MOLECULAR BEACON PROBE BASED PROMOTER MOTIFS VALIDATION IN ANOXIA RESPONSIVE DIFFERENTIALLY EXPRESSED GENES AND THEIR IN SILICO INTERACTION STUDIES WITH AP2/EREBP TF IN RICE (ORYZA SATIVA L.)

GOPAL KUMAR Prajapati, DEV MANI Pandey*
Department of Bio-Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand-835215
Email: dmpandey@bitmesra.ac.in

Received: 10 Dec 2014 Revised and Accepted: 30 Dec 2014

ABSTRACT

Objective: Progressive evolution in molecular biology revealed the differential expression of genes and their regulatory mechanism in rice under anoxia. In addition to that the consensus promoter motifs (GCC and TCC box) were identified in differentially expressed genes (DEGs) from microarray analysis through in silico study. These promoter motifs need to be validated and their interaction study with the transcription factors (TFs) are essential.

Methods: To unravel the regulatory mechanism in rice during anoxia, we identified and validated the promoter motifs through Molecular Beacon Probes (MBP) based Real Time PCR. In silico protein-DNA interaction was studied between highly up-regulated APETALA2/Ethylene-responsive element binding proteins (AP2/ERBP) TF under anoxia and validated promoter motifs through the HADDOCK and SiteMap module.

Results: It was identified that consensus promoter motif GCC and TCC box were present in highly up-regulated methyl-transferase domain containing protein gene (MT) and down-regulated RhoGAP domain containing protein gene (RG), respectively.

Conclusion: These promoter motifs were validated through MBP and further their interaction with AP2/ERBP shows the significant binding affinity towards GCC and TCC box present on MT and RG, respectively.

Keywords: DEGs, Anoxia, DEG, MBP, HADDOCK, SiteMap.

INTRODUCTION

Rice is one of the most important basic food crops. More than half of the world’s population depends upon rice [1]. However, abiotic stress is a major limiting factor of rice productivity worldwide [2]. Abiotic stress like submergence, drought, salinity, cold, anoxia are the most prominent factors which affect the plant growth and development. Importantly, among natural hazard flooding plays most hazardous role for the standing crop leads to death during complete submergence for 1 to 2 weeks of most rice cultivars [3]. In consequence of submergence rice plant suffers from oxygen deficiency. Germination of the rice coleoptile under anoxia is highly infrequent characteristic by extending the coleoptile, above the water surface is a key feature of rice to sustain under anaerobic condition [4-6]. However, low oxygen stresses regulate the different metabolic pathways and differential expression of genes [7]. Moreover, various studies in response to anoxia [8, 9], anaerobic response elements (AREs) with their binding sites [10] and factors regulating the wide range of differential expression of genes in anoxic rice coleoptile have also been reported [8]. However, under anoxia it is still largely unknown key regulatory mechanisms of rice coleoptile germination and elongation along with the differential expression of the genes.

During rice germination and coleoptile elongation TFs MYB,zip, AP2/ERF and ZnF play a potential role in controlling the transcription of sucrose metabolism and fermentation genes under anaerobic condition [6]. Moreover the differentially expressed genes (DEGs) and TF family, including AP2-EREBP, MYB, bHLH, WRKY, zip and NAC were identified which are involved in salinity and submergence stresses [11]. More specifically, AP2-EREBP/ERFs TF has been found to be involved in growth, development, metabolic regulation under biotic and abiotic responses [12]. This superfamily TF divided into subfamily TF AP2, RAV, CBF/DREB and ERF genes, which play a variety of roles throughout the plant life cycle and key regulator in various biotic and environmental stresses. TFs encoded by AP2/EREBP genes contain the highly conserved AP2/ERF DNA binding domain [13-14]. However, proteins encoded by ERF subfamily genes bind to the core motif AGCGGCC (GCC box) mainly a response to pathogenesis and wounding [15-17]. A similar study was reported on tomato, Ethylene- Responsive Factor (ERF) transcription factor Pt4, which binds the GCC box (cis-element) that is present in the promoters of many Pathogenesis-Related (PR) genes [18]. Whereas, CBF/DREB ERF subfamily gene TF also recognizes C-repeats cis-acting element, A/GCGGC, which is often associated with ABA, drought and cold responsive genes [19, 20]. Additionally, in rice Submergence1 (Sub1) locus encoding three ERF transcriptional regulators. Sub1 TF gene is a key breakthrough for the submergence tolerance, which often regulates other genes by binding to their consensus promoter motif, GCC box [21]. Similarly, it has been reported from the promoters of anoxerob stress responsive genes statistically significant, common and consensus promoter motifs are detected by in silico analysis in majority of promoters [22]. Further, in silico study of anoxia coleoptile rice microarray data [8] reveals that consensus promoter motif GCC box (GCCGCC) and the TCC box (TCTC) was highly conserved in the promoter of up-regulated differently expressed genes (DEGs) and down-regulated DEGs respectively [23]. Furthermore, GCC in the UR-DEG promoter of Ubiquinol Cytochrome C chaperone gene (UCC) identified by MEME (v 4.5.0, http://meme.nbcr.net/meme_4_5_0/cgi-bin/meme.cgi) online tool and their validation done through MBP based on Real Time PCR [24].

Several probes based techniques have been reported like Molecular beacon, Minor groove binding (MGB) assays used to identify the specific sequences in the nucleic acids based on the RealTime PCR [24-28]. More specifically, MBP increases the sensitivity and precision over the conventional PCR without post-reaction analysis for the detection and as well as quantification of target genes [29].

The genes and their regulatory TF are central to the expression of the functional genes under abiotic condition. Remarkably, genes facilitate their expression by binding different transcription factor in the promoter region. Hence the interaction of the TFs with the cis-element are the key feature of the gene regulation and its expression [30]. In these complexes amino acids and nucleotide sequences have participated in the formation of the interactive structure of Protein-DNA complexes, which determine the many functional
characteristics [31]. On the basis of sequence and structural information several methods and software have been developed for prediction of the binding site and interacting residues in TFs [32,33]. Subsequently the prediction of complex structure and their active binding site through computational approach are becoming progressively important [34-36]. Moreover, several approaches like geometric hashing method [37], Fast Fourier, correlation techniques [38] and HADDOCK program [39-40] used for the Protein-DNA interaction study.

The intermolecular docking study reported the Brassica Napus DREB1 protein has a GCC binding domain which bind to six nucleotides GCC box (A/GCCGAG) [41]. Similarly, Haddock used for the interactive study of the CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-fiprotein [42]. Furthermore, the similar interaction study of sub 1 gene protein with the GCC box promoter motif of UCC done through HADDOCK server in rice [24]. In Arabidopsis, HARDY (ATHRD) gene has Ap2/ERF domain had docked with the GCC box promoter motifs of several drought responsive genes [43]. Since the interaction of the regulatory protein and DNA involves the cis-element in the regulation of the various biological processes, hence these complex biological protein DNA structures need to be recognized. In this study, we validated and identify the GCC and TCC box promoter motifs by using Molecular Beacon Probe (MBP) [25], in the promoter sequences with the respective TF family gene. We carried out detailed in silico interaction analysis of the up-regulated MT gene (LOC_Os06g05910) expression increased 29 fold) also has GCC box likewise down-regulated RG gene (LOC_Os12g05900) expression decreased -2 fold) expressed genes. The GCC and TCC box was found in the promoter of UR-DEGs and DR-DEGs respectively [23-24] after analysis through MEME (+e.5.0) (http://meme. ncbi. net/meme/cgi-bin/meme. cgi). It was observed that promoter sequence of up-regulated MT gene (LOC_Os06g05910) expression increased 15 fold) also has GCC box likewise down-regulated RG gene (LOC_Os12g05900) (expression decreased -2 fold) has TCC box, which were further used to retrieve their promoter sequence from eukaryotic promoter database (EPD) (http://www. epd. isb-sib. ch/seq_download. html) Consensus promoter motifs GCC and TCC box found in the MT (UR-DEG) and RG (DR-DEG) respectively. For the validation of GCC and TCC box in the promoter of the MT and RG respectively, the specific MBP and primers [25] were designed by using Beacon Designer 7 (BD7, PREMIER Biosoft, USA) as reported in our previous study [24].

For designing of specific MBP and primer, the promoter sequences of the MT and RG of 600 nt length (-499 to +100) were retrieved from the EPD (http://www. epd. isb-sib. ch/seq_download. html) and used to design the specific MBP with stem sequences at 5' and 3' end (highlighted/underlined) (Table-1) and primers (Table-2) by Beacon Designer7. Further validation analysis was carried out by using MBP based Real Time PCR (Applied Biosystems 7500 Fast-Real-Time PCR Systems, USA).

**Table 1: Molecular beacon probe sequence of MT (UR-DEG) and RG (DR-DEG)**

<table>
<thead>
<tr>
<th>DEGs</th>
<th>Length of MBP (nt)</th>
<th>MBP with stem sequence (underlined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT (Os06g05910)</td>
<td>24</td>
<td>5‘-GCCGATCGCGGCCTGCGGATCGCG-3’</td>
</tr>
<tr>
<td>RG (Os12g05900)</td>
<td>29</td>
<td>5‘-GCCGATCTTCCTCCGTCCTGCGGATCGCG-3’</td>
</tr>
</tbody>
</table>

**Table 2: Primer sequences of UR-DEG (MT) and DR-DEG (RG)**

<table>
<thead>
<tr>
<th>DEGs</th>
<th>Left primer (5’-3’)</th>
<th>Right primer (5’-3’)</th>
<th>Amplicon size</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT (Os06g05910)</td>
<td>AGTGCATTTGGTGCTGCTCA</td>
<td>TCGGGGGATGAGCTTCACC</td>
<td>107</td>
</tr>
<tr>
<td>RG (Os12g05900)</td>
<td>CATGATTAGGGAGGATT</td>
<td>CGGAGGGTCCTAAATAAG</td>
<td>162</td>
</tr>
</tbody>
</table>

**Isolation of genomic DNA from rice plant**

Rice seeds of Azurca (Japonica sp.) were grown at room temperature (Fig. 1A) after surface sterilization (0.1% HgCl2) and dark incubation (48 at 36°C). The genomic DNA was isolated from rice seeds using CTAB (2X) method and subjected to RNase treatment (Fig. 1B). The quantity and quality checked in Biophotometer (Eppendorf, USA) followed by 0.8% agarose gel electrophoresis.

**Fig. 1:** (A) Fourteen days old rice seedling (B) RNase treated genomic DNA isolated from rice leaves

**Validation of consensus promoter motifs using a molecular beacon probe based on real time PCR**

For validation of the promoter motifs of the DEGs, specific MBP (table-1) and primers (table-2) used. Isolated genomics DNA was used as a template, whereas the GCC and TCC box containing MBP used as a probe for detection of consensus promoter motifs in DEGs. In PCR reaction volume of 15-20 µl (1X Taq buffer, 1 unit Taq polymerase, 0.2 mm dNTPs, 3-4 mM MgCl₂, 0.45µM primer, 3-10ng g DNA and 0.3-0.8µg MBP) at optimized PCR condition (95°C for 4-10 min; 40 cycles of 10-15s at 95°C, 20-35s at 60°C, and 30-45s at 72°C). PCR amplification was carried out in Real Time PCR System (Applied Biosystems 7500 Fast-Real-Time PCR Systems, USA). For the detection of GCC box in the promoter of MT (LOC_Os06g05910) gene probe (MBP) of GCC box, primers and target genomic DNA as well as with non-template control (NTC) and negative control having MB of TCC box were used in PCR amplification. Likewise, for TCC box promoter motif detection in RG (LOC_Os12g05900) gene promoter, TCC box MBP was amplified along with non-template control (NTC) and negative control having MB of GCC box using primers and target genomic DNA. The Ct value obtained from the Real Time PCR data.

**In silico protein-DNA interaction studies**

For the protein-DNA interaction study the UR-DEG AP2/EREBP (LOC_Os03g22170) TF (expression increased 29 fold), its protein

**MATERIALS AND METHODS**

**Identification of consensus promoter motif in DEGs and designing of specific MBP and primers**

Anoxic rice coleoptiles microarray result [8] used for the identification of DEGs. In our previous work, we identified the up-regulated differently (expression increased ≥2 fold) expressed genes (UR-DEGs) and down regulated differently (expression decreased ≥-2 fold) expressed genes. The GCC and TCC box was found in the promoter of UR-DEGs and DR-DEGs respectively [23-24] after analysis through MEME (+e.5.0) (http://meme. ncbi. net/meme/cgi-bin/meme. cgi). It was observed that promoter sequence of up-regulated MT gene (LOC_Os06g05910) expression increased 15 fold) also has GCC box likewise down-regulated RG gene (LOC_Os12g05900) (expression decreased -2 fold) has TCC box, which were further used to retrieve their promoter sequence from eukaryotic promoter database (EPD) (http://www. epd. isb-sib. ch/seq_download. html) Consensus promoter motifs GCC and TCC box found in the MT (UR-DEG) and RG (DR-DEG) respectively. For the validation of GCC and TCC box in the promoter of the MT and RG respectively, the specific MBP and primers [25] were designed by using Beacon Designer 7 (BD7, PREMIER Biosoft, USA) as reported in our previous study [24].

For designing of specific MBP and primer, the promoter sequences of the MT and RG of 600 nt length (-499 to +100) were retrieved from the EPD (http://www. epd. isb-sib. ch/seq_download. html) and used to design the specific MBP with stem sequences at 5’ and 3’ end (highlighted/underlined) (Table-1) and primers (Table-2) by Beacon Designer7. Further validation analysis was carried out by using MBP based Real Time PCR (Applied Biosystems 7500 Fast-Real-Time PCR Systems, USA).
sequences retrieved from the TIGR (http://rice.plantbiology.msu.edu/cgi-bin/ORF_infoPage.cgi). Further, its structure was predicted by using I-TASSER (http://zhanglab.ccmb.med.umich.edu/ITASSER/) which built the 3D models on multiple-threading alignments. The accuracy of protein prediction is based on the confidence score (C-score) of the model by I-TASSER [45]. The best protein model generated by I-TASSER was run on Ramachandran Plot Analysis (RAMPAGE) (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [46] for evaluation of the stability of protein models. Moreover, for the construction of a 3D DNA structure of promoter motif DNA, 25 nt sequence of the promoter motif having GCC box and the TCC box required [24, 42]. 3D-DART (DNARic:Driven DNA Analysis and Rebuilding tool) server (http://haddock.science.uu.nl/services/HADDOCK/) was used for generating 3D custom made structural model of the validated promoter motif for both MT and RG DEGs.

The 3D model of DNA having varies bend angle ranging from the 0-40° for each constructed 3D DNA model. Hence five 3D Model generated for each MT and RG DEGs. Consequences for in silico interaction studies, both 3D DNA models of gene promoter motifs and AP2/EREBP TF model were run on the online HADDOCK web server (http://haddock.science.uu.nl/services/HADDOCK/haddock.php). Further for validation of the interacting protein-DNA complex molecular structure predicted by HADDOCK server, analyzed in SiteMap module of the Schrödinger Suite [36].

RESULTS AND DISCUSSION

Rice (Oryza sativa L.) represents the third most important food grain crop in the world behind wheat and corn [47]. However the various abiotic stresses reduce the crop productivity [48-49]. Abiotic stresses control the expression of the many genes and their product as well as TFs for their regulation [50]. In plants low oxygen stress stimulates the composite metabolic pathways and genetic programs, including the differential expression of a great number of genes [7]. The gene expression studies revealed the up-regulation of genes coding for transcription factors under low oxygen stresses [51]. Microarray analysis has been used to study differential expression of various genes in abiotic and biotic stresses. Moreover, in silico study of differentially expressed genes reveals many possible functions of the genes during the different stresses, hence the validation required for the obtained results. The gene promoter contains cis-elements which play a central role in genes regulation contains the essential nucleotide sequences and transcription start site.

In silico study about promoter motifs of the differentially expressed genes in anoxia have been reported [9,22,23]. However the anoxia responsive DEGs have the consensus promoter motifs (GCC and TCC box) in their promoter, reported using MEME analysis, in our previous study [23]. Further identification and validation study on a GCC box (GCCGCC) in the anoxia responsive differentially expressed UCE gene have been done [24]. We also identified the occurrence of the GCC box (GCCGC) in MT and the TCC box (CTCCCTCCTCCTC) in RG gene in their promoter motif (fig. 2).

Fig. 2: A snapshot repeated set of GCC-box (GCCGCCGCCGC) and the TCC box (CTCCCTCCTCCTCCTC) positioned in the promoter region of MT (EP06921) and RG (EP02492) DEG, respectively, which ranging from 200 to 600 bp of the promoter region and analyzed by MEME (v.4.5.0)

To identify the promoter motifs in a set of sequences web-accessible bioinformatics tools are being used routinely by molecular biologists, such as MEME [52]. Similar studies on the identification of promoter motif reported in rice [23,24,53] and in Zea Mays [54] using MEME. Several reports described methyltransferases involve in gene expression, genome stability and the DNA methylation in plants like in Arabidopsis [55], maize [56], rice [57] and in wheat [58]. It involves in embryonic development (GO: 0009790), metabolic process (GO: 0008152) and transferase activity (GO: 0016740) in plants. While RG has imperative secondary functions in plants like catabolic (GO: 0006139), metabolic process (GO: 0006139), signal transduction (GO: 0007165) and enzyme regulator activity (GO: 0030234). Moreover the RhoGAP is peripheral membrane proteins which control over the cell surface-associated actin cytoskeleton, contributing to the formation of social systems as diverse as lamellipodia and filopodia of animal cells, yeast buds, and plant root hairs and/or pollen tubes [59,60,61]. So, the presence and experimental detection of GCC and TCC box need to be validated in MT (UR-DEG) and RG (DR-DEG) becomes essential. The validation of the promoter sequences was achieved through MBP (table-1) and their specific primers (table-2) based on Real Time PCR. The genomic DNA extracted (fig. 1B) from the Rice seedlings of Azucena (japonica sp.) (fig. 1A) using CTAB (2X) method. DNA concentration and quality checked in Biophotometer (Eppendorf, USA) and 0.8% agarose gel electrophoresis, respectively.

Fig. 3: Molecular beacon probe based real time PCR result for MT having a GCC box in its promoter region (-499 to +100). Graph indicating the relation between Delta Rn vs Cycle number (A) and Rn vs Cycle number (B). Curves a’&b’ depict amplification of the gene with GCC probe. Curves c’ & d’ are the amplification of TCC probe and e’&f’ are amplifications of NTC (Non template control)
Further genomic DNA was used for the validation of Promoter motifs in DEG. The presence of GCC and TCC box promoter motif in the promoter region of the of MT and RG DEG validated through the designed MBP respectively, based on Real Time PCR assay. The graph (DRn vs cycle number (Fig. 3A) and Rn vs Cycle number (Fig. 3B) generated by the inbuilt Real Time PCR software. The gene amplification of MT observed through the GCC box containing MBP (Fig. 3) have avg. Ct values 30.62 (table-3). The GCC box containing MBP probe was detected during PCR amplification, however amplification of the NTC (Non template control) and the negative control TCC box containing MBP was undetected, confirming the GCC box presence in MT gene, after analysis in the Real Time PCR assay.

The presence of the TCC box detected in the promoter region after PCR amplification reaction of the RG, have 30.13 avg Ct value (table-3). The graph DRn vs cycle number (Fig. 4A) and Rn vs Cycle number (fig. 4B) generated by the inbuilt software. The presence of the negative control GCC box containing MBP and NTC were undetected in a real time PCR assay, confirming the TCC box presence in RG gene. Similarly the GCC box validated in anoxia responsive differentially expressed UCC gene with MBP [24]. Moreover the sensitivity and accuracy of MBP have been reported earlier [29, 62]

Transcription factors (TFs) are the key regulator which controls the expression of clusters of genes through the specific binding site present in the genes promoter’s site of the respective target genes [63]. Under biotic and abiotic responses AP2-EREBP/ERFs TF has been found to be involved in growth, development and metabolic regulation [12]. AP2/ERF superfamily proteins act as a transcriptional regulator plays a essential role in gene expression in response to the hormone, biotic and abiotic factors, symbiotic interactions, cell specialization, and stress signalling pathways in plants [64-65]. The rice ERF transcription factor OsERF922 binds specifically to the GCC box sequence, and acts as a transcriptional activator in rice plant cells [66]. However, in plants AP2/ERF superfamily TF interact specifically with widely conserved AGCCGCC motifs (GCC box). However base pair mutation decreases the binding affinity of the ERF TF [17]. Gene expression controlled by the AP2/ERF superfamily TF negatively or positively with the interaction of the GCC box promoter motif. However reduces its expression when G residue in GCC box replaced by T residue [67]. Moreover, in plant the mutation of the core sequence in the promoter region of GCC box reduces the binding activity of TF reported by several researchers [68-69]. In consequence, protein sequence and structure for the interaction study of the TFs with the promoter motif is needed to identify for understanding the regulation mechanism of various biological process. Hence, the 3D structure of the AP2/EREBP TF (Loc_Os03g22170) which is not available in the PDB database, generated from the I-TASSER (http://zhanglab.csb.med.umich.edu/I-TASSER/) sever. I-TASSER predicted the five best models on the basis of the confidence score (C-score), the estimated TM-score and RMSD [45]. Best two predicted models (Model1 and Model 2) from the I-TASSER run in the RAMPAJE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) for evaluation of the stability of protein models. Protein model 2 of AP2/EREBP TF is a more stable structure having 68.5% residues is in the most favorable region and 22.8% in allowed region (fig. 5). Hence the model 2 of AP2/EREBP TF used for the further interactive study.

**Table 3: Ct value chart of MT and RG genes**

<table>
<thead>
<tr>
<th>DEGs</th>
<th>Replicate</th>
<th>Template</th>
<th>Molecular beacon</th>
<th>Ct value</th>
<th>Avg Ct value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT (Os01g05910) R1</td>
<td>Template</td>
<td>GCC box</td>
<td>30.9</td>
<td>30.62</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>Template</td>
<td>GCC box</td>
<td>30.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>Template</td>
<td>TCC box</td>
<td>30.82</td>
<td>30.13</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>Template</td>
<td>TCC box</td>
<td>29.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For in silico study of the protein-DNA interaction, the 3D model of protein as well as 3D model of DNA was required. The MT and RG gene promoter DNA of 25 nt length used for the generation of the 3D model (fig. 7) by the 3D DART server. Five 3D DNA models generated for each MT and RG gene promoter DNA having GCC and TCC box respectively.
Protein-DNA interaction proteins are central for many processes in living cells, especially transcriptional regulation and DNA modification. To understand the important biological process and working genomes, it is essential to understand the interaction at the macromolecular level [70-71]. Therefore, structural determination of the protein-DNA complexes and the factors that regulating interaction is essential [72]. In an interaction study of macromolecular complexes HADDOCK server was used [24,42,73]. The both DNA and protein, with their respective binding sites run into the HADDOCK server. The HADDOCK result of the interacting molecules between MT gene Promoter DNA (containing GCC box) and AP2/EREBP TF (table-4) generated. Similarly, HADDOCK result generated for the interacting molecules between RG gene Promoter DNA (containing TCC box) and AP2/EREBP TF (table-5). Prediction of the best interaction model based on HADDOCK score, which is based on the RMSD, van der wall energy, electrostatic buried surface area and Z-score. The lowest HADDOCK score showed the favorable interaction. The result showed the best interaction between the MT gene promoter DNA and AR2/EREBP TF in the model (IAPMTGMB2-BS1) has HADDOCK score -112.9 +/ - 9.0 (table-4), while the interaction model (IAPRGTBTFM2-BS1) of RG gene promoter DNA and AP2/EREBP TF has HADDOCK score -120.7 +/- 5.9 (table-5).

**Table 4: Protein-DNA docking Models of docked AP2/EREBP (LOC_Os03g22170) TF gene with DNA segment containing a GCC box of UR-DEG, MT (LOC_Os06g05910) by HADDOCK**

<table>
<thead>
<tr>
<th>Interaction</th>
<th>HADDOCK score</th>
<th>RMSD</th>
<th>Van der Waals energy</th>
<th>Electrostatic energy</th>
<th>Desolvation energy</th>
<th>Restraints violation energy</th>
<th>Buried Surface Area</th>
<th>Z- Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAPMTGMB2-BS1</td>
<td>-112.9 +/- 9.0</td>
<td>11.1 +/ - 0.2</td>
<td>-619.9 +/- 6.6</td>
<td>-438.1 +/- 22.6</td>
<td>25.6 +/- 9.4</td>
<td>110.1 +/- 27.36</td>
<td>21681.1 +/- 158.6</td>
<td>-2.4</td>
</tr>
<tr>
<td>IAPMTGATM2-BS1</td>
<td>86.1 +/- 14.6</td>
<td>20.1 +/- 0.5</td>
<td>-627.8 +/- 4.4</td>
<td>-273.7 +/- 43.3</td>
<td>5.5 +/- 3.9</td>
<td>1980.5 +/- 33.25</td>
<td>2110.6 +/- 148.6</td>
<td>-2.8</td>
</tr>
<tr>
<td>IAPMTGATB2-BS1</td>
<td>94.9 +/- 6.9</td>
<td>16.0 +/- 0.4</td>
<td>-582.7 +/- 4.6</td>
<td>-351.4 +/- 44.5</td>
<td>21.8 +/- 6.5</td>
<td>2015.0 +/- 57.07</td>
<td>18345 +/- 149.1</td>
<td>-2.2</td>
</tr>
<tr>
<td>IAPMTGCTMB2-BS1</td>
<td>97.1 +/- 13.4</td>
<td>25.4 +/- 0.2</td>
<td>-547.2 +/- 8.0</td>
<td>-317.3 +/- 31.1</td>
<td>3.2 +/- 5.0</td>
<td>2120.8 +/- 27.28</td>
<td>15514 +/- 106.5</td>
<td>-2</td>
</tr>
<tr>
<td>IAPMTGDM2-BS1</td>
<td>119.2 +/- 16.0</td>
<td>15.5 +/- 0.6</td>
<td>-53.5 +/- 8.2</td>
<td>-149.6 +/- 59.4</td>
<td>1.2 +/- 11.0</td>
<td>20083.9 +/- 45.67</td>
<td>13731 +/- 195.1</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

Keys: I- Interaction; AP- AP2/EREBP (LOC_Os03g22170) TF; MT- methyltransferase domain containing protein gene (LOC_Os06g05910); G- GCC box (A/B/C/D) /T- 10-40º bend angle; M2- Protein model 2; BS1- Binding site.

**Table 5: Protein-DNA docking Models of docked AP2/EREBP (LOC_Os03g22170) TF gene with DNA segment containing a TCC box of the DR-DEG, RG (LOC_Os12g05900) by HADDOCK server**

<table>
<thead>
<tr>
<th>Interaction</th>
<th>HADDOCK score</th>
<th>RMSD</th>
<th>Van der Waals energy</th>
<th>Electrostatic energy</th>
<th>Desolvation energy</th>
<th>Restraints violation energy</th>
<th>Buried Surface Area</th>
<th>Z- Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAPRGTM2-BS1</td>
<td>-102.7 +/- 10.9</td>
<td>4.2 +/-</td>
<td>-68.0 +/- 10.6</td>
<td>-289.1 +/- 32.1</td>
<td>12.0 +/- 6.0</td>
<td>110.5 +/- 32.04</td>
<td>20664.4 +/- 225.2</td>
<td>-1.7</td>
</tr>
<tr>
<td>IAPRGTATM2-BS1</td>
<td>-108.3 +/- 9.1</td>
<td>16.6 +/-</td>
<td>-73.3 +/- 7.5</td>
<td>-305.3 +/- 12.0</td>
<td>13.6 +/- 3.2</td>
<td>160.0 +/- 37.04</td>
<td>2306.6 +/- 123.9</td>
<td>-1.6</td>
</tr>
<tr>
<td>IAPRGTB2TM2-BS1</td>
<td>-120.7 +/- 5.9</td>
<td>15.2 +/-</td>
<td>-68.4 +/- 2.5</td>
<td>-417.9 +/- 27.5</td>
<td>14.9 +/- 5.6</td>
<td>163.7 +/- 42.39</td>
<td>22328.2 +/- 44.4</td>
<td>-2</td>
</tr>
<tr>
<td>IAPRGCTM2-BS1</td>
<td>-104.8 +/- 12.4</td>
<td>0.4</td>
<td>-76.4 +/- 10.8</td>
<td>-272.9 +/- 29.5</td>
<td>15.3 +/- 6.5</td>
<td>108.0 +/- 15.74</td>
<td>2204.6 +/- 167.7</td>
<td>-1.6</td>
</tr>
<tr>
<td>IAPRGDTM2-BS1</td>
<td>-110.9 +/- 7.3</td>
<td>6.4 +/-</td>
<td>-73.9 +/- 4.0</td>
<td>-373.4 +/- 32.3</td>
<td>19.8 +/- 2.0</td>
<td>177.9 +/- 62.35</td>
<td>2231.4 +/- 68.0</td>
<td>-2.2</td>
</tr>
</tbody>
</table>

Keys: I- Interaction; AP- AP2/EREBP (LOC_Os03g22170) TF; RG- RhoGAP domain containing protein (LOC_Os12g05900); T- TCC box; A/B/C/D /T- 10-40º bend angle; M2- Protein model 2; BS1- Binding site. The structural visualization of these protein-DNA interaction models done by using Chimera 1.9 for the model IAPMTGMB2-BS1 (fig. 8) and model IAPRGTB2TM2-BS1 (fig. 9). In the complex protein-DNA model (IAPMTG2-B51), AP2/EREBP TF binds with the linear DNA segment (fig. 8A) whereas in the model (IAPRGTBTFM2-BS1) it binds with the 20º bend 3D DNA model (Fig. 9A).
Fig. 8: (A) Protein-DNA docking model (IAPMTGM2-BS1) of the AP2/EREBP TF (LOC_Os03g22170) and DNA segment containing GCC box promoter motif of the UR-DEG, MT (LOC_Os06g05910) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas nucleotide Adenine in red, Guanine represented in forest green, Cytosine in a golden and thymine in blue color at chain B. The structure was generated using Chimera 1.9. (B): Enlarged molecular view of rectangle, area was generated using Site Map module of Schrödinger Suite.

Fig. 9: (A) Protein-DNA docking model (IAPRGTBTM2-BS1) of the AP2-EREBP TF (LOC_Os03g22170) and DNA segment containing TCC box promoter motif of the DR-DEG, RG (LOC_Os12g05900) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas nucleotide Adenine in red, Guanine represented in forest green, Cytosine in a golden and thymine in blue color at chain B. The structure was generated using Chimera 1.9 (B): Enlarged molecular view of rectangle, area was generated using the Site Map module of Schrödinger Suite.

A further validation study of the HADDOCK generated complex protein-DNA model analyzed through SiteMap. The HADDOCK results for the model IAPMTGM2-BS1 (fig. 8A) and model IAPRGTBTM2-BS1 (fig. 9A) showing the binding interaction between the DNA and the protein. In a Site Map analysis of model IAPMTGM2-BS1 (fig. 8B) the binding residue SER213 and MET1 binds with the cytosine17, VAL212 to guanine15, GLN223 to guanine19 with Hydrogen bond. All binding residue is present in the protein active binding site. Similarly, in model IAPRGTBTM2-BS1 (fig. 9B). The binding residue SER224 bind with the thymine10 residue and GLN225 DNA binding with cytosine9 with H-bond. The residues present in the protein active site bind with the DNA sequence with H-bonding predicted by SiteMap for complex model IAPMTGM2-BS1 (fig. 8B) and IAPRGTBTM2-BS1 (fig. 9B). Similarly, the report on the active binding site prediction of the flexible loop PIRI02 kinase (as plausible novel anti-malarial drug target) which can interact with appropriate ligands was identified computationally by SiteMap module [74]. The SiteMap program [36] can successfully suggest possible binding sites in protein. [75, 76]

CONCLUSION
The present study is proposed to show the relation between the TF and promoter motifs of anoxia responsive DEGs. The identification and validation of the promoter motif sequences in MT (UR-DEG) and RG (DR-DEG) genes were done successfully through in silico study and MBP based Real Time PCR analysis, respectively.

Anoxia responsive AP2/EREBP TF [LOC_Os03g22170] has shown the good interaction between the MT and RG genes. However, the result revealed that the AP2/EREBP TF binding affinity towards the TCC box in the RG gene promoter is more as compared to the GCC box promoter of MT gene. Hence the present study reveals the validation of the in silico study of the promoter motifs of MT and RG genes by MBP is reliable. Moreover, their interaction study with transcription factor shows that it might regulate the differential expression of these genes under rice in anoxia. Further the validation of interacting molecules will help to understand the molecular level of organization and their regulation mechanism.

ACKNOWLEDGEMENT
This work was financially supported by University Grants Commission (UGC), GOI, and New Delhi for the major research project [F. No. 37-113/2009 (SR)]. Mr. Gopal Kumar Prajapati worked as Project Fellow and financially supported from the above mentioned UGC sponsored project. Use SiteMap module of Schrödinger Suite facilities provided by Department of Pharmaceutical Sciences and Technology, BIT Mesra. Mr. Surendra Singh is kindly acknowledged for his concept in using SiteMap module of Schrödinger for this study. Dr. Ashutosh Kumar is kindly acknowledged for his concept in using Molecular Beacon for this study. Central Rainfed Upland Rice Research Station (CRURRS), Hazaribagh-825301, Jharkhand, India is gratefully acknowledged for providing the Azucena rice seeds used in this experiment.
CONFLICT OF INTERESTS
Declared None

REFERENCES


129


56. Steward N, Kusano T, Sano H. Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic Acids Res 2000;28:3250–9.


