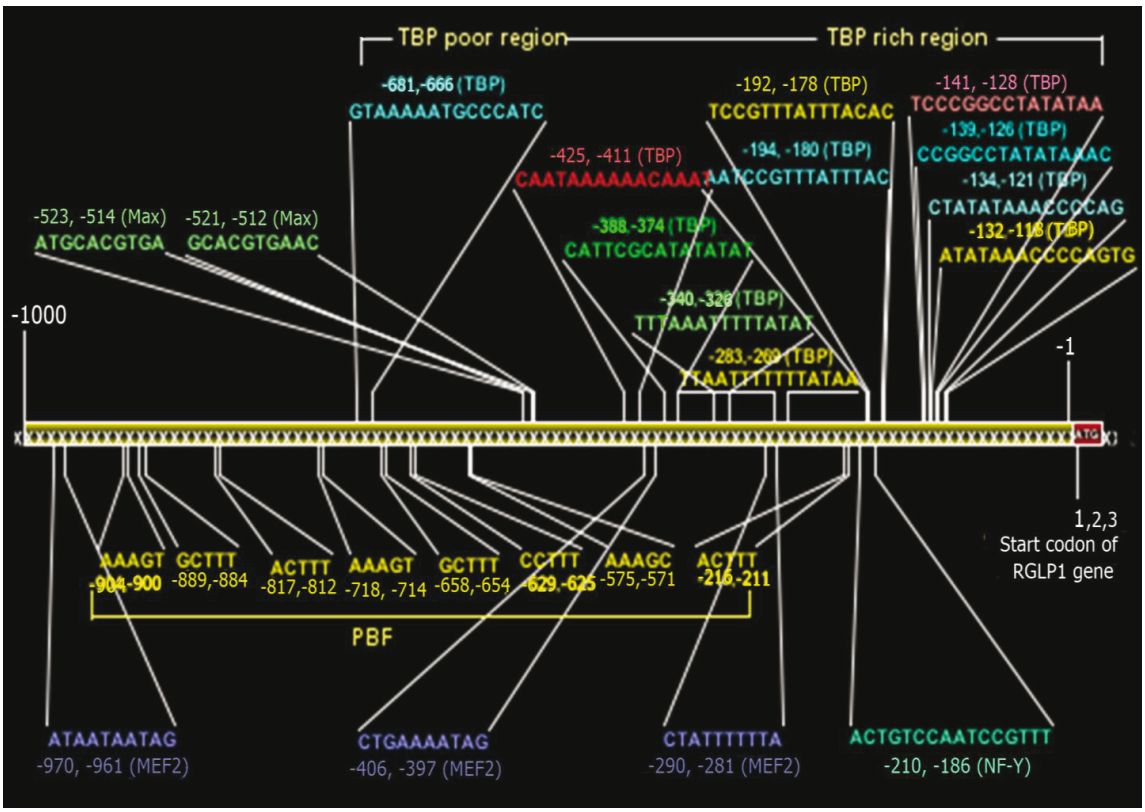

Figure 2. A map of consensus regulatory elements present on both the plus and the minus strands of rice *OSGLP1* promoter sequences (Acc; KJ807466, KJ807467, KJ807468, KJ807469, KJ807470, Os08g0460000 (AP004586), and AP004707).

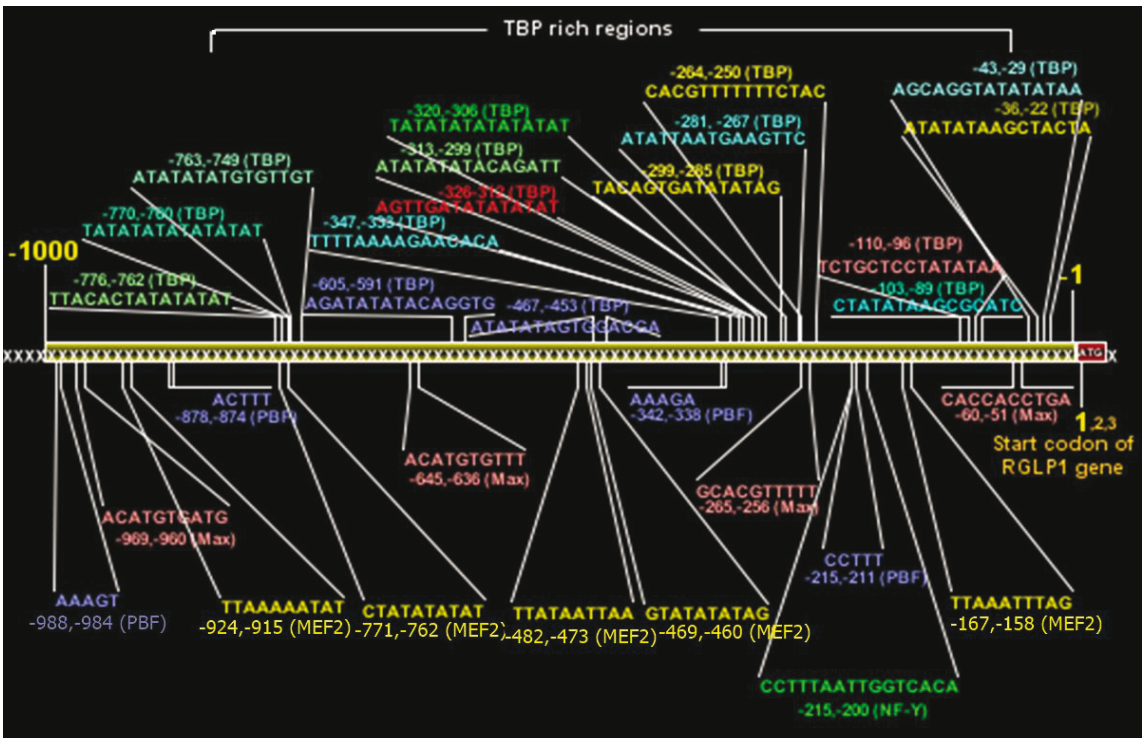

Figure 3. A map of consensus regulatory elements presents on both the plus and minus strands of rice putative germin A promoter sequences (Acc; KJ807471, KJ807472, KJ807473, KJ807474, KJ807475, AF141878, AP005531, and AP005505).

to bind. These positions were: –14 to –35, –228 to –244, –272 to –393, –509 to –532, and –659 to –673 (Figure 4, upper portion). The position variations of other TF's regulatory elements (Max, PBF, MEF2, and NF-Y) for these promoters are also shown in the lower portion of Figures 2–4. It was found that these promoter sequences contained similar regulatory elements; however, these elements were different with respect to positions and numbers in each promoter sequence.

3.3. Interaction of transcription factors with regulatory elements (docking)

To understand the molecular mechanisms of gene regulation, it is necessary to identify protein–DNA interactions as they play a vital role in regulation (Si et al., 2011). In order to study DNA–protein interactions in the present study, five TFs were selected to dock. Prior studies indicate that two thirds of the interactions between residues of proteins and DNA are hydrogen bonds (Angarica et al., 2008). Thus, our study was preferentially focused on hydrogen bonding between protein and DNA residues. The details of the results generated by the HADDOCK webserver are shown in Table 2.

3.4. Evaluation of docking results

Gene expression is regulated by binding of TFs to its specific DNA control elements to maintain cell growth, differentiation, and development. To understand the transcription control of the *OSGLP1* gene, the structures of many TFs and their DNA complexes have been examined. Several highly conserved protein residues were identified and their DNA contact mechanisms were elucidated. TBPs are general eukaryotic transcription factors that participate in the initiation of RNA synthesis by all three eukaryotic RNA polymerases. The carboxy-terminal portion of the TBP is a unique DNA-binding motif/protein fold, adopting a highly symmetric alpha/beta structure that resembles a molecular saddle with two stirrup-like loops (Nikolov and Burley, 1994).

The transcription factor TBP recognizes its corresponding regulatory elements 5′-CTATATAAACCCCAG-3′, 5′-AATCCGTTTATTTAC-3′, and 5′-TTTAAATTTTTATAT-3′ in a long strand of DNA, showing high specificity with them (Figures 5–7). After recognition, formation of hydrogen bonds (shown by yellow dots in Figures 5–7) between them produces a

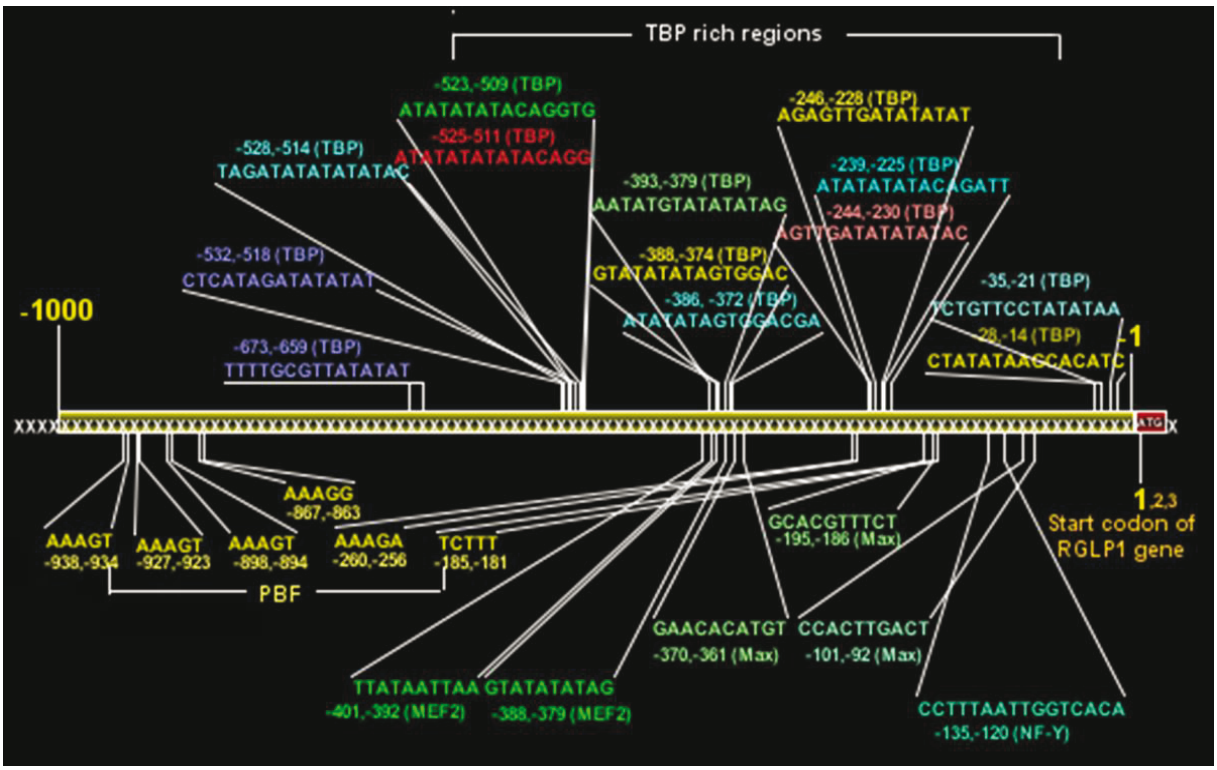

Figure 4. A map of consensus regulatory elements present on both the plus and minus strands of rice putative *OSGLP1* promoter (Acc; EU742684).

Table 2 The HADDOCK web server results. Rows represent information about the models of protein-DNA complexes and their corresponding energies. For each TF, binding affinities with DNA binding sites are shown. The HADDOCK score is a combination of buried surface area, Van der Waals, electrostatic, desolvation, and restraint violation energies.

TF/DNA	HADDOCK Score	Cluster size	RMSD (A*)	Van der Waals energy (Kcal/Mol)	Electrostatic energy (Kcal/Mol)	Desolvation energy (Kcal/Mol)	Restrains violationenergy (Kcal/Mol)	Buried surface area
TBP _ CTATATAAACCCCAG	–84.4	13.0	20.2	–60.1	–852.0	80.4	657.5	5162.0
TBP _ AATCCGTTTATTTAC	63.9	10.0	18.4	–83.2	–712.2	55.8	2341.8	2333.7
TBP _ TTTAAATTTTTATAT	55.1	4.0	89.3	–55.8	–731.3	99.0	1581.7	2049.4
PBF _ GCTTT	42.4	4.0	4.5	–64.9	–401.3	89.5	968.8	1829.3
NF-Y _ ACTGTCCAATCCGTTT	60.5	50.0	7.6	–50.1	–283.9	5.8	11.4	981.5
Max _ ATGCACGTGAAC	–105.3	20.0	3.1	–51.4	–596.9	48.9	171.6	1557.5
MEF2 _ CTGAAAATAG	–122.5	15.0	4.6	–63.8	–560.4	62.5	1731.5	1646.6

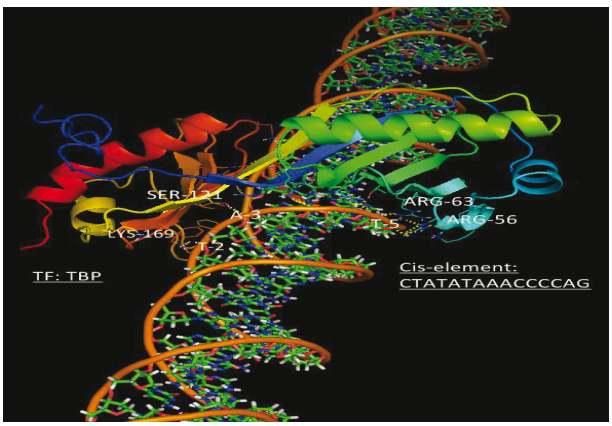

Figure 5. Recognition of the corresponding regulatory element by its TF 'TBP' when docked with a long strand of *OSGLP1* promoter region (-159//GCTCGATCTCGCGTCTCCTCC-CGGCCTATATAAACCCCAGTGGCGTTGCACTGAGCC-CTA//-101). The hydrogen bonding interaction between TF and the Cis-element is shown by yellow dots.

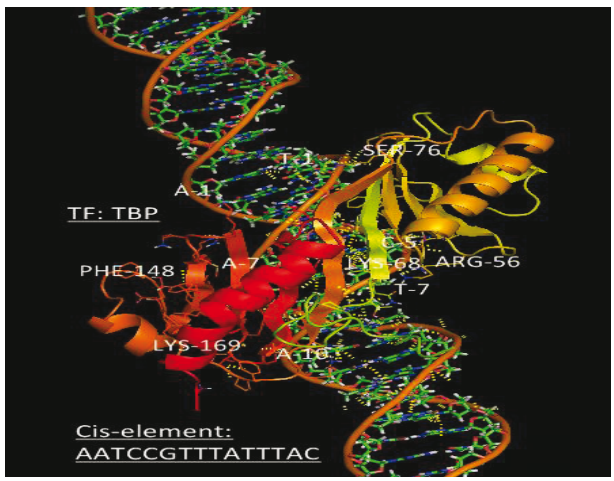

Figure 6. Recognition of the corresponding regulatory element by its TF 'TBP' when docked with a long strand of *OSGLP1* promoter region (-219//AATAACTTTCCCCGTTACACTGTCCAATC-CGTTTATTTACACGAGAACGTACGTACGCGC//-160).

strong complex. More dots between residues represent a larger distance, showing weak bonds between them, and vice versa. In the TBP_5′-CTATATAAACCCCAG-3′ complex, ADE3 has made a hydrogen bond with SER121 (Figure 5). A double hydrogen bond is formed between THY2 and LYS169, as shown in Figure 5. ARG56 was found constructing 2 hydrogen bonds, one with THY5 and the other with ADE5. THY5 is also involved in the formation of another hydrogen bond with ARG63. In the TBP_5′-AATCCGTTTATTTAC-3′ complex, two hydrogen bonds were formed between THY1 and SER76. Similarly, CYT5 bonds with ARG56, ADE7 with PHE148, THY7 with LYS68, and ADE10 with LYS169 were noted. However, in the TBP_5′-TTTAAATTTTTATAT-3′ complex, THY1 bonded with LYS55, THY11 with ARG48, ADE12 with LYS86, and THY19 with SER167 (Figure 7). The TF 'TBP' binds differentially with different regulatory elements. The interaction of TBP with different Cis-elements showed that TBP has more affinity with element 5′-AATCCGTTTATTTAC-3′, with which it forms a much stronger complex than with the other elements. Therefore, this may be the 1st recognition site for TBP among a long stretch of DNA.

The present study showed that TBP not only recognized the (-30) TATA box element but a wide range of TATA box elements located at different locations of the upstream

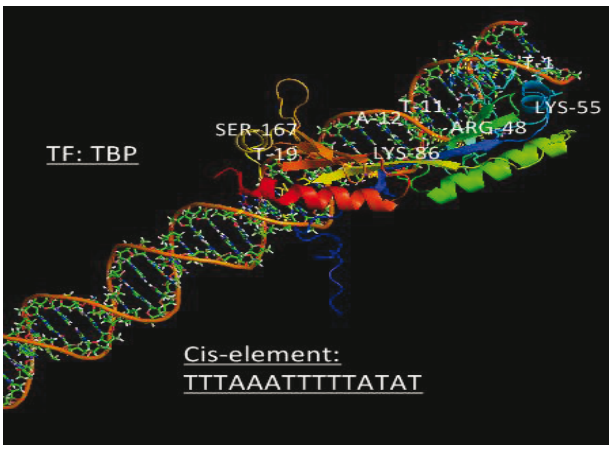

Figure 7. Recognition of the corresponding regulatory element by its TF 'TBP' when docked with a long strand of *OSGLP1* promoter regions (-340//TTTAAATTTTTATATTTGTG-GAGTGATATATTTCATATTAATCTATATTACTATTTTT-TAATTTTTTTATAA//-269).

region of the promoter. TBP is mainly responsible for Pol II transcription, while its role in Pol I and Pol III transcription is less characterized. Genes encoding TBP have been cloned from organisms ranging from bacterial to human, all of which share a conserved C terminal segment of 180 residues that are responsible for all TBP biological functions and interactions. However, the N terminus was not involved in these interactions. The results showed that TBP binds with DNA exclusively by ADE and THY, through minor-groove interactions, similar to the work of Kim and Burley (1994). The full structure of TBP looks like a saddle and consists of four convex alpha helices in the upper surface and ten antiparallel beta sheets in its undersurface. A 3D structure of a full length TBP of *Arabidopsis thaliana* was retrieved from the PDB. Its C terminus interactions with different TATA elements were characterized and we found that only a few amino acids were present in the undersurface concave beta sheets that are involved in DNA binding, while the upper surface helices did not participate in such binding. This study presents the detailed binding interactions of TBP-TATA box complexes and hypothesizes which TATA element has a more stable interaction. The reason that TBP binds with TATA boxes and not GCGC boxes may be the fact that the width of the minor groove is more similar in regions of AT bps rather than of GC, and only in the absence of the protruding amino groups of guanines the intimate contact between the minor groove and the TBP saddle surface is possible (Juo et al., 1996).

Structural information of other protein-DNA complexes was analyzed to check the number of actual and possible hydrogen bonds among them. The structural representations of other complexes are shown in Figures 8–11 and details of possible hydrogen bonds are denoted by yellow dots. Arginine was found to be the most reactive residue in TFs, while in regulatory elements the most sensitive residue was adenine. The second most reactive residue was lysine, with high probability of interactions with adenine and thymine. Those protein-DNA residues

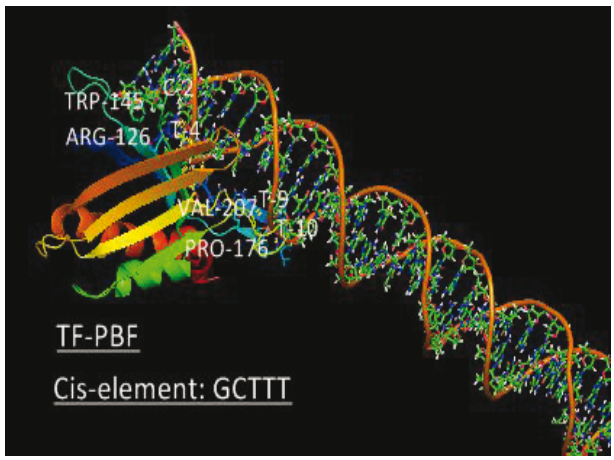

Figure 8. Recognition of the corresponding regulatory element by its TF 'PBF' when docked with a long strand of *OSGLP1* promoter region (-660//GGGCTTTCCCAAATAATGCAGAAATC-CCAGACCTTTTCTTCATCATAGCCCGTCCTCTTCATG-GACCACGGTGTACAGCTTGCTTAAAGC//-571).

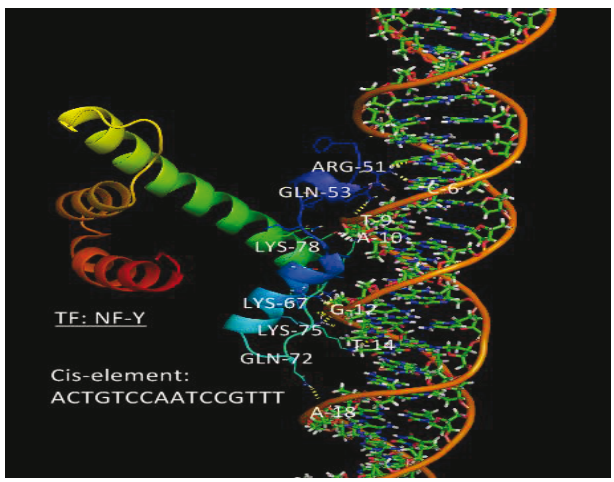

Figure 9. Recognition of the corresponding regulatory element by its TF 'NF-Y' when docked with a long strand of *OSGLP1* promoter region (-228// AATAACTTTCCCCGTTACACTGTCCAATC-CGTTTATTTACACGAGAACGTACGTACGCGC //-169).

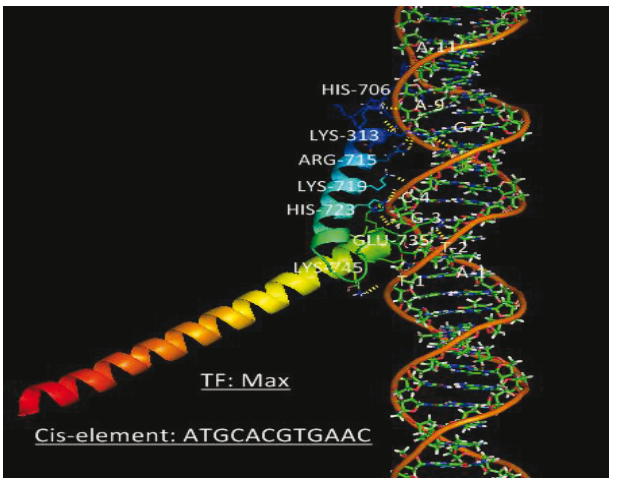

Figure 10. Recognition of the corresponding regulatory element by its TF 'Max' when docked with a long strand of *OSGLP1* promoter region (-536//GCCATCCGCATGCATGCACGTGAACTACTCGATCCACCCC//-495).

Figure 11. Recognition of the corresponding regulatory element by its TF 'MEF2' when docked with a long strand of *OSGLP1* promoter region (-417//AACAAATTAAACTGAAAATAGTTATCGCACATTCG//-383).

that have a distance of less than 3.5 Å are assumed to be realistic for hydrogen bond formation (Si et al., 2011). The studies of hydrogen bond formation are significant because these hydrogen bonds confer stability and specificity in the protein–DNA complexes (Coulocheri et al., 2007).

Regulation differences of the TFs can be recognized by its binding mechanism when bound to DNA (Sheng et al., 2002). Our results indicate that obvious binding differences occur in the studied TFs, showing that only a few regulatory elements (those that can create more stable bonds) will initially be involved in *OsGLP1* and *OsGerA* genes regulation. The structural comparison of different TFs and their DNA complexes indicates that even highly conserved TFs can adopt different local structures when they contact another DNA binding sequence, and their interaction and stability depend on the distance and number of hydrogen bonds formed.

The study concludes that *OsGLP1* and *OsGerA* genes are regulated by promoter sequences that are very dissimilar from each other. For all promoter sequences, higher numbers of TBP regulatory elements were found in close vicinity of the corresponding genes, showing that TBP most probably binds very near to the transcription start site. Binding studies reveal that the DNA binding domain of a conserved C terminus TBP recognizes the AATCCGTTTATTTAC with higher affinity (by forming more hydrogen bonds) as compared with other studied TFs. Preferential recognition of the AATCCGTTTATTTAC by TBP may be due to the TTTATTT sequence being symmetrical end for end, and being identical upon inversion, while other sequences are not symmetrical. The study also found that TBP can only bind to ADE and THY; this may be due to the absence of the protruding amino groups of guanines.

Acknowledgement

We are thankful to Higher Education Commission, Pakistan, for providing financial assistance.

References

Agarwal G, Rajavel M, Gopal B, Srinivasan N (2009). Structure based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. Plos One 4: e5736.

Angarica VE, Pérez AG, Vasconcelos AT, Collado-Vides J, Contreras-Moreira B (2008). Prediction of TF target sites based on atomistic models of protein-DNA complexes. BMC Bioinformatics 9: 436.

Banerjee J, Maiti MK (2010). Functional role of rice germin-like protein1 in regulation of plant height and disease resistance. Biochem Bioph Res Co 394: 178–183.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000). The Protein Data Bank. Nucleic Acids Res 1: 235–242.

Breen J, Bellgard M (2010). Germin-like proteins (GLPs) in cereal genomes: gene clustering and dynamic roles in plant defence. Funct Integr Genomic 10: 463–476.

Burley SK (1996). X-ray crystallographic studies of eukaryotic transcription initiation factors. Philos T Roy Soc B 35: 483–489.

Carrillo M, Goodwin P, Leach J, Leung H, Cruz CV (2009). Phylogenomic relationships of rice oxalate oxidases to the cupin superfamily and their association with disease resistance. Qtl. Rice 2: 67–79.

Carter C, Thornburg RW (1999). Germin-like proteins: structure, phylogeny, and function. J Plant Biol 42: 97–108.

Carter C, Thornburg RW (2000). Tobacco nectarin I. Purification and characterization as a germin-like, manganese superoxide dismutase implicated in the defense of floral reproductive tissues. J Biol Chem 275: 36726–36733.

Coulocheri SA, Pigis DG, Papavassiliou KA, Papavassiliou AG (2007). Hydrogen bonds in protein-DNA complexes: where geometry meets plasticity. Biochimie 89: 1291–1303.

Cramer GR, Ergül A, Grimplet J, Tillett RL, Tattersall EAR, Bohlman MC, Vincent D, Sonderegger J, Evans J, Osborne C et al (2007). Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomic 7: 111–134.

Davidson R, Reeves P, Manosalva P, Leach J (2009). Germins: a diverse protein family important for crop improvement. Plant Sci 3: 43–55.

De Vries SJ, Dijk MV, JBonvin AMJ (2010). The HADDOCK web server for data-driven biomolecular docking. Nature Protocols. In: Grasser, KD, editors. Regulation of Transcription in Plants: Vol. 29. Blackwell Publishing 5: 883–897.

Dunwell JM, Gane PJ (1998). Microbial relatives of seed storage proteins: conservation of motifs in a functionally diverse superfamilly of enzymes. J Mol Evol 46: 147–154.

Dunwell J, Gibbings JG, Mahmood T, Naqvi S (2008). Germin and germin-like proteins: evolution, structure, and function. Crit Rev Plant Sci 27: 342–375.

Durbin R, Eddy SR, Krogh A, Mitchison G (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids. 1st ed. Cambridge, UK: Cambridge University Press.

Federico ML, Iñiguez-Luy FL, Skadsen RW, Kaeppler HF (2006). Spatial and temporal divergence of expression in duplicated barley germin-like protein-encoding genes. Genetics 174: 179–190.

Fitch W (1971). Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20: 406–416.

Houde M, Diallo AO (2008). Identification of genes and pathways associated with aluminum stress and tolerance using transcriptome profiling of wheat near-isogenic lines. BMC Genomics 9: 400.

Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, Dickerson RE (1996). How Proteins Recognize the TATA Box. J Mol Biol 261: 239–254.

Ke Y, Han G, He H, Li J (2009). Differential regulation of proteins and phosphoproteins in rice under drought stress. Biochem Bioph Res Co 379: 133–138.

Kim HJ, Pesacreta TC, Triplett BA (2004). Cotton-fiber germin-like protein. II: Immunolocalization, purification, and functional analysis. Planta 218: 525–535.

Kim JL, Burley SK (1994). 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. Nat Struct Biol 1: 638–653.

Knecht K, Seyffarth M, Desel C, Thurau T, Sherameti I, Lou B, Oelmüller R, Cai D (2010). Expression of bvglp-1 encoding a germin-like protein from sugar beet in *Arabidopsis thaliana* leads to resistance against phytopathogenic fungi. Mol Plant Microbe In 23: 446–457.

Lou Y, Baldwin IT (2006). Silencing of a germin-like gene in *Nicotiana attenuata* improves performance of native herbivores. Plant Physiol 140: 1126–1136.

Lu M, Han Y, Gao J, Wang X, Li W (2010). Identification and analysis of the germin-like gene family in soybean. BMC Genomics, 11: 620.

Mahmut Ç (2000). Germin, an oxalate oxidase, has a function in many aspects of plant life. Turk J Biol 24: 717–724.

Manosalva P, Davidson R, Liu B, Zhu X, Hulbert S, Leung H, Leach J (2008). A germin-like protein gene family functions as a complex qtl conferring broad-spectrum disease resistance in rice. Plant Physiol 149: 286–296.

Membré N, Bernier F, Staiger D, Berna A (2000). Arabidopsis thaliana germin-like proteins: common and specific features point to a variety of functions. Planta 211: 345–354.

Nikolov DB, Burley SK (1994). 2.1 A resolution refined structure of a TATA box-binding protein (TBP). Nat Struct Biol 1: 621–637.

Park CJ, Kim KJ, Shin R, Park JM, Shin YC, Paek KH (2003). Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. Plant J 37: 186–198.

Requena L, Bornemann S (1999). Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. Biochem J 343: 185–190.

Richards EJ (1997). Preparation of plant DNA using CTAB. In: Ausubel F, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, editors. Short protocols in molecular biology. New York, NY, USA: John Wiley, pp. 2.10–2.11.

Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res D91–D94.

Sheng W, Rance M, Liao X (2002). Structure comparison of two conserved HNF-3/fkh proteins HFH-1 and genesis indicates the existence of folding differences in their complexes with a DNA binding sequence. Biochemistry-US 41: 3286–3293.

Si J, Zhang Z, Lin B, Schroeder M, Huang B (2011). Meta DB site: a meta approach to improve protein DNA binding sites prediction. BMC Syst Biol 1: S7.