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.)

Gopal Kumar Prajapati¹, Neha Kashyap¹**, Ashutosh Kumar¹ and Dev Mani Pandey¹*

¹ Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India
**Present Address: Department of Biotechnology, Ranchi University, Ranchi-834001, Jharkhand, India

*Corresponding author: Dev Mani Pandey; e-mail: dmpandey@bitmesra.ac.in

ABSTRACT
Rice is the primary staple food for more than half of the world's population but very sensitive to various abiotic stresses (submergence) causing crop loss. During abiotic stresses various genes are differentially expressed to cope up with the stress conditions. The identification of Sub1A gene was a major breakthrough for the submergence tolerance which often regulates other genes by binding to their consensus promoter motifs such as GCC box. It was observed that Ubiquinol Cytochrome C chaperone (UCCC) gene was among many up-regulated differentially co-expressed genes having GCC box as a conserved motif. The primary role of UCCC gene is oxidative respiration but also has imperative secondary functions in plants. Therefore, UCCC gene was selected for the identification of GCC Box in the promoter region using Molecular Beacon Probe based Real Time PCR and their interaction with Sub1A protein. Real Time PCR analysis confirmed the presence of GCC box. Subsequently, the interaction of Sub1A with GCC box was studied through HADDOCK server. Protein-DNA interaction thus, suggested significant binding affinity of Sub1A towards GCC box present in the promoter region of UCCC gene.

Keywords: 3D-DART, Differentially expressed genes, HADDOCK, I-TASSER Oryza sativa, Sub1A, Ubiquinol Cytochrome C Chaperone gene

INTRODUCTION
Rice is a most important crop for human consumption with production in over 150 million hectares yielding almost 600 million megagrams annually [1]. It is a very important staple food that feeds more than half the world's population [2]. Rice thrives in water-logged conditions and can tolerate submergence at levels that would kill other crops [3] but is highly sensitive to salinity among cereals [4], and is susceptible to drought and cold [3]. Rice is a semi-aquatic plant and well adapted to survive partial flooding conditions. However, it is damaged when submerged for a relatively longer period of time [5-7].

A plant when in submerged condition inhibits aerobic respiration and photosynthesis, and stimulates a variety of responses that can enhance survival, acts as a switch from aerobic to anaerobic respiration [8]. The anaerobic stress stimulates the composite metabolic pathways by the differential expression of a large number of genes [9] including genes coding for transcription factors [10] and signal transduction components [11]. Studies on differential expression of genes have also been explained by microarray experiments in anoxic rice coleoptiles [12]. The regulation of gene expression in response to oxygen deprivation has been described in Oryza sativa [12-16].
Further, it was discovered that three ethylene response factors (ERFs) were identified within the Sub1A locus in tolerant rice varieties (e.g. FR13A) determining it as the major determinant of tolerance. Therefore, identification of the Sub1 locus and elucidation of its role in the adaptation of rice to submergence is a breakthrough in plant adaptation to anaerobiosis [6]. Xu et al. [16] reported that FR13A Sub1 region encodes three transcription factors (Sub1A, Sub1B and Sub1C) belonging to the B-2 subgroup of the ethylene response factors (ERFs) \( / \) ethylene-responsive element binding proteins (EREBPs)/apetala 2-like proteins (AP2). Study on molecular marker survey and expression analyses of Sub1A in rice has also been reported [17]. In Arabidopsis, gene expression study revealed the modulation of gene expression occurred positively or negatively by interaction of ERF-TF with GCC Box [18]. Chakravarty et al. [19] described 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. Kumar et al. [20, 21] reported that consensus promoter motif GCC box (GGCCGC) was highly conserved in the promoter of up-regulated differentially expressed Genes (DEGs) in rice under anoxia. These transcriptional regulations of DEGs were facilitated by the interaction of TFs with promoter motifs/cis-regulatory elements which provided an insight to the vast molecular mechanisms of co-expressed/regulated genes during stress.

Cytochrome C is a small peripheral, nuclear encoded membrane protein located in inter-membrane space of mitochondria. It functions in the catalytic transfer of electrons between respiratory complexes III and IV [22]. The functional importance and unique intragranellar position of cytochrome c molecule has been investigated in animals and is reported to contain cAMP response element (CRE) and nuclear respiratory factor (NRF)-binding sites [23]. CRE is a cytochrome c gene promoter involved in cAMP-dependent expression and NRF is involved in the coordinating activities of nuclear and mitochondrial genes [24]. However, little is known about the cytochrome c gene in plants with the exception that it has been cloned and sequenced in Arabidopsis and rice. The regulation of rice cytochrome c gene OsCc1 (accession no. M63704) and its promoter activities in transgenic rice have been examined [25]. They demonstrated that OsCc1 is expressed in most of the tissues, and its expression is particularly high in the non-photosynthetic parts of plants such as roots, calli, and suspension cells. A fusion gene was constructed for evaluating the OsCc1 promoter with the gfp gene and introduced into rice. The activity of this gene in various tissues and cell types of transgenic plants was analyzed in comparison with other promoters by conventional dot-blot hybridization techniques [25]. Chaperone genes are encoded proteins in both prokaryotes and eukaryotes that bind to nascent or unfolded polypeptides and ensure correct folding or transport [26]. It is well known that plant heat- shock proteins (HSPs) and other chaperones play major roles both in response to adverse environmental conditions and in various developmental processes [27], rather they are also responsible for protein folding, assembly, translocation & degradation [28].

DNA–protein interactions are pivotal for many biological activities. These interactions are fundamental for gene expression and DNA modifications and their function in regulating [29] and determining the structure of protein-DNA complexes have been understood many biological processes [30]. Although, laboratory methods for protein-DNA interaction studies are very expensive and time-consuming, therefore by doing computational analysis possibilities of finding results in short time increases. Docking has been a powerful tool till date when protein-DNA interactions comes into play and there are many softwares available which can perform this type of study such as PISA, PROMOTIF, X3DNA, ReadOut, DDNA, DCOMPLEX [31]. It is very much essential to perform systematic docking for the prediction of protein-DNA complexes [32] which were well supported by different method/techniques/program such as geometric hashing method [33], Fast Fourier correlation techniques [34] and HADDOCK program [35,36]. Henceforth, protein-DNA docking has been studied using one of the advanced versions of HADDOCK software [37]. HADDOCK (High Ambiguity Driven protein-protein DOCKing) program which starts with a similar rigid body docking of the two partners based on interaction surface definition, a semi-flexible simulated annealing stage followed by a water-refinement step [35].

DNA protein interaction studies have been reported to establish the involvement of conserved sequence GCGGCC box motif of DNA with Protein [38]. Protein–nucleic acid interactions therefore play a crucial role in central biological processes, ranging from the mechanism of replication, transcription and recombination to enzymatic events utilizing nucleic acids as substrates [39, 40]. Pandey and Kumar [39] reported the protein-DNA interaction of CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-1 protein using HADDOCK server. Chen et al. [41] reported that OsBP-73, a rice gene, encodes a novel DNA-binding protein with a SAP-like domain and their results suggest that OsBP-73 may play an important role in the regulation of cell. Since, during abiotic and biotic stresses regulatory mechanism involve the promoter motifs/cis-regulatory elements in various cellular mechanisms. Hence, the presence of GCC box in Ubiquinol Cytochrome C Chaperone (UCCC) gene promoter needs to be validated using a suitable molecular technique. In present study, first time we have tried to validate the GCC box (GCGGCC) in UR-DEG UCCC gene by using MBP based Real-Time PCR amplification assay to detect nucleic acid sequences [42-45]. In this study we have identified the presence of GCC box in UCCC Gene and also tried to establish the relationship between the sub1A protein interactions.
with the GCC box of the UCCC gene. For that we generated 3D structure of Sub1A protein by I-TASSER. DNA model of UCCC Gene promoter sequence having core GCCGCCG motif was generated by 3D-DART. Interaction of Sub1A and GCCGCCG motif was studied by HADDOCK server. Eventually, the involvement in regulation by interacting with core GCC box motif with Transcription factor was found out.

**MATERIALS AND METHODS**

**Identification of consensus promoter motif and DEG**

For the identification of DEG’s, online available microarray result of anoxic rice coleoptiles reported by Lasanthi-Kudahettige et al. [12] was used, which was stated in our previous work [21]. From the microarray result, genes were shortlisted as UR-DEGs (up-regulated DEGs, expression increased by equal or more than two-fold, ≥2X). Promoter sequence of -499 to +100 bp was retrieved from Eukaryotic Promoter Database (EPD) for each shortlisted UR-DEGs, DR-DEGs and UC-DEGs. Promoter motifs as well as their consensus promoter motifs were analyzed through MEME (v 4.5.0). It was also observed that ubiquinol-cytochrome C chaperone family protein gene (AK068288 or Os07g30790) was up-regulated during anoxia and also have detected GCC box motif in their promoter region.

**Designing of Molecular Beacon probes and gene specific primers**

To identify the presence of the detected GCC box promoter motif in the up-regulated ubiquinol-cytochrome C chaperone family protein gene, Molecular Beacon probe (MBP) and its specific primers were designed. Promoter sequence of this gene having length -499 to +100 was retrieved from the Eukaryotic Promoter Database (http://www.epd.isb-sib.ch/seq_download.html). Considering the location of GCC box sequences, MBPs and their specific primers were designed using Beacon Designer 7 (BD7, PREMIER Biosoft, USA). As per BD7 protocol to maintain the optimum difference of annealing temperature (Tm) between probe and gene specific primers, MBP length was adjusted having the sequence of GCCGCCGCCG rather than the core sequence GCCGCCG that was consensus motif in UR-DEGs [21]. Molecular Beacon compatibility score for probe and primers were also considered while its designing. Designed MBPs and primers were procured from GeneLink™ (USA) and Hysel India Pvt. Ltd, respectively.

**Isolation of Genomic DNA from Rice**

Rice seeds of Azucena (japonica sp.) were surface sterilized with 0.1% of HgCl2 and incubated in dark for 48 h at 35°C. Sterilized seeds [Figure-1(a)] were grown in pot at room temperature. Genomic DNA was isolated from rice seedlings [Figure-1(b)] using CTAB (2X) method and subjected to RNase treatment by standard protocol. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked in 0.8% agarose gel.

**Amplification profiling through Real Time PCR**

The identification of conserved promoter motif in the selected UR-DEGs was carried out by using MBP based Real Time PCR. The PCR amplification was performed in the total reaction volume of 15 µl (1X Taq polymerase, 0.2mM dNTPs, 3mM MgCl2, 0.45µM primer, 3ng gDNA and 0.3µM MBP) at optimized PCR condition (95°C for 4 min; 40 cycles of 15 s at 95°C, 35 s at 60°C, and 45 s 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 UR-DEGs, TCC box containing probe used as a reference.

**1K Promoter sequence and Protein/Genomic sequence retrieval**

1K Promoter sequence of Ubiquinol Cytochrome C Chaperonee gene (Os07g30790) and Protein sequence of Sub1A was retrieved from the TIGR release version6.1 (ftp://ftp.plantbiology.msim.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/).

**3D structure prediction of Sub1A protein**

The 3D structure of Sub1A was not available in PDB database, therefore its structure was predicted using I-TASSER (http://zhanglab.cshl.edu/I-TASSER/). I-TASSER is a hierarchical protein structure modeling approach based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA). I-TASSER (The iterative threading assembly refinement) server determines 3D structures of protein based on the sequence-to-structure-to-function paradigm algorithm. It predicts secondary structure, tertiary structure and functional annotations on ligand-binding sites, enzyme commission numbers and gene ontology terms. The accuracy of prediction is based on the confidence score of the modeling [46,47]. C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations.

**Construction of 3D DNA structure of promoter region having GCC-box motif**

Construction of 3D DNA structure of promoter region having GCC-box motif was performed with slight modification as described by Pandey and Kumar [38]. To study protein-DNA interaction a 3D model of DNA fragment (promoter region of 25 bases having core GCCGCCGmotif) was required. Therefore, 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) server was used for generating custom 3D structural model of DNA and its PDB file. The promoter fragment of 25 nt long having core GCCGCCG motif of UCCC gene was used for the 3D model of DNA. DNA was bended at angle of 40° with 5° tilt between 11-14 nucleotide (GCCG). 3D-DART uses the DNA rebuild functionality of software package 3DNA [48] and extends its functionality with tools to change the global
conformation of the DNA models from a sequence to a base-pair step parameter file [49].

**In-silico protein-DNA interaction studies**

Construction of 3D DNA structure of promoter region having GCC-box motif was performed with slight modification as described by Pandey and Kumar [38]. For in-silico protein-DNA interactions studies between 3D structure of Sub1A (1-TASSER generated 3D models) and 3D structure of DNA fragment having core GCC box motif (3D-DART generated models), HADDOCK web server (http://haddock.science.uu.nl/services/HdeADDOCK/haddockserver-easy.html) was used. HADDOCK (High Ambiguity Driven protein-protein Docking) is an information-driven flexible docking approach for the modeling of bimolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods with the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. These AIR files have information about active residues (directly involved in the interaction) in protein as well as in the DNA model. Result with the lowest HADDOCK score and Z-Score were considered as the best interaction between these molecules [50].

**RESULTS AND DISCUSSION**

Rice is the second largest produced cereal for half of the world population, posing itself as the staple food crop (http://www.irri.org/index.php). Harsh environmental conditions lead to partial to complete destruction of rice, among them flooding is considered as third major issue [51, 52]. Rice has a semiaquatic origin therefore; identification of traits associated with submergence tolerance with molecular techniques has been initiated and achieved. In this context, understanding the molecular mechanisms for submergence tolerance was very well supported by gene expression profiling through microarrays, quantitative Real-Time PCR analyses and Real-Time PCR based nucleic acid sequence detection [12, 14, 45 and 53-55] and transcriptome analysis using massively parallel signature sequencing [56, 57].

A large number of genes are differentially expressed during the submergence induced low oxygen stress involved in complex biochemical and genetic pathways [58, 9], ethylene biosynthesis [59] and enzymes encoded for sugar metabolism, glycolysis, and fermentation pathways in rice [60]. Plant promoters are the key component for studying the mechanism of transcriptional regulation. A core promoter contains the essential nucleotide sequences for the regulation of gene function known as transcriptional regulatory elements (cis- and trans-regulatory elements) and Transcription Start Site (TSS). The presence of transcriptional regulatory elements helps to regulate the function of transcription factors and their expression during normal and unfavorable (abiotic and biotic) conditions. Therefore, identification of these regulatory elements is very much essential. There are lots of reports about the identification and characterization of stress-responsive cis-regulatory elements [21 and 61-63].

In our previous study, consensus promoter motifs were identified using MEME that are common in their promoter region of differentially co-expressed genes in rice seedling under anoxia [20, 21]. The MEME detected GCC box (GCCGCC) as a consensus promoter motif in promoters of UR-DEGs with IC value of 16.6 bits, E-value 1.4e-056, and width length of 11 nucleotides (Figure-2). Similarly, Sharma et al. [54] identified 223 types of CREs associated with 40 rice sperm co-expressed genes by analysing 1-kb upstream regions with the help of MEME. Doi et al. [64] also identified 7514 motifs from 1-kb promoter of auxin-induced Arabidopsis prha homeobox gene using MEME.

**Figure 1.** (A) Surface sterilized Azucena (japonica sp.) seeds; (B) Rice seedlings grown in tray; (C) RNase treated genomic DNA of Azucena leaves

**Figure 2.** Identified consensus promoter motif GCC box in UR-DEGs with IC value by MEME (v 4.5.0)

**Figure 3.** Position of identified consensus promoter motif GCC box (green colored box) in UR-DEG ubiquinol-cytochrome C chaperone family protein gene of Eukaryotic Promoter Database Id

It was also observed that ubiquinol-cytochrome C chaperone family protein gene was up-regulated during anoxia and also have detected GCC box motif in their promoter region (Figure-3). Gene ontology suggested that the ubiquinol-cytochrome C chaperone family protein gene (Os07g30790) is associated with some biological process (GO:0008150) related to tissue respiration and found in mitochondrion as cellular component (GO:0005739). In principle mitochondria’s
primary roles are the oxidation of organic acids through the tricarboxylic acid cycle and the synthesis of ATP coupled to the transfer of electrons from reduced NAD+ to oxygen via the electron transport chain. Beside these primary functions, it also has important secondary functions in plants, like synthesis of nucleotides, amino acids, lipids, and vitamins [65-67]. Undertaking transcription and translation [68] through their own genome [69], actively import proteins and metabolites from the cytosol [70], influence programmed cell death [71], and respond to cellular signals such as oxidative stress [72,73].

The mitochondrial respiratory chain (MRC) and oxidative phosphorylation (OXPHOS) system are composed of five enzymatic complexes (I to V) present in the inner mitochondrial membrane and two mobile electron carriers (ubiquinone and cytochrome c). Electrons are donated from reducing equivalents, NADH and FADH2, to complex I (NADH: ubiquinone oxidoreductase) and complex II (succinate: ubiquinone oxidoreductase), respectively, and flow down an electrochemical gradient in the MRC until complex IV (cytochrome c: oxygen oxidoreductase), which catalyzes the reduction of molecular oxygen, the final acceptor of electrons, to water. Complexes I, III (ubiquinol: cytochrome c oxidoreductase; cytochrome bc1 complex), and IV use the energy liberated by the electron flux to pump protons from the mitochondrial matrix to the intermembrane space, generating a proton gradient across the mitochondrial inner membrane that is used by complex V to drive the synthesis of ATP from ADP and inorganic phosphate [74, 75]. Therefore, presence and experimental detection of GCC box in UR-DEGs becomes very much essential, which was achieved through MBP based Real Time PCR, an accurate and advance molecular technique.

Consequently, MBPs and their specific primers were designed using BD7 (PREMIER Biosoft, USA). Depending upon the parameters of BD7, forward and reverse primers for Ubiquinol-cytochrome C chaperone family protein gene were 5’-CCTCCTAGTTCGTCGTCACAA-3’ and 5’-TCGAGGGTTGGACTTCACC-3’, respectively. MBP for the validation of GCC box containing UR-DEGs was 5’-[6-FAM]CGCGATCGGCCGCCGGATCGGG[BBQ-1]-3’. Reporter dye 6-FAM (6 – Carboxyfluorescein) at 5’ and quencher BHQ1 (Black Hole Quencher@-1) at 3’ was used for designing the MBPs.

Rice seedlings of Azucena (Japonica sp.) were grown and sampled for genomic DNA isolation using CTAB (2X) method and RNA contamination was removed by the standard protocol of RNase treatment. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked at 0.8% agarose gel [Figure 1(c)].

Good quality of isolated gDNA was further used for the MBP based Real Time PCR assay with their specific primers for the identification of conserved GCC box promoter motif in up-regulated ubiquinol-cytochrome C chaperone family protein gene. The PCR assay confirmed the presence of GCC box in the promoter region of the gene. The graph (Rn vs cycle number and Dissociation curve) generated by the inbuilt software suggested the amplification of gene with the GCC MBP (Figure 4) has been occurred having Ct values 28.03 and 29.19 for the two replicates respectively (Table-1). The Rn vs cycle number graph (Figure 4A) indicate the a’ & b’ curves for amplification of selected genes with GCC probe; curves c’ & d’ for the amplification of genes with TCC probe and e’ & f’ curves for non-template control (NTC). In this experiment the TCC probe was used as a reference which didn’t showed any amplification suggesting the presence of GCC box (rather than TCC box) in the promoter of ubiquinol-cytochrome C chaperone family protein gene. Similar works has also been reported on nucleic acid sequence detection, sensitivity, accuracy and reliability of MBP [43, 44 and 76-80]. Similarly, a dissociation curve (Figure 4B) for the amplification of GCC and TCC box was also prepared indicating the amplification of only two specific products. Dissociation curves a’’ & b’’ depict the two independent replications for the amplification of GCC probe while c’’ & d’’ for TCC probe (Figure 4B).

![Figure 4.](http://bioinfo.aizeonpublishers.net/content/2013/5/bioinfo213-222.pdf)
activate gene expression in an ethylene-dependent manner [81]. It was reported that GCC-box work as an ethylene-responsive element that is essential in some cases for the regulation of transcription [81]. Hao et al. [82] reported that numerous members of the ERF family interact specifically with AGCCGCC through the conserved ERF domain. Fujimoto et al. [18] described that maltose binding protein–AtERF fusion proteins bind with GCC box sequence (AGCCGCC) and binding activity was abolished when both G residues within the GCC box were replaced by T residues (ATCCCTCC). Similarly, using electrophoresis mobility shift assay (EMSA) Cheong et al. [83] concluded that OsEREBP1 specifically binds to the GCC box (AGCCGCC) motif but not to the mutated GCC box (ATCCCTCC). Additionally, ethylene response factors are also involved in regulating jasmonate-responsive gene expression by interacting with the GCC-box. And introduction of point mutations into GCC-box sequence substantially reduced jasmonate responsiveness [84]. Using EMSA and transient expression assay TiERF1 protein binds with GCC box and modulate the defense response by up-regulating transcripts of a subset of genes with the GCC box present in their promoters [85].

<table>
<thead>
<tr>
<th>S. No</th>
<th>Replicates</th>
<th>Template</th>
<th>Molecular Beacon</th>
<th>Ct value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>R1</td>
<td>Template</td>
<td>GCC box</td>
<td>28.03</td>
</tr>
<tr>
<td>b</td>
<td>R2</td>
<td>Template</td>
<td>GCC box</td>
<td>29.19</td>
</tr>
<tr>
<td>c</td>
<td>R1</td>
<td>Template</td>
<td>TCC box</td>
<td>-</td>
</tr>
<tr>
<td>d</td>
<td>R2</td>
<td>Template</td>
<td>TCC box</td>
<td>-</td>
</tr>
<tr>
<td>e</td>
<td>R1</td>
<td>No</td>
<td>GCC box</td>
<td>-</td>
</tr>
<tr>
<td>f</td>
<td>R2</td>
<td>No</td>
<td>GCC box</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Best predicted model with their C-Score, TM Score and RMSD value where C-Score is the confidence score for the predicted model, TM-score is a measure of global structural similarity between query and template protein and Root Mean Square Deviation is the RMSD between residues that are structurally aligned by TM-align.

<table>
<thead>
<tr>
<th>Best Predicted Model</th>
<th>Locus Id</th>
<th>C Score</th>
<th>TM score</th>
<th>RMSD value (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Os09g1148</td>
<td>-4.06</td>
<td>0.28±0.09</td>
<td>15.8±3.2A</td>
</tr>
</tbody>
</table>

It is well known that Sub1A gene was responsible for the submergence tolerance in rice [16]. Therefore the protein sequence and structure of Sub1A was required. The protein sequence of Sub1A was retrieved from the TIGR (v6.1) (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudo_molecules/version_6.1/). The 3D structure of Sub1A was not available in PDB database, therefore its structure was predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER server predicts and displays various features in different sections for best model studies. It was considered that the prediction and generation of the best model based on C-Score, their structural analogs and binding sites. The quality of the generated models are estimated based on a confidence score (C-score), ranges from -5 to 2 where a high value signifies a model with a high confidence and vice-versa. C-score is highly correlated with TM-score and RMSD. Therefore, TM-score and RMSD are known standards to measure the accuracy of structure modeling thereby measuring structural similarity between two protein structures. RMSD is an average distance of all residue pairs in two structures and is sensitive to local errors (i.e., a mis-orientation of the tail) which occurs inspite of the correct global topology hence, TM-score must be used for solving these errors A TM-score >0.5 indicates a model of correct topology. Roy et al. [86] predicted the structures of three human GPCRs complexes using I-TASSER with a RMDS of 1.6Å, 2.27Å and 2.82Å to the crystal structures in the transmembrane region. The models predicted by I-TASSER were based on the best 10 threading templates available on RCSB PDB. The best predicted model is selected on the basis of confidence score; TM-Score as well as RMSD value (Table -2). The C score value for the best predicted model which is model 1 of Sub1A was -4.06 and furthermore, highly similar structures in PDB (as identified by TM-align) were identified and listed in Table-3. Template proteins with similar binding sites for Sub1A are listed in Table -4. The best binding site is predicted on the basis of C score LB (Range = 0-1) and BS-Score (>1) values. A higher score C score indicates a more reliable ligand-binding site prediction and BS-score reflects a significant local match between the predicted and template binding site [46,87]. Qin and Zhou [29] suggested that binding site prediction is a useful tool for building structural models for protein-DNA complexes and for experimental design and validation. Two best predicted binding sites for Sub1A 3D was taken for further interaction studies (Table-4). Structure of Sub1A protein predicted by I-TASSER and visualized by VMD tool has been shown in Figure-5.

Figure 5. Result showing 3D structure of Sub1A protein predicted by I-TASSER. The coloring method is based on secondary structure. The pink color represents α-helix and yellow color represents β-strand and deep sky blue color represents the coil in the 3D structure.
Table 3. Identified best two structural analogs in PDB where coverage represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein. Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

<table>
<thead>
<tr>
<th>PDB Hit</th>
<th>TM Score</th>
<th>RMSD (Å)</th>
<th>IDEN</th>
<th>Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4he8L</td>
<td>0.443</td>
<td>5.67</td>
<td>0.038</td>
<td>0.72</td>
</tr>
<tr>
<td>3rkL</td>
<td>0.441</td>
<td>5.75</td>
<td>0.046</td>
<td>0.724</td>
</tr>
</tbody>
</table>

For in-silico protein-DNA interaction studies 3D structure of protein along with DNA was required. Consequently, 3D structure of DNA segment (25 nt long) containing identified and validated GCC Box promoter motif of UCCC gene (Os7g30790) was generated by 3D-DART server. GCC-Box promoter motif positioned at 10-15 nucleotides have been shown in Figure-6. DNA was bended at angle of 40° with 5° tilt between 11-14 nucleotides (GCGG).

Figure 6. 3D structure of linear DNA segment generated by 3D-DART server of GCC-Box promoter motif positioned at 8-17 nucleotide. In DNA model red color represents Adenine and pink color represents Guanine and sea green color represents cytosine and gold yellow color represents Thymine.

Protein-DNA interactions are the physical basis of gene expression and DNA modification for vital biological activities [29]. Because there is no simple mapping code between DNA base pairs and protein amino acids, the prediction of protein-DNA interactions is a challenging problem. Therefore, HADDOCK can make use of a broader array of restraints, including those derived from biochemical and biophysical data [37]. Determining the structure of protein-DNA complexes and clarifying the factors that regulating their interaction is essential to better understand many biological processes [30]. A review describing the experimental strategies currently employed to solve structures of protein-DNA complexes and to analyze their dynamics has been published [88]. Protein-DNA interactions facilitate the fundamental functions of living cells and are universal in all living organisms [89].

To determine the protein-DNA interactions the Easy interface of HADDOCK web server was used [50]. Before going for docking AIR files were generated for both the interacting molecules having information about the active binding sites of Sub1A protein as well as in the DNA model. The HADDOCK score is the weighted sum of van der Waals energy (negative indicating favorable interactions), electrostatic energy (negative indicating favorable interactions), distance restraints energy (only unambiguous and AIR (ambig-restraints), direct RDC restraint energy, inter vector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, buried surface area (negative weight indicate a better interface), binding energy, desolvation energy. Meanwhile, the solution structures are analyzed for their intermolecular hydrogen bonds and intermolecular hydrophobic contacts by HADDOCK, the solutions are clustered according to the interface ligand RMSDs. The Z-score indicates the standard deviations from the average of a particular cluster in terms of HADDOCK score.

Table 4. Template protein for similar binding sites. Binding sites represents the amino acid positions

<table>
<thead>
<tr>
<th>C-score</th>
<th>PDB Hit</th>
<th>TM-score</th>
<th>RMSDÅ</th>
<th>IDEN</th>
<th>Coverage</th>
<th>BS-score</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.31</td>
<td>1gccA</td>
<td>0.236</td>
<td>1.5</td>
<td>0.629</td>
<td>0.248</td>
<td>1.67</td>
<td>102, 103, 104, 106, 108, 116, 118, 120, 141</td>
</tr>
<tr>
<td>0.26</td>
<td>1gccA</td>
<td>0.236</td>
<td>1.5</td>
<td>0.629</td>
<td>0.248</td>
<td>1.76</td>
<td>106, 108, 110, 112, 118, 125, 127, 129, 130</td>
</tr>
</tbody>
</table>

Table 5. HADDOCK score of protein and UCCC gene (Os7g30790) segment containing GCC-Box depicted by HADDOCK server.

<table>
<thead>
<tr>
<th>Model</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>SAUGCM1-BS1</td>
<td>33.1 ± 3.8</td>
<td>8.5 ± 0.6</td>
<td>-55.0 ± 7.3</td>
<td>-640.1 ± 63.6</td>
<td>44.1 ± 10.6</td>
<td>1719.7 ± 29.43</td>
<td>1645.3 ± 95.6</td>
<td>-1.2</td>
</tr>
<tr>
<td>SAUGCM1-BS2</td>
<td>12.4 ± 5.7</td>
<td>16.6 ± 0.6</td>
<td>-61.9 ± 33.3</td>
<td>-613.6 ± 30.2</td>
<td>32.0 ± 6.8</td>
<td>1650.2 ± 19.23</td>
<td>1912.0 ± 117.0</td>
<td>-1.2</td>
</tr>
<tr>
<td>SAUGCM2-BS1</td>
<td>27.8 ± 3.6</td>
<td>8.7 ± 0.2</td>
<td>-60.5 ± 59.0</td>
<td>-574.1 ± 24.4</td>
<td>34.4 ± 2.5</td>
<td>1687.2 ± 49.37</td>
<td>1741.7 ± 115.4</td>
<td>-1.2</td>
</tr>
<tr>
<td>SAUGCM2-BS2</td>
<td>11.7 ± 4.4</td>
<td>0.9 ± 0.5</td>
<td>-56.2 ± 3.0</td>
<td>-597.9 ± 63.0</td>
<td>18.5 ± 5.7</td>
<td>1689.9 ± 39.38</td>
<td>1804.5 ± 70.5</td>
<td>-1.7</td>
</tr>
</tbody>
</table>

For the prediction of best interaction, different binding sites and models of Sub1A were docked with 3D structure of DNA segment having promoter motif of UCCC gene (Table-5). The result showed best model SAUGCM2-BS2 (Table-5) has HADDOCK score of 11.7 +/- 4.4 and Z-score of -1.7 when interacting with bended DNA segment (Figure-7).
Figure 7. Result showing interaction between 3D structure of Sub1A protein predicted by I-TASSER and DNA model (promoter region having complementary GCCGCC core motif, CGGCGG) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 9, 10, 12, 13, 15 and 16 position while green color represents Guanine at 8, 11, 14, and 17. Interacting residues (THR at 130 and TRP at 110, 127 and ARG at 106,108, 112, 118, 125 and GLY at 129 positions respectively) in protein model represented with red color.

CONCLUSION

The aim of the present study was to identify consensus promoter motifs and their interaction with Sub1A. Previous study suggested that the consensus promoter motif GCC box sequence was found in the up-regulated promoter of Ubiquinol the cytochrome c chaperone family gene. Identification of consensus promoter motif in the promoter of up-regulated gene was done by using MBP based real time PCR. Real time PCR result showed that the presence of GCC box in promoter sequence through amplification whereas no template controls (NTC) and negative control (TCC box in promoter sequence) showed no amplification. In Silico study on 3D structure prediction and protein-DNA interaction of UCCC gene in rice has been studied and revealed that core GCC box (GCCGCC) was identified in antisense strand of UCCC gene promoter. It predicted a good binding affinity of Sub1A protein with Core GCC box promoter motif. Thus it can be concluded that Computational analysis for Sub1A protein and interaction with promoter motif have been performed successfully. Furthermore, to confirm the above interaction validation study is very much required in future.

Acknowledgement

This work was financially supported by University Grants Commission 10 (UGC), GOI, New Delhi for the major research project [F.No. 37-113/2009 (SR)]. Mr. Gopal Kumar Prajapati is working as Project Fellow and financially supported from the above mentioned UGC sponsored project. Use of Beacon designer 7 software for designing of probe was generously provided by PREMIER Biosoft, USA. Central Rainfed Upland Rice Research Station (CRURRS), Hazaribagh-16 825301, Jharkhand, India is gratefully acknowledged for providing the Azucena rice seeds used in this experiment.

REFERENCES


© 2013; AIZEON Publishers; All Rights Reserved
This is an Open Access article distributed under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

http://bioinfo.aizeonpublishers.net/content/2013/5/bioinfo213-222.pdf

222