1. Introduction

Plants, being sessile, are exposed to a wide variety of environmental stresses. These environmental stresses result in decreased productivity and yield losses. More than 50% yield forfeiture in major crops around the world is caused by drought and salinity (Bray, 2004). These environmental stresses are becoming serious challenges for prime agricultural output. Plants respond to environmental stress in a highly complex and integrated way involving an array of molecular, physiological, and biochemical changes (Bartels and Sunkar, 2005). Major metabolic responses to abiotic stresses including drought tolerance are perception, signal transduction, gene expression, and metabolic changes revealing drought tolerance (Agarwal et al., 2006). The genes induced by drought stress not only work in cell protection by protein production but are also involved in regulation of downstream genes for signal transduction. These gene products are usually divided into 2 groups: proteins that protect cells from the effects of water stress, and others that regulate the signal transduction by modulating gene expression and hence are probably involved in stress tolerance (Fowler and Thomashow, 2002; Seki et al., 2002). These include many transcription factors (TFs), including dehydration-responsive element binding (DREB) protein (Agarwal et al., 2006; Yamaguchi-Shinozaki and Shinozaki, 2009). TFs regulate many abiotic stress-related genes in plants and enhance tolerance against abiotic stress (Kirch et al., 2005; Lee et al., 2006). TFs bind with the cis-elements of stress-related gene promoters and upregulate many downstream genes (Agarwal and Jha, 2010). Microarray studies in many plants show that there are copious pathways that autonomously respond to abiotic stress in an abscisic acid (ABA)-dependent or ABA-independent manner (Umezawa et al., 2006). Some cis-acting elements like DRE/CRT function in ABA-dependent or ABA-independent gene expression during abiotic stresses (Nakashima and Yamaguchi-Shinozaki, 2010). ABA-independent gene expression involves 2 regulons: the cold binding factor/dehydration-responsive element (CBF/DRE), and the NAC and zinc-finger homeodomain (ZF-HD). DREB genes are among the most studied groups of

Abstract: The family of AP2/ERF plant specific transcription factors, including dehydration-responsive element binding (DREB) proteins, has been reported to activate and bind with the dehydration responsive element/C repeat (DRE/CRT) of stress-inducible gene promoters. We amplified DREB1A from Oryza sativa var. IR6 (807 bp) and subjected it to protein homology modelling and phylogenetic analysis. BLAST sequence analysis indicated high similarity (99%) with the O. sativa japonica group. Phylogenetic analysis was carried out with other 12 DREB sequences from cereals/grasses using Clustal Omega. Protein secondary structure and 3D models were determined using SOPMA and MODELLER v9.10, respectively. PROSA was used to determine the excellence and consistency of models. The results indicated that the Z-score value was ~2.94 and 79.5% of the residues were found in the favoured region. A psi/phi Ramachandran plot was determined using PROCHECK. The molecular docking of the GCC-box binding domain (GBD) of the protein under investigation was done using atomic coordinates of Arabidopsis thaliana GBD- and GCC-box containing DNA, respectively. The tertiary structure of the modelled protein and template were found to be very similar as predicted by a root mean square deviation of 0.968 Å. It seemed that both the proteins interacted with the major groove of DNA using β-sheets. We attempted to expose the molecular basis of O. sativa var. IR6 DREB1A protein interaction with the target promoter sequence. The results highlighted that this gene could be a good candidate for production of abiotic stress-tolerant crop(s) in the future.

Key words: Abiotic stress, DREB1A, homology modelling, docking

* Correspondence: muhammadnawaz@gcuf.edu.pk
TFs related to biotic and abiotic stress tolerance in plants. They are involved in expression of many stress-inducible genes imparting abiotic stress tolerance (Hussain et al., 2011). Signal transduction pathways in low temperature and dehydration involve DREB1 and DREB2 proteins respectively in Arabidopsis thaliana. Several DREB1/ DREB2 homologous genes have been isolated from many plants including wheat, rice, barley, rye, sorghum, and oat (Nakashima et al., 2009). The DREB proteins have a GCC-box binding domain that is a specific nucleotide sequence of DRE (Agarwal and Jha, 2010). Conclusively, the DRE binding protein transcription pathway is a promising candidate to further explore mechanisms of drought tolerance (Pasquali et al., 2008).

It is important to investigate the structural and functional relation of transcription factors to understand the molecular mechanism for recognition and expression of target genes at the genome level (Garg et al., 2008). Computational tools are used to understand the physicochemical properties of proteins leading towards structure/function prediction. In silico approaches offer solutions to the problems like time constraints, high costs, and more labour usually faced in experimental methods. In the present study we have carried out in silico analysis and protein homology modelling of the DREB1A gene isolated from O. sativa var. IR6, a drought-resistant nonaromatic rice cultivar of Pakistan. The derived protein was docked with a DNA double helix having a GCC-box to reveal the residues involved in protein–DNA interface. It is anticipated that these findings will offer more insights into the structural and functional roles of the DREB1A protein involved in the abiotic stress tolerance mechanism.

2. Materials and methods

The seeds of O. sativa var. IR6 were taken from the Rice Research Station of Kala Shah Kako, Pakistan. Seeds were germinated in small plastic bags having mud and compost in a 1:1 ratio in a growth room at 27 °C under cool white fluorescent light.

2.1. Genomic DNA isolation

Three to 5 young growing leaves from each plant were collected in a triplicate and transported to the laboratory immediately after being placed in liquid nitrogen. Genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method (Navaz et al., 2009).

2.2. Primer design and PCR amplification

The complete coding sequence of DREB1A was retrieved from the National Center for Biotechnology Information (NCBI). Primers were designed using the free online tool Primer3 (http://frodo.wi.mit.edu/) to amplify this sequence from the genomic DNA of the indigenous rice cultivar. The sequence of the forward and the reverse primers was as follows:

DREB1AF: GGAGCAACGAGAACCACA
DREB1AR: GCATCGGAAGCCAGAAAAGA

The reaction mixture (25 µL) used for the amplification of the 807-bp gene contained 2.5 µL of 10X PCR buffer, 2.5 µL of 25 mM MgCl₂, 2.0 µL of 2.0 mM dNTPs, 25 ng/µL genomic DNA, and 1.5 U Taq polymerase (Fermentas, USA). The amplification reaction was performed in Mastercycler Gradient (Eppendorf) programed for 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 60 s, followed by a final single step extension at 72 °C of 5 min. The amplified product was confirmed on 2% agarose gel, prepared in 1X TBE buffer, and detected by poststaining with ethidium bromide (20 ng/400 mL of distilled water). The amplified product was eluted from gel using a DNA elution kit (QIAGEN, the Netherlands) and purified using the Wizard Genomic DNA Purification Kit. The purified DNA product was used as a template for dye terminator cycle sequencing reaction. The sequencing was done in an ABI Prism 310 genetic analyser.

2.3. Sequence analysis

BLAST searches were carried out to align our isolate with already existing DREB1A sequences in the database. Multiple sequence alignments were performed to find the conserved regions in the predicted amino acid sequence. The phylogenetic tree was constructed with the already reported C-repeats/DRE binding factors from different cereals/grasses using Clustal Omega. The ProtParam tool (Wilkins et al., 1999) was used to determine the molecular weight, theoretical isoelectric point (pI), atomic composition, amino acid composition, estimated half-life, instability index, aliphatic index, extinction coefficient, and grand average of hydrophobicity (GRAVY). Secondary structure analysis was done using SOPMA (Geourjon and Deléage, 1995). The 3D models were predicted using Python-based protein modelling software, Modeller v9.10 (Eswar et al., 2007). After generating 3D models, the psi/phi Ramachandran plot was determined using PROCHECK, which helped in evaluating backbone conformation. The Z-score was determined by the PROSA web tool. The model was further evaluated with ERRAT (Colovos and Yeates, 1993). Furthermore, visualisation of the generated model was performed using UCSF Chimera 1.5.3.

2.4. Molecular docking

The structure of O. sativa var. IR6 DREB1A was further explored using a molecular docking and interaction approach. Energy of the O. sativa var. IR6 DREB1A structure was minimised using MOE software before docking. The conserved domains were identified using the pfam and SMART databases. The GCC-box containing...
double helical DNA was retrieved from the PDB database using PDB ID: 1GCC. The docking was carried out with the identified conserved domain of the DREB1 gene using the Hex server after removing the already bound domain (Macindoe et al., 2010). After DNA–protein docking, docked complexes were subjected to binding analysis. UCSF Chimera software was used to visualise binding among docked molecules.

3. Results and discussion
The isolated DNA from O. sativa var. IR6 was run on 1% agarose gel for qualitative analysis and for qualitative analysis a NanoDrop 2000C spectrophotometer (Thermo Scientific, USA) was used. The results revealed that the DNA concentration of 25 ng/µL resulted in the best amplification of the 807-bp fragment (Figure 1). The sequencing of osIR6DREB1A indicated high-quality results implied by sharp peaks. The genomic DNA sequence was compared with the cDNA clone and was found not to contain any introns as evident from the results of a previous study (Xiong and Fei, 2006). The sequenced data of osIR6DREB1A showed 63% to 99% homology with different cereals, the maximum being with the O. sativa var. japonica DREB1A gene (accession no. AP006859), which was used to design primers (Table). BLAST (Altschul et al., 1997) searches revealed that the amplified gene contained a conserved DNA-binding domain having high homology with the ethylene-responsive element binding protein (EREBP/AP2) domain. Multiple sequence alignment of osIR6DREB1A protein signposted high homology at the nuclear localisation signal at the N-terminal, moderate homology at the AP2 domain, and low conservation at the acidic C-terminal. It was fond that DREB1A proteins generally in cereals and specifically in rice had conserved valine at the 14th and 19th positions of the AP2/EREBP domain (Figure 2), probably essential for protein recognition and binding to target DNA (Sakuma et al., 2006). Phylogenetic analysis of the selected accessions from multiple sequence alignment of osIR6DREB1A using Clustal Omega (Sievers et al., 2011)

<table>
<thead>
<tr>
<th>DRE/species</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E-value</th>
<th>Identity</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCBF3 O. sativa japonica group</td>
<td>470</td>
<td>470</td>
<td>100%</td>
<td>3e-166</td>
<td>99%</td>
<td>AAQ23983.1</td>
</tr>
<tr>
<td>BREB1B O. brachyantha</td>
<td>335</td>
<td>335</td>
<td>100%</td>
<td>4e-113</td>
<td>79%</td>
<td>ABG73450.1</td>
</tr>
<tr>
<td>DREB1A like S. italicca</td>
<td>286</td>
<td>286</td>
<td>100%</td>
<td>1e-93</td>
<td>68%</td>
<td>Xp004957404.1</td>
</tr>
<tr>
<td>DREB factor 3 L. perenne</td>
<td>281</td>
<td>281</td>
<td>100%</td>
<td>5e-92</td>
<td>65%</td>
<td>AAX57275.1</td>
</tr>
<tr>
<td>Ap2/ERFB protein Z. mays</td>
<td>275</td>
<td>275</td>
<td>89%</td>
<td>3e-89</td>
<td>74%</td>
<td>DAA62352-1</td>
</tr>
<tr>
<td>DREB1A F. arundinacea</td>
<td>271</td>
<td>271</td>
<td>100%</td>
<td>1e-87</td>
<td>65%</td>
<td>CAG30550.1</td>
</tr>
<tr>
<td>DREB factor 6 F. pratensis</td>
<td>267</td>
<td>267</td>
<td>100%</td>
<td>2e-86</td>
<td>64%</td>
<td>ABL96271.1</td>
</tr>
<tr>
<td>CBF1 protein A. biuncialis</td>
<td>264</td>
<td>264</td>
<td>100%</td>
<td>6e-85</td>
<td>67%</td>
<td>CBX87015.1</td>
</tr>
<tr>
<td>Hypothetical protein S. bicolor</td>
<td>263</td>
<td>263</td>
<td>100%</td>
<td>7e-85</td>
<td>67%</td>
<td>XP00246269.1</td>
</tr>
<tr>
<td>DREB1A like B. distachyon</td>
<td>255</td>
<td>255</td>
<td>100%</td>
<td>1e-81</td>
<td>68%</td>
<td>XP003578468.1</td>
</tr>
<tr>
<td>CBFIII aD-6 T. aestivum</td>
<td>251</td>
<td>251</td>
<td>100%</td>
<td>3e-80</td>
<td>66%</td>
<td>ABK55360.1</td>
</tr>
<tr>
<td>C/Repeat binding 6 H. vulgare</td>
<td>231</td>
<td>231</td>
<td>100%</td>
<td>2e-73</td>
<td>67%</td>
<td>ACA29489.1</td>
</tr>
</tbody>
</table>
resulted in 2 main clusters, clusters 1 and 2. Cluster 2 was subdivided into 2 subclusters, 2A and 2B. Subcluster 2B divided the accessions into 3 groups, separating the genus *Oryza*. The accession under investigation, *osIR6DREB1A*, was found to be a close relative of *osDREB1A* from *O. sativa* var. indica, suggesting the common origin of structure and function (Figure 3).

### 3.1. Insights from the structure

The structure-based sequence analysis studies of the DREB1 protein derived from the genome of *O. sativa* var. IR6 from Pakistan revealed that the DREB1 protein had a molecular weight of 25376.0 Da and pI of 5.41. An isoelectric point below 7 indicated a negatively charged protein. The instability index was computed to be 59.27, classifying the protein as unstable. The N-terminus of the sequence was considered to be M (Met). The GRAVY of −0.422 indicated that the protein was hydrophilic. The secondary structure revealed that it had 31.38% alpha helices, 8.8% beta turns, 3.35% extended strands, and 51.88% extended coils. A previous study had indicated that the protein's 3D structure was very important in understanding the protein interactions, functions, and localisations (Parasuram et al., 2010). Homology modelling is the most common structure prediction method. Moreover, finding a best matching template using similarity searching programs like PSI BLAST against a PDB database has been considered the basic step in homology modelling, the most common structure prediction method. Templates were selected based on their sequence similarity with query sequence. PDB ID 2GCC_A, with 44% homology, was selected as a template for the structure prediction using MODELLER v9.10 software (Figure 4). After protein structure prediction, the quality and reliability of structure was assessed by several structure assessment

![Figure 2. Sequence and modular domain analysis of translated osIR6DREB1A. Nuclear localisation signal region (NLS) at the N-terminal is indicated with blue dotted underline. The AP2 conserved domain is indicated with red thick underline highlighted with 14- and 19-position valine.](image-url)
methods, including Z-score and Ramachandran plots. The Z-score is indicative of overall model quality and is used to check whether the input structure is within the range of scores typically found in native proteins of similar size. PROSA was used to find the Z-score of the predicted structure. The Z-score of the protein was –2.94. PROCHECK was used to determine the Ramachandran plot to assure the quality of the model (Laskowski et al., 1993) The result of the Ramachandran plot showed 79.5% of residues in the favoured region (Figure 5). The Z-scores and Ramachandran plot confirmed the quality of the homology model of the osIR6DREB1 protein.

Figure 4. The osIR6DREBA gene-derived protein 3D model using MODELLER v9.10.

Figure 5. Ramachandran plot of 3D protein structure derived from osIR6DREB1A gene.
3.2. Domain analysis and molecular docking

Positive amino acids like Lys, Arg, and/or His have been considered very important in DNA–protein interaction studies. The known binding site of the GCC-box binding domain (GBD) of *A. thaliana* strongly supported this concept (Allen et al., 1998). A putative DNA-binding conserved domain was identified in the present study. This DNA-binding domain was also found in transcription regulators in plants, such as APETALA2 and EREBP. The GCC-box containing double helical DNA was docked with the DNA-binding domain of the *osIR6DREB1* gene-derived protein using the Patchdock online server. The GCC-box containing double helical DNA was docked with the DNA-binding domain of the *osIR6DREB1* gene-derived protein using the Hex online server. The Hex online server provides different orientations of docked complexes on the basis of docking correlation by root mean square deviation and steric clashes. The top-ranking docked complex is chosen on the basis of docking score, which is based on the energy of the complex. In this case, the best complex with a minimum E value of −331 was selected for further analysis. After postdocking analysis, it was assured that the DNA-binding domain bound efficiently with the GCC containing DNA (Figure 6). The molecular docking

![Diagram](image)

**Figure 6.** Docking of DREB1GCC binding domain with DNA double helix with conserved residues AGCCGCC. a) β-strand showed interaction with the conserved residues GCC box. b) Surface orientation of the interaction between GCC binding domain and GCC box. c) Interacting residues of *O. sativa* var. IR6 DREB1A protein were found to be Arg68 and Gly105.
studies revealed that like the GBD of *A. thaliana* (Allen et al., 1998), the *O. sativa* var. IR6 DREB1A GBD also used its 3-stranded antiparallel β-sheet to bind with the major groove of the DNA. In general, the α-helices of the zinc finger-containing protein were found to be involved in DNA–protein interactions (Tan et al., 2003). Conversely, it was also reported that plant-origin GBD and other DNA-interacting proteins may also use their β-sheet(s) for this purpose (Mazarei et al., 2002). The GBD of *O. sativa* var. IR6 had 3 antiparallel β-sheets recognising the nonpalindromic sequence of 9 consecutive DNA base pairs. This was found to be in good agreement with results of previous studies (Allen et al., 1998; Ouellet et al., 1998).

Furthermore, the proposed structure of *O. sativa* var. IR6 DREB1A GBD was well preserved with *A. thaliana*. This was found to be in good agreement with results of previous studies (Allen et al., 1998; Ouellet et al., 1998).

In conclusion, the plant responses and adaptations to different abiotic stresses could be best understood using multidisciplinary tools and techniques including physiology, biochemistry, and genomics. The engineering of biosynthetic pathways related to different abiotic stresses in plants is being considered as the most promising method to improve stress tolerance (Hong et al., 2000). Transcription regulation of abiotic stress-related genes is a potential area of interest for improving stress tolerance in plants; nevertheless, identification of transcription factors controlling the sustained response is the key to success. The results of DREB1A amplification and in silico characterisation from *O. sativa* var. IR6 could be utilised in molecular docking studies using various permutations of DNA to reveal more insight about DNA–protein interactions. Moreover, this could be explored in genetic transformation of elite sugarcane (*Saccharum officinarum* L.) cultivars to improve drought tolerance.

**Acknowledgements**

This study was a part of PhD work supported by the Higher Education Commission of Pakistan under the Indigenous PhD Fellowship Programme. The cooperation of the Agricultural Biotecnology Research Institute, Ayub Agriculture Research Institute, Faisalabad, for technical help and support is gratefully acknowledged. The authors also thank Dr LJ (Luit) de Kok, Laboratory of Plant Ecophyology, University of Groningen, the Netherlands, for his valuable suggestions and kind support.

References


