Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 7, Issue 3, 2015

Original Article

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, the 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 AGCCGCC (GCC box) mainly a response to pathogenesis and wounding [15-17]. A similar study was reported on tomato, Ethylene- Responsive Factor (ERF) transcription factor Pti4, 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/GCCGAC, 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 anaerobic 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 (TCCTCC) 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 binging (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/GCCGAC) [41]. Similarly, Haddock used for the interactive study of the CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-1protein [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 motif by using Molecular Beacon Probe (MBP) [25], in the *MT* gene (LOC_Os06g05910) and RG gene (LOC_Os12g05900) respectively, founded on our preliminary work [44] by the Real Time PCR. We carried out detailed *in silico* interaction study of validated promoter motifs with AP2/EREBP TF. We generated the 3D DNA model for the validated promoter motifs by 3D DART and protein model of AP2/EREBP TF Protein by using I-TASSER. A further interaction study was carried out through HADDOCK severs. Eventually, we examine the comparative interaction relation between the two promoter sequences with the respective TF family gene.

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 (v4.5.0) (http: //meme. nbcr. 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. isbsib. 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)

DEGs	Length of MBP (nt)	MBP with stem sequence (underlined)
MT (Os06g05910)	24	5'-CGCGATCGCCGCCGCCGGATCGCG-3'
RG (Os12g05900)	29	5'-CGCGATCCTCCTCCTCCTCCTCGATCGCG- 3'

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

DEGs	Left primer (5'-3')	Right primer (5'-3')	Amplicon size
MT (Os06g05910)	CCTCCTAGTTCGTCCGTCAA	TCGAGCCTGGACTTCACC	107
RG (Os12g05900)	CATCATTAGCGGAGGATT	CGGAGGTGGCTAAATAAC	162

Isolation of genomic DNA from rice plant

Rice seeds of Azucena (Japonica *sp.*) were grown at room temperature (Fig. 1A) after surface sterilization (0.1% $HgCl_2$) and dark incubation (48 at 36 °C). The genomic DNA was isolated from rice seedlings 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.

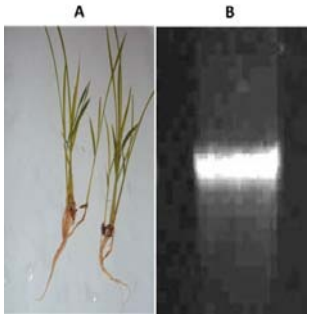

Fig. 1: (A) Fourteen days old rice seedling (B) RNase treated genomic DNA isolated from rice leaves

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 0.3-0.8µM MBP) at optimized PCR condition (95°C for 4-10 min; 40-45 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

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 (3DNA-Driven DNA Analysis and Rebuilding Tool) server (http: //haddock. science. uu. nl/dna/dna. php) 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 *UCC* gene have been done [24]. We also identified the occurrence of the GCC box (CGCCGCCGCCG) in *MT* and the TCC box (CTCCTCCTCCTCCTC) in *RG gene* in their promoter motif (fig. 2).

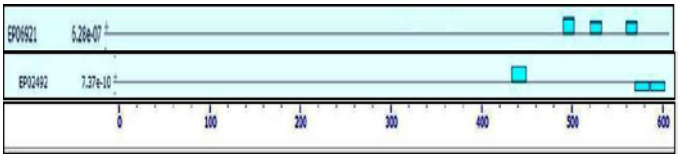

Fig. 2: A snapshot repeated set of GCC-box (CGCCGCCGCCG) and the TCC box (CTCCTCCTCCTCCTC) positioned in the promoter region of *MT* (EP06921) and *RG* (EP02492) DEG, respectively, which ranging from 200 to 600 bp of the promoter region and analyzed by MEME (v4.5.0)

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: 0009056), 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.

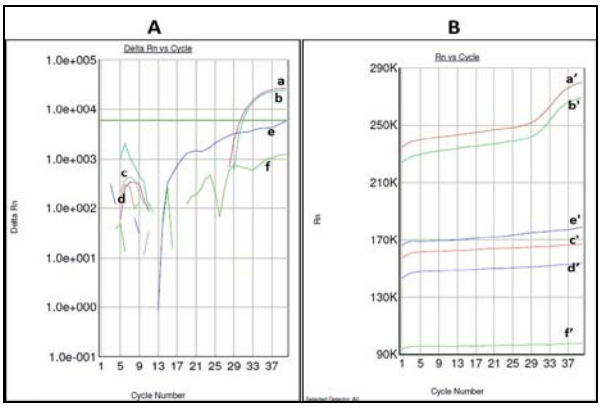

Fig. 3: Molecular beacon probe based real time PCR result for *MT* having a GCC box in its promoter region (-499 to +100). Graph indicating the relation between Delta Rn *vs* Cycle number (A) and Rn vs Cycle number (B). Curves *a'*&*b'* depict amplification of the gene with GCC probe. Curves *c'* & d *'* are the amplification of TCC probe and *e'*&*f'* are amplifications of NTC (Non template control)

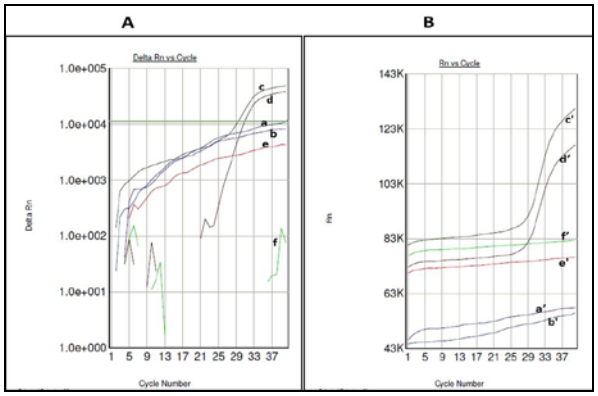

Fig. 4: Molecular beacon based real time PCR result for *RG* having a TCC box in its promoter region (-499 to +100). Graph indicating the relation between Delta Rn vs Cycle number (A) and Rn vs Cycle number (B). Curves *a′*&*b′* depict amplification of the gene with GCC probe. Curves *c′* & d′ are the 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 (Figur-3A) and RN v's 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 containg MBP probe was detected during PCR amplification, however amplification of the NTC (Non template control) and the negative control TCC box containg 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]

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

DEGs	Replicate	Template	Molecular beacon	Ct value	Avg Ct value
MT (Os06g05910)	R1	Template	GCC box	30.9	30.62
	R2	Template	GCC box	30.34	
RG (Os12g05900)	R1	Template	TCC box	30.82	30.13
	R2	Template	TCC box	29.43	

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. ccmb. med. umich. edu/I-TASSER/) severe. 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 RAMPAGE (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.

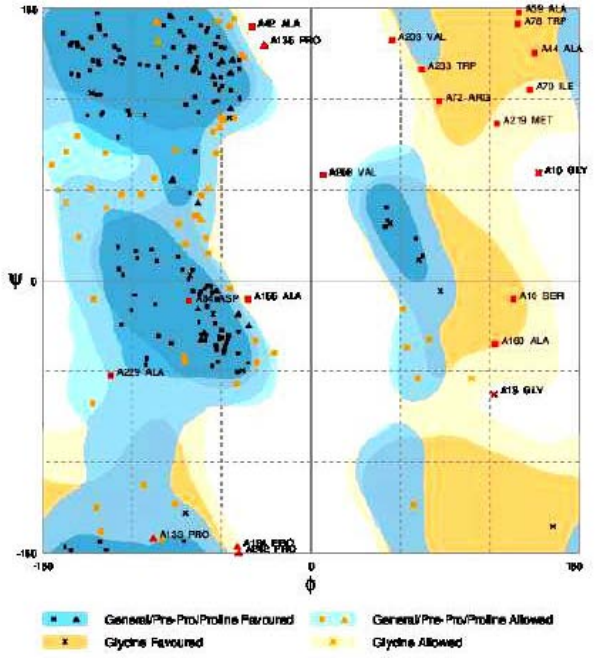

Fig. 5: Ramachandran plot assement of the AP2/EREBP (LOC_Os03g22170) TF (Protein model 2)

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.

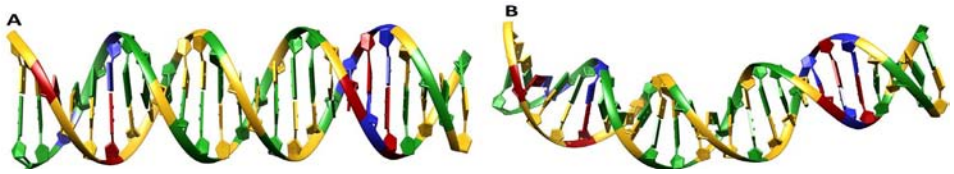

Fig. 7: 3D structure of the linear DNA segment of *MT* gene promoter motif. GCC-Box promoter motif positioned at 9- 18 nucleotide of 25 nt long sequences of the *MT* gene promoter DNA, generated by 3D-DART server. 3D DNA model structure varies between 0-40º of bending. Linear 3D model (Figure–7A) and bended 3D (40º) model (fig. 7B) represented for *MT* gene promoter DNA. In DNA model red color represents Adenine, green color represents guanine, gold color represents cytosine and the blue color represents thymine. The structure was generated using Chimera 1.9

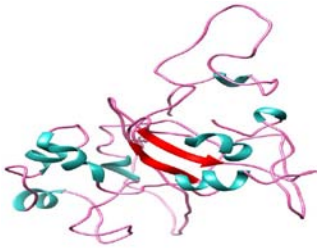

Fig. 6: 3D structure of AP2/EREBP (LOC_Os03g22170) TF (Model 2) predicted by I-TASSER. The coloring method is based on secondary structure. The red color represents strand, pink color is coiled and cyan color represents the helix on AP2/EREBP TF. The structure was generated using Chimera 1.9

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 site 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 HADDDOCK 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 (IAPMTGM2-BS1) has HADDOCK score -112.9 +/- 9.0 (table-4), while the interaction model (IAPRGTBTM2-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

Interaction	HADDOCK score	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-Score
IAPMTGM2-BS1	-112.9 +/- 9.0	11.1 +/- 0.2	-61.9 +/- 6.6	-438.1 +/- 22.6	25.6 +/- 9.4	110.1 +/- 27.36	2168.1 +/- 158.6	-2.4
IAPMTGATM2-BS1	86.1 +/- 14.6	19.1 +/- 0.5	-62.7 +/- 4.4	-273.7 +/- 43.3	5.5 +/- 3.9	1980.5 +/- 33.25	2110.6 +/- 148.6	-2.8
IAPMTGBTM2-BS1	94.9 +/- 6.9	26.0 +/- 0.4	-58.2 +/- 4.6	-351.4 +/- 44.9	21.8 +/- 6.5	2015.0 +/- 57.07	1834.5 +/- 149.1	-2.2
IAPMTGCTM2-BS1	97.1 +/- 13.4	25.4 +/- 0.2	-54.7 +/- 8.0	-317.3 +/- 31.1	3.2 +/- 5.0	2120.8 +/- 27.28	1551.4 +/- 106.5	-2
IAPMTGDTM2-BS1	119.2 +/- 16.0	15.5 +/- 0.6	-53.0 +/- 8.2	-149.6 +/- 59.4	1.2 +/- 11.0	2008.9 +/- 45.67	1373.1 +/- 195.1	-1.4

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

Interaction	HADDOCK score	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-Score
IAPRGTM2-BS1	-102.7 +/- 10.9	4.2 +/- 2.6	-68.0 +/- 10.6	-289.1 +/- 32.1	12.0 +/- 6.0	110.5 +/- 32.04	2066.4 +/- 225.2	-1.7
IAPRGTATM2-BS1	-104.8 +/- 9.1	16.6 +/- 0.1	-73.3 +/- 7.5	-305.3 +/- 12.0	13.6 +/- 3.2	160.0 +/- 37.04	2306.6 +/- 123.9	-1.6
IAPRGTBTM2-BS1	-120.7 +/- 5.9	15.2 +/- 0.7	-68.4 +/- 2.5	-417.9 +/- 27.5	14.9 +/- 5.6	163.7 +/- 42.39	2232.8 +/- 44.4	-2
IAPRGTCTM2-BS1	-104.8 +/- 12.4	20.6 +/- 0.4	-76.4 +/- 10.8	-272.9 +/- 29.5	15.3 +/- 6.5	108.0 +/- 15.74	2204.6 +/- 167.7	-1.6
IAPRGTDTM2-BS1	-110.9 +/- 7.3	6.4 +/- 0.4	-73.9 +/- 4.0	-373.4 +/- 32.3	19.8 +/- 2.0	177.9 +/- 62.35	2231.4 +/- 68.0	-2.2

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 model2; BS1- Binding site. The structural visualization of these protein-DNA interaction models done by using Chimera 1.9 for the model IAPMTGM2-BS1 (fig. 8) and model IAPRGTBTM2-BS1 (fig. 9). In the complex protein-DNA model (IAPMTGM2-BS1), AP2/EREBP TF binds with the liner DNA segment (fig. 8A) whereas in the model (IAPRGTBTM2-BS1) it binds with the 20° bend 3D DNA model (Fig. 9A).

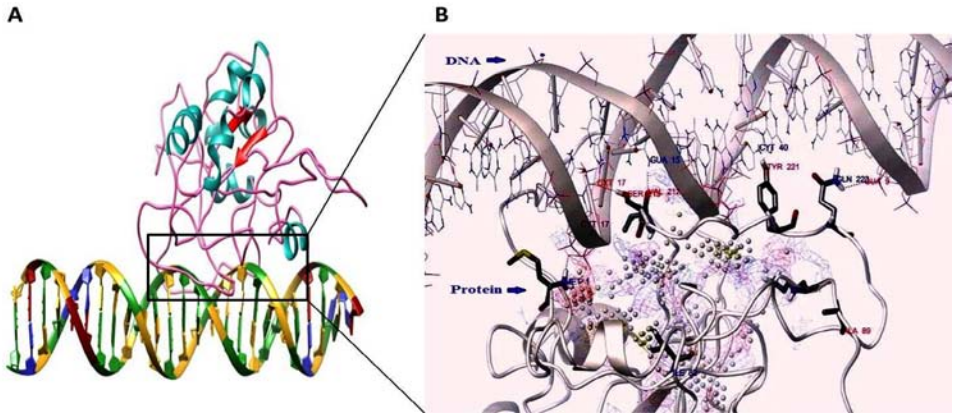

Fig. 8: (A) Protein-DNA docking model (IAPMTGM2-BS1) of the AP2/EREBP TF (LOC_Os03g22170) and DNA segment containing GCC box promoter motif of the UR-DEG, *MT* (LOC_Os06g05910) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas nucleotide 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

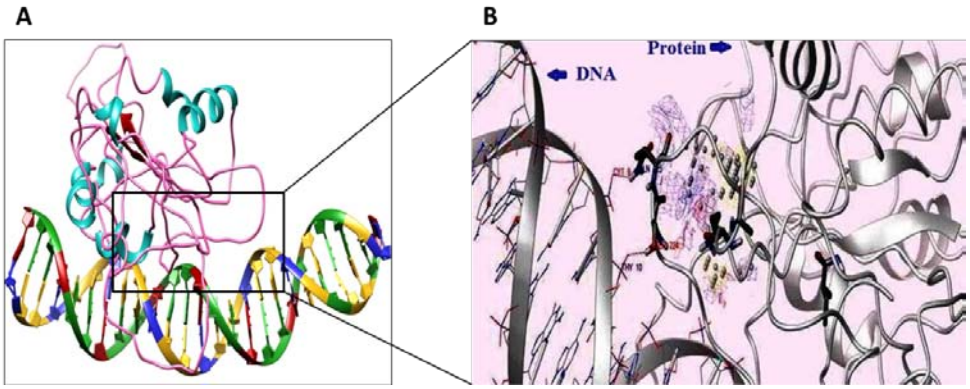

Fig. 9: (A) Protein-DNA docking model (IAPRGTBTM2-BS1) of the AP2-EREBP TF (LOC_Os03g22170) and DNA segment containing TCC box promoter motif of the DR-DEG, *RG* (LOC_Os12g05900) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas 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 PfRIO2 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 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

1. Liu F, Xu W, Wei Q, Zhang Z, Xing Z, Tan L, *et al*. Gene expression profiles deciphering rice phenotypic variation between *nipponbare* (Japonica) and 93-11 (Indica) during oxidative stress. PLoS ONE 2010;5(1):e8632.
2. Mizoi J, Yamaguchi-Shinozaki K. Molecular approaches to improve rice abiotic stress tolerance. Methods Mol Biol 2013;956:269-83.
3. Perata P, Voesenek LA. Submergence tolerance in rice requires Sub1A, an ethylene-response-factor-like gene. Trends Plant Sci 2007;12:43–6.
4. Magneschi L, Perata P. Rice germination and seedling growth in the absence of oxygen. Ann Bot 2009;103:181.
5. Hattori Y, Nagai K, Ashikari M. Rice growth adapting to deepwater. Curr Opin Plant Biol 2011;14:100-5.
6. Lakshmanan M, Mohanty B, LimS-H, Ha S-H, Lee D-Y. Metabolic and transcriptional regulatory mechanisms underlying the anoxic adaptation of rice coleoptile. AoB Plants 2014;6:1-13.
7. Vartapetian BB, Jackson MB. Plant adaptations to anaerobic stress. Ann Bot 1997;79:3.
8. Lasanthi-Kudahettige R, Magneschi L, Loreti E, Gonzali S, Licausi F, Novi G, *et al*. Transcript profiling of the anoxic rice coleoptiles. Plant Physiol 2007;144:218-31.
9. Pandey DM, Deb R, Kumar A, Japan T. Study on the functional characterization of low oxygen stress inducible genes in rice((Oryza sativa L) Proceeding of 5th. Intewrnational symposium on rice functional genomics (ISRFG' Epochal; 2007. p. 110.
10. Paul AL, Schuerger AC, Popp MP, Richards JT, Manak MS, Ferl RJ. Hypobaric biology: arabidopsis gene expression at low atmospheric pressure. Plant Physiol 2004;134:215.
11. Garg R, Verma M, Agrawal S, Shankar R, Majee M, Jain M. Deep transcriptome sequencing of wild halophyte rice, Porteresia coarctata, Provides novel insights into the salinity and submergence tolerance factors. DNA Res 2013;1-16.
12. Hussain SS, Kayani MA, Amjad M. Transcription factors as tools to engineer enhanced drought tolerance in plants. Biotechnol Prog 2011;27(2):297-306.
13. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. Biol Chem 1998;379:633-46.
14. Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, Sasaya T, *et el*. Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. Plant Cell Physiol 2011;52(2):344-60.
15. Buttner M, Singh B. *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. Proc Natl Acad Sci USA 1997;94:5961–6.
16. Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell 1995;7:173–82.
17. Hao DY, Ohme-Takagi M, Sarai A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem 1998;273(41):26857-61.
18. Chakravarthy S, Tuori RP, D Ascenzo MD, Fobert PR, Després C, Martin GB. The tomato transcription factor Pti4 regulates defense-related gene expression via GCC box and non-GCC box cis elements. Plant Cell 2003;15:3033.
19. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. Trends Plant Sci 2005;10:88-94.
20. Stockinger EJ, Gilmour SJ, Thomashow MF. *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcription activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA 1997;94:1035–40.
21. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, *et al*. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nat 2006;442:705-8.
22. Mohanty B, Krishnan SPT, Swarup S, Bajic VB. Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann Bot 2005;96:669.
23. Kumar A, Sharma V, Smita S, Sahu N, Shankaracharya, Vidyarthi AS, *et al*. In Silico analysis of motifs in promoters of differentially expressed genes in rice (Oryza sativa L.) under anoxia. Int J Bioinfo Res Appl 2009;5(5):523-47.
24. Prajapati GK, Kashayp N, Kumar A, Pandey DM. Identification of GCC box in the promoter region of ubiquinol cytochrome c chaperone gene using molecular beacon probe and its *in silico* protein-DNA interaction study in rice (*Oryza sativa* L.). Int J Comp Bioinfo *In Silico* Mod 2013;2(5):213-22.
25. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 1996;14:303-8.
26. Andersen CB, Holst-Jensen A, Berdal KG, Thorstensen T, Tengs T. Equal Performance of taq man, MGB, Molecular beacon, and SYBR green-based detection assays in detection and quantification of roundup ready soybean. J Agric Food Chem 2006;54:9658-63.
27. Kamboj A, Pateriya AK, Mishra A, Ranaware P, Kulkarni DD, Raut AA. Novel molecular beacon probe-based Real-Time RT-PCR assay for diagnosis of crimean-congo hemorrhagic fever encountered in India. Bio Med Res Int 2014;27:496219.
28. Gohring J, Jacak J, Barta A. Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. Plant Cell 2014;26:754–64.
29. Lata P, Ram S, Agrawal M, Shanker R. Real Time PCR for the rapid detection of vanA gene in surface waters and aquatic macrophyte by molecular beacon probe. Environ Sci Technol 2009;43:3343-8.
30. Qin S, Zhou HX. Structural models of Protein-DNA complexes based on interface prediction and docking. Curr Protein Peptide Sci 2011;12:531-9.
31. Leung KS, Wong K-C, Chan TM, Wong MH, Lee KH, Lau CH, *et al*. Discovering protein–DNA binding sequence patterns using association rule mining. Nucl Acids Res 2010;38(19):6324-37.
32. Nagarajan R, Ahmad S, Gromiha MM. Novel approach for selecting the best predictor for identifying the binding sites in DNA binding proteins. Nucleic Acids Res 2013;41(16):7606–14.
33. Tomovic A, Oakeley EJ. Position dependencies in transcription factor binding sites. Bioinfo 2007;23:933-41.
34. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis-an improved methodological approach. J Comput Chem 2007;28(3):644–54.
35. Setny P, Bahadur RP, Zacharias M. Protein-DNA docking with a coarse-grained force field. BMC Bioinfo 2012;13:228.
36. Site Map-V-2.5, V 2.5, Schrödinger, LLC: New York, NY; 2011.
37. Banitt I, Wolfson HJ. Para dock: a flexible non-specific DNA–rigid protein docking algorithm. Nucleic Acids Res 2011;39(20):135.
38. Aloy P, Moont G, Gabb HA. Modelling repressor proteins docking to DNA. Proteins 1998;33(4):535–49.
39. Van Dijk M, van Dijk ADJ, Hsu V. Information driven protein-DNA docking using, HADDOCK: it is a matterof flexibility. Nucleic Acids Res 2006;34(11):3317–25.
40. van Dijk M, Bonvin AMJJ. Pushing the limits of what is achievable in protein-DNA docking: bench marking HADDOCK's performance. Nucleic Acids Res 2010;38(17):5634–47.
41. Qamarunnisa S, Hussain M, Jabbeen, Raza S, Khanani MH, Azhar A, *et al*. *In silico* studies on structure-fuction of DNA GCC-Box binding domain of *Brassica Nnapus* DREB1 protein. Pak J Bot 2012;44(2):493-500.
42. Pandey DM, Kumar A. 3D structure prediction and protein-DNA interaction of CCCH-type zinc finger transcription factor gene in Rice (*Oryza sativa* L.). Int J Comp Bioinfo In Silico Modeling 2013;2(2):94-103.
43. Shanker AK, Maddaala A, Kumar MA, Yadav SK, Maheswari M, Venkateswarlu B. In silico targeted genome mining and comparative modelling reveals a putative protein similar to an Arabidopsis drought tolerance DNA binding transcription factor in Chromosome 6 of Sorghum bicolor genome. Interdiscip Sci 2012;4(2):133-41.

44. Prajapati GK, Kumar A, Pandey DM. Poster presentation on molecular beacon probe based promoter motif detection & in silico protein-DNA interaction studies during submergence in rice (Oryza sativa L. In 7th International rice genetics symposium manila philippines; 2013. p. 85.
45. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf 2008;9:40.
46. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, *et al*. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 2003;50(3):437-50.
47. Binod P, Sindhu R, Singhania RR, Vikram S, Devi L, Nagalakshmi S, *et al*. Bioethanol production from rice straw: an overview. Bioresour Technol 2010;101:4767–74.
48. Rehman S, Harris PJC, Ashraf M. Stress environments and their impact on crop production. In: Ashraf M, Harris PJC, eds. Abiotic stresses: plant resistance through breeding and molecular approaches. Haworth Press: New York; 2005. p. 3-18.
49. Ashraf M, Athar HR, Harris PJC, Kwon TR. Some prospective strategies for improving crop salt tolerance. Adv Agron 2008;97:45–110.
50. Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. Rice 2012;5:6.
51. Liu FL, Van Toai T, Moy LP, Bock G, Linford LD, Quackenbush J. Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant Physiol 2005;137:1115–29.
52. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, *et al*. MEME SUITE: tools for motif discovery and searching. Nucl Acids Res 2009;37:202-8.
53. Doi K, Hosaka A, Nagata T. Development of a novel data mining tool to find cis-elements in rice gene promoter regions. BMC Plant Biol 2008;8:20.
54. Liseron-Monfils C, Lewis T, Ashlock D, McNicholas PD, Fauteux F, Strömvik M, *et al*. Promzea: a pipeline for discovery of co-regulatory motifs in maize and other plant species and its application to the anthocyanin and phlobaphene biosynthetic pathways and the maize development atlas. BMC Plant Biol 2013;13:42.
55. Vongs A, Kakutani T, Martienssen RA, Richards EJ. *Arabidopsis thaliana* DNA methylation mutants. Sci 1993;260:1926–8.
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.
57. Teerawanichpan P, Chandrasekharan M, Jiang Y, Narangajavana J, Hall T. Characterization of two rice DNA methyltransferase genes and RNA imediated reactivation of a silenced transgene in rice callus. Planta 2004;218:337–49.
58. Thomas M, Pingault L, Poulet A, Duarte J, Throude M, Faure S, *et al*. Evolutionary history of Methyltransferase 1 genes in hexaploid wheat. BMC Genomics 2014;15:922.
59. Rivero F, Cvrcková F. Origins and evolution of the actin cytoskeleton. Adv Exp Med Biol 2007;607:97-110.
60. Hall A, Rho GT. Pases and the control of cell behaviour. Biochem Soc Trans 2005;33:891-5.
61. Ridley AJ. Rho GT. Pases and actin dynamics in membrane protrusions and vesicle trafficking. Trends Cell Biol 2006;16:522-9.
62. Ye Q, Zhuang H, Zhou C. Detection of naphthalene by real-time immuno-PCR using molecular beacon. Mol Cell Probes 2009;23:29.
63. Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses, including drought, cold, and heat. Front Plant Sci 2009;5:170.
64. Song X, Li Y, Hou X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. pekinensis). BMC Genomics 2013;14:573.
65. Rashid M, Guangyuan H, Guangxiao Y, Hussain J, Xu Y. AP2/ERF transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. Evol Bioinf Online 2012;8:321-55.
66. Liu D, Chen X, Liu J, Ye J, Guo Z. The rice ERF transcription factor OsERF922 negatively regulates resistance to Magnaporthe oryzae and salt tolerance. J Exp Bot 2012;63(10):3899-911.
67. Fujimoto SY, Ohta M, Usui A. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell 2000;12:394-404.
68. Cheong YH, Moon BC, Kim JK. BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. Plant Physiol 2003;132:1961-72.
69. Sessa G, Meller Y, Fluhr R. A GCC element and a G-box motif participate in ethylene-induced expression of the PRB-lb gene. Plant Mol Biol 1995;28:145-53.
70. Donald JE, Chen WW, Shakhnovich EI. Energetics of protein-DNA interactions. Nucleic Acids Res 2007;35(4):1039-47.
71. Kauffman C, Karypic G. Computational tools for protein–DNA interactions. WIREs Data Mining Knowl Discov 2012;2:14–28.
72. Chou CC, Rajasekaran M, Chen C. An effective approach for generating a three Cys2His2zinc-finger-DNA complex model by docking. BMC Bioinf 2010;11:334.
73. deVries SJ, van Dijk M, Bonvin AMJJ. The HADDOCKweb server for data-driven Biomolecular docking. Nat Protoc 2010;5:883-97.
74. Parveen A, Chakraborty A, Konreddy AK, Chakravarty HC, Sharon A, Trivedi V, *et al*. Skeletal hybridization and PfRIO-2 kinase modeling for synthesis of a-pyrone analogs as anti-malarial agent. Eur J Med Chem 2013;70:607-12.
75. Nayal M, Honig B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. Proteins 2006;63:892-906.
76. Halgren TA. Identifying and characterizing binding sites and assessing drug ability. J Chem Inf Model 2009;49(2):377-89.