Rubber Science, **28**(2): 147-158, 2015 RESEARCH ARTICLE

IN SILICO CHARACTERIZATION OF A CALCIUM-DEPENDENT PROTEIN KINASE FROM *HEVEA BRASILIENSIS* REVEALS PROSPECTIVE FEATURES FOR CONFERRING MULTIPLE STRESS TOLERANCE

Suni Anie Mathew, R. Supriya and A. Thulaseedharan[1*]

Rubber Research Institute of India, Kottayam, Kerala, India
[1]School of Biological Sciences, Central University of Kerala, Riverside Transit Campus, Padanakkad-671 314, Kasargod, Kerala, India

Received: 06 February 2015 Accepted: 20 May 2015

Mathew, S.A., Supriya, R. and Thulaseedharan, A. (2015). *In silico* characterization of a calcium-dependent protein kinase from *Hevea brasiliensis* reveals prospective features for conferring multiple stress tolerance. *Rubber Science*, **28**(2): 147-158.

Identifying potential genes imparting stress tolerance is an important step for developing rubber clones that can survive in stress-prone geographical locations. This will help in increasing rubber production by extending the area under *Hevea* plantations. Calcium-dependent protein kinase (*cdpk*) is one such gene involved in multiple stress signaling pathways. In the present study, a *cdpk* gene and its promoter region were isolated from the high latex-yielding *Hevea* clone RRII 105 and characterized. The intron pattern analysis of the genomic sequence classified the gene into Group II subfamily of CDPK proteins. The *in silico* analysis predicted the myristoylation site, palmitoylation sites, presence of nuclear localization signal and subcellular localization, hinting its role in signal transduction, protein-protein interactions and shuttling mechanisms during stress. The sequence analysis of the promoter region showed stress-responsive *cis*-elements that help in regulating gene expression. The sequence alignment and 3D modeled protein structure superposition of the isolated *cdpk* with *Arabidopsis cdpk21* is also predicted which is useful in identifying the orthologous nature between the two proteins, contributing to their functional similarity involved in multiple stress signaling. These results suggest that the isolated *Hevea cdpk* gene confers features for imparting multiple abiotic stress tolerance.

Keywords: Abiotic stress, Calcium-dependent protein kinase, *Hevea brasiliensis*, Natural rubber, Protein superposition

INTRODUCTION

Natural rubber is used in the manufacture of more than 35,000 products (Parthasarathy *et al*., 2006) making it a "hot" commodity (Fox and Castella, 2013). The perennial tree, *Hevea brasiliensis* is the exclusive commercial source of natural rubber (NR) owing to its abundance in the tree, good quality and ease of harvesting (Oh *et al*., 2000) compared to other latex producing plants. However, the high

*Correspondence: A. Thulaseedharan (Email: thulaseedharan@rubberboard.org.in)

consumption rates encourage researchers to conduct studies on increasing NR production by maximizing product quality, improving production capacity, reducing product waste/recycling and enhancing productivity. Though other alternatives like synthetic rubber are available in the market and useful to some extent, the main raw material of large-scale industries such as the automobile industry is basically natural rubber. One of the plausible solutions for improving rubber production capacity is to extend areas under rubber cultivation. The traditional areas have already been saturated with existing *Hevea* plantations. This has led to rubber development programmes in the non-traditional areas (Saraswathyamma, 1993). In spite of the marginal increase in the overall rubber production, the traditionally high latex yielding *Hevea* clones such as RRII 105 do not perform well in unfavourable climatic conditions present in these regions (Vinod *et al.,* 1996). Genetic improvement of *Hevea* clones to withstand abiotic stress is, therefore, a major area of research focus.

In the present scenario, it is essential to study potential genes that respond to environmental stresses, so as to genetically improve rubber plants with these genes through genetic engineering. However, in the environment, plants are not challenged by a single stress. Cellular homeostasis of a cell is disrupted during stress; which is then restored when different genes come into play. Different stress pathways share common secondary signals and therefore, one primary stress condition may activate multiple signaling pathways. These pathways may connect or interact with one another using shared components generating networks (Knight and Knight, 2001). Finally, stress-responsive genes having different functions are activated together to bring the cell back to normal. Therefore, the approach of genetic transformation should not be limited to a single action gene which may help in overcoming only a particular effect of stress, ignoring accompanying effects of one or more stresses. When it is crucial to make a plant tolerant to multiple stresses; instead of transferring a single-action gene that produces a specific stress protein it can be a more promising approach to transfer a gene that regulates several genes involved in multiple stresses (Bhatnagar-Mathur *et al.,* 2008).

Genes triggered at the early phase of a stress signaling pathway like protein kinases are regulatory in nature activating several genes in the late phase including those that produce transcription factors. One of the major stress-activated early phase genes reported is calcium-dependent protein kinase (CDPK) due to their rapid biochemical activation under multiple stresses (Ludwig *et al.,* 2004) and could be a potential candidate gene for multiple stress tolerance. There are numerous reports on *cdpk* genes, widely studied in *Arabidopsis* and other annual plants such as rice and wheat. The over-expression of rice gene *Oscdpk7* conferred tolerance to cold, salt and drought stresses (Saijo *et al.,* 2000). A wheat *cdpk* gene was found to respond to multiple stresses suggesting the idea of *cdpk* genes being converging points in multiple stress signaling pathways (Li *et al.,* 2008). Calcium-dependent protein kinases NtCDPK2 and NtCDPK3 responded to various stress stimuli (Witte *et al.,* 2010). A *cdpk1* mRNA sequence isolated from *H. brasiliensis* clone RRIM 600 was relatively more expressed in laticiferous cells and played a significant role in rubber biosynthesis (Zhu *et al.,* 2010). Further studies regarding CDPK proteins related to abiotic stress signaling are scanty in this tree species.

In the present study, the *cdpk* gene from high yielding and relatively stress

susceptible *Hevea* clone RRII 105 was isolated and characterized. The plausibility of the gene to be an ideal candidate for multiple stress tolerance was explored through bioinformatic tools. The isolation and characterization of the promoter region was also done since the identification of *cis*-acting promoter elements is important in recognizing its role. The intron pattern analysis aided classification of the gene while the comparison of the modeled protein structure with functionally characterized *Arabidopsis cdpk21* protein structure gave insights into the functional capabilities of the isolated *Hevea cdpk* gene.

MATERIALS AND METHODS

Medium mature, uninfected leaf samples were collected from *H. brasiliensis* clone RRII 105 growing in the experimental gardens of Rubber Research Institute of India. All chemicals used for the study were from M/s. Sigma Aldrich, USA and enzymes from M/s. MERCK, Bangalore. The GFX Gel Band purification kit was procured from M/s. GE Healthcare, USA while Improm II Reverse transcription (cDNA synthesis) kit and pGEM®-T Easy vector transformation kit were purchased from M/s. Promega, USA.

Genomic DNA isolation

The genomic DNA was isolated from 0.1g leaf samples, using the modified CTAB method (Doyle and Doyle, 1990). The quality of the isolated DNA sample was analyzed using 0.8 per cent agarose gel and the DNA concentration was determined by a nano drop - spectrophotometer (M/s. Nanodrop, USA).

RNA isolation and cDNA synthesis

RNA was isolated from 0.1g leaf samples of *H. brasiliensis* clone RRII 105 (Venkatachalam *et al.,* 1999). The RNA samples were electrophoresed on 0.8 per cent agarose gel for 2h at 50V. Aliquot of 20 µl was then treated with 2U of RNase-free DNase enzyme to degrade traces of genomic DNA contamination. cDNA was prepared from 1µg of these DNA-free RNA samples using ImpromII mRNA Reverse Transcription kit according to the manufacturer's protocol.

PCR amplification

Forward and reverse primers for amplification of full-length of *cdpk* gene was designed based on the previously reported (Zhu *et al.,* 2010) mRNA sequence of *Hevea* CDPK1 (NCBI #EU581818.2) isolated from *H. brasiliensis* clone RRIM 600. The forward primer was F1 – 5′ ATG GGT TGT TGT AGC AGC 3′ whereas the reverse primer was R1 – 5′ CTA GAA AAG CTT TCC TGG 3′. PCR for cDNA amplification was performed with 1X Taq buffer (Buffer A with 1.5 mM $MgCl_2$), 100 µM dNTPs, 0.5U Taq polymerase enzyme, 250 nM each of the above forward and reverse primers and 0.1X dilution of cDNA template prepared from 1µg RNA. PCR amplification of full length *cdpk* gene was done from the genomic DNA using the same concentration of reaction components except for Taq DNA polymerase enzyme being 0.25U used as a combination with 0.5U XT-20 DNA polymerase. PCR amplification from cDNA was done at an initial denaturation of 95 °C for two min followed by 36 cycles of 95 °C for 1 min, 55 °C for one min, 72 °C for one min and a final extension of 72 °C for 10 min. The amplification of *cdpk* genomic sequence was also done at the same PCR cycle conditions except for the extension being prolonged for 3 min. The amplicons from cDNA and genomic DNA samples were electrophoresed on 1.2 per cent and 1.0 per cent agarose gels, respectively alongwith *Eco*RI/*Hind*III double digest λDNA marker. The amplicons were eluted and purified from the

agarose gel using the GFX Gel band purification kit.

From the reported *Hevea cdpk* promoter sequence (NCBI #JF06669.1) from RRIM 600, the forward primer F2 - 5′ GCA CTT GCG GAT ACA CCG TAT 3′ and reverse primer R2 - 5′ CAACAACCCATCTTGCACTCA 3′ were designed for the amplification of *cdpk* promoter sequence from genomic DNA of RRII 105. The PCR reaction components were the same as used for the amplification of the gene at PCR conditions of initial denaturation of 95 °C for 2 min followed by 36 cycles of denaturation at 94 °C for one min, 55 °C for one min, 72 °C for one min and a final extension at 72 °C for ten min. The amplicons were electrophoresed on 1.0 per cent agarose gel alongwith *Eco*RI/*Hind*III double digest λDNA marker.

Cloning of PCR products and sequencing

The amplicons were mixed with cloned pGEM®-T Easy vector for overnight ligation according to manufacturer's instructions and transformed into competent DH5α *E.coli* cells. After shaking incubation at 37 °C for two hours, the cells were plated on solid LB agar medium and incubated at 37 °C overnight. Recombinant plasmids were then isolated from the transformed colonies using the alkaline lysis method (Sambrook *et al.*, 1989) for plasmid isolation. The plasmids were electrophoresed on 0.8 per cent agarose gel to check integration of inserts into the specific plasmids. The plasmid samples were sequenced at M/s. Macrogen, Korea.

Bioinformatic analysis

The isolated nucleotide sequences were analyzed using blastn (http:// blast. ncbi. nlm. nih. gov/Blast.cgi) and ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2) programmes. The alignment to compare the isolated sequence with other Group II members of the *Arabidopsis* was done through ClustalW analysis. The promoter elements in the *cdpk* promoter sequence were detected using the PLACE tool (http://www.dna.affrc.go.jp/PLACE/). The amino acid sequence was deduced using ExPASY translate tool (http://web.expasy.org/translate/) and analyzed using the Conserved Domain (CD) search tool (www.ncbi.nlm.nih.gov/structure/cdd/wrpsb.cgi) to identify the sequence of these domains, active sites and EF hands. The Myristoylator tool (http://web.expasy.org/myristoylator) and Palmitoylator tool (http://web.expasy.org/palmitoylator) predicted the myristoylation and palmitoylation sites, respectively. With the help of Plant-mPLoc tool (http://www.csbio.sjtu.edu.in/bioinf/plant-multi/) and NLS Prediction software cNLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi), the subcellular localization of the protein was virtually determined. The protein sequences were submitted in the Protein Model Port (http://www.proteinmodelportal.org/) which generates protein models using Swiss-MODEL workspace (http://swissmodel.expasy.org/workspace/) and Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/). The superposition and comparison of the two models were done using FATCAT software program (http://fatcat.burnham.org/).

RESULTS AND DISCUSSION

Isolation and characterization of *cdpk* cDNA, *cdpk* gene and promoter region from *Hevea* clone RRII 105

Through semi-quantitative PCR amplification using primers F1 and R1, a 1.6 kb cDNA fragment was obtained (Fig.1A) from cDNA which was subsequently cloned into pGEM®T Easy vector and transformed into *E.coli* cells. The recombinant plasmids

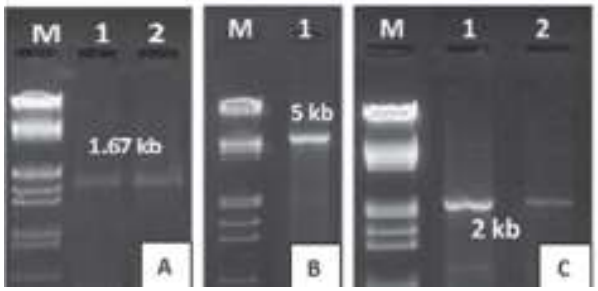

Fig.1. Gene fragments that are amplified using *cdpk* gene-specific primers visible on agarose gels A – 1.67 kb *cdpk* fragment amplified from cDNA using primers F1 and R1; M – Lambda Eco RI/Hind III double digest marker, 1, 2 – amplified fragments from cDNA samples; B – 5 kb *cdpk* gene fragment amplified from *Hevea* clone RRII 105 genomic DNA; M –Lambda Eco RI/Hind III double digest marker, 1 - amplified fragments from genomic DNA sample; C – 2 kb *cdpk* promoter region amplified from *Hevea* clone RRII 105 genomic DNA using primers F2 and R2; M –Lambda Eco RI/Hind III double digest marker, 1,2 - amplified fragments from genomic DNA sample

were isolated and sequenced to obtain 1671 bp showing 99 per cent identity with the *cdpk* nucleotide sequence of *Hevea* RRIM 600 (EU581818.2), 88 per cent with *Ricinus communis*, 86 per cent with *Populus* sp., 85 per cent with *Theobroma cacao* and more than 80 per cent with other plant species.

The full length gene amplified from genomic DNA through semi-quantitative PCR amplification using the primers F1 and R1 produced a 5.0 kb fragment (Fig.1B), cloned into pGEM®T Easy vector was transformed into *E.coli* DH5α cells. The isolated recombinant plasmids were sequenced to obtain a fragment containing a total of 5152 bp. The sequencing of such a large fragment was done through primer walking using the following primers. Initial forward and reverse reactions were done using the M13F and M13R universal primers as these are the flanking sequences in the vector. From the obtained fragments, subsequent primers were designed from the forward and reverse directions as each consequent fragments were sequenced from either direction. The primers used for sequencing were: 5′ GCA AGG ATG AGG GAG CAA TG 3′, 5′GCT CCT GAA GTA TTG AGG CGG 3′, 5′GGA GGT GCA GCA TCA GAC AAG 3′ in the forward direction and 5′CGC CTC AAT ACT TCA GGA GCC 3′, 5′ GCC ACT CTT GTC AGT GTC C 3′ in the reverse directions. Blastn analysis of the gene sequence showed 99 per cent similarity with CDPK1 mRNA sequence of *H. brasiliensis* (RRIM 600) and with CDPK sequences of other species such as *R. communis*, *P. trichocarpa*, *Panax ginseng* and *Vitis vinifera* except for the introns. The clustalW alignment of the isolated *cdpk* cDNA sequence with isolated *cdpk* genomic sequence showed that the gene contains eight exons and seven introns with GT-AG splice junctions. The exons aligned perfectly with the cDNA sequence.

For the characterization and functional prediction of the gene, analysis of the promoter region is also significant. In this view, the promoter region of the gene was isolated from genomic DNA through PCR amplification using primers F2 and R2 (Fig. 1C); cloned and sequenced. The isolated region of the promoter sequence consists of 2091bps and analyzed using the PLACE software to identify the promoter *cis*-elements that are involved in the regulation of gene expression. The study of these *cis*-elements also helps in determining whether the isolated gene is stress-responsive or not. The major stress-responsive elements identified include DRECRECOREAT element at position 342, MYBCORE elements at positions 1021 and 1364, WRKY elements at positions 109, 562 and 694. The presence of DRECOREAT, MYBCORE and WRKY stress responsive *cis*-elements hints that this gene may be responsive to multiple stresses such as drought, cold, salt as well as pathogenic

stresses (Rushton *et al.,* 2010; Ayadi *et al.,* 2004). Differential expression of CDPKs has been observed in response to diverse stimuli, including abscisic acid (ABA), cold, drought, salinity, heat, elicitors and pathogens, which correlates with the presence of stress-responsive *cis*-elements in rice *CDPK* gene promoters (Boudoscq and Sheen, 2013). When the promoter region was combined with the obtained genomic sequence of *cdpk* and subjected to blastn analysis with the whole genome shotgun contigs of *H. brasiliensis* (RRIM 600), it showed 99 per cent similarity with two contigs NCBI#AJJZ010952744.1, AJJZ010952745.1 showing query coverage of 95 per cent and 4 per cent respectively. The contigs when aligned together was found to have an overlapping region depicting that the two contigs represented a contiguous sequence indicating that the isolated promoter region is specific to the isolated gene sequence. The isolated genomic sequence is now available in the NCBI database under the accession NCBI: KJ939362.

Characterization of the deduced amino acid sequence

The 1.6 kb cDNA sequence spans the entire open reading frame encoding a protein sequence of 556 amino acids. The blastp analysis of the deduced amino acid sequence mainly showed 99, 85 and 80 per cent identities with CDPK protein sequences of *H. brasiliensis* (RRIM 600), *R. communis* and *T. cacao,* respectively. With CD (Conserved Domain) search tool the four characteristic domains of CDPK proteins - the N-terminal domain, the catalytic Ser/Thr kinase domain containing the catalytic active site, the autoinhibitory junction domain and the calmodulin-domain with four EF hands for binding the calcium ions were identified (Fig. 2). The CDPK amino acid sequence is sub-divided into 12 subdomains characterized by consensus sequences. These subdomains were identified by comparing the deduced amino acid sequence with the sequence of CDPK isozymes isolated from corn root by Takezawa *et al.* (1996). With the ScanProsite tool the sequences of these domains, functional active sites and EF domains were identified (Fig. 2). The deduced amino acid sequence has a molecular weight of 62 kDa with a pI of 6.36 as calculated using the ProtParam tool. The blastn analysis of the *cdpk* cDNA and genomic sequence along with the identification of conserved domains on the deduced amino acid sequence confirms that the isolated sequence is a *cdpk* gene.

The post-translational modifications of the deduced amino acid sequence were also predicted. The myristoylator tool indicates that the N-terminal glycine residue at position 2 can be myristoylated. The palmitoylator tool predicts the N-terminal cysteine residue at position 3 or 4 can be palmitoylated. The presence of a bipartite Nuclear Localization Signal (NLS) was identified within the junction domain with the help of cNLS mapper (Fig. 2B) tool. The subcellular localization of the protein was predicted to be in the nucleus with the PlantmPLoc tool which could probably be attributed by the presence of NLS.

The prediction of N-terminal myristoylation and palmitoylation sites in the deduced amino acid sequence is in consistent with the fact that majority of the CDPK proteins in *Arabidopsis* (Cheng *et al.,* 2002), rice (Asano *et al.,* 2005) and canola (Zhang *et al.,* 2014) have an N-terminal glycine residue which is myristoylated and N-terminal cystiene residue which is palmitoylated. The irreversible cotranslational lipid modification of myristoylation is

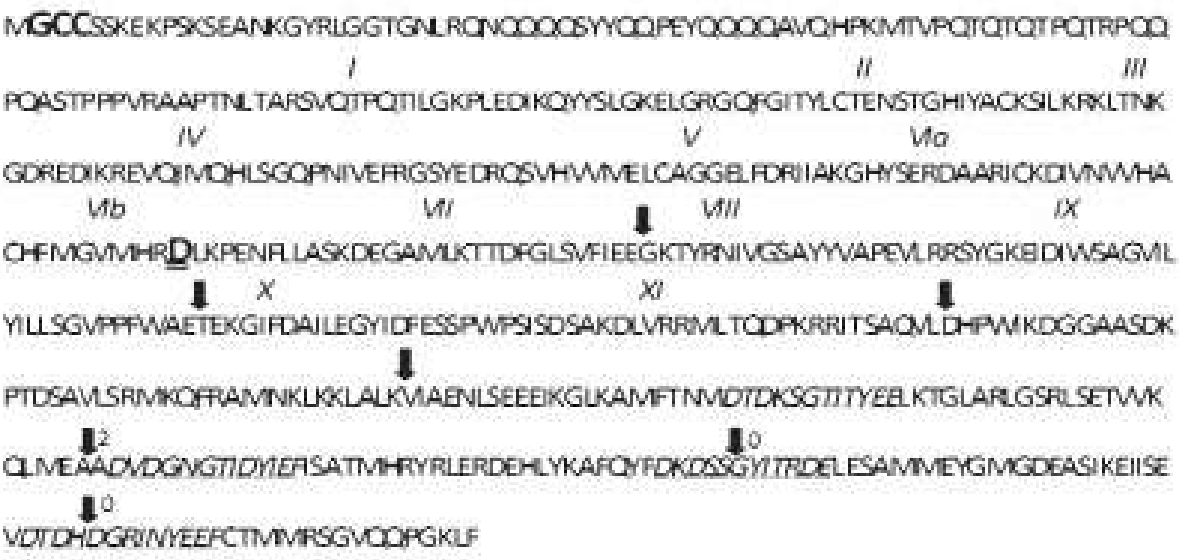

Fig. 2. The amino acid sequence of the isolated *Hevea cdpk* gene was deduced using the ExPASY (http://web.expasy.org/translate/) translate tool and characterized.
The sequence consists of 556 amino acids. Conserved Domain (CD) search tool identified the characteristic domains of the protein sequence consisting of kinase domain (region 150 – 275; dotted underlined) in which the active site aspartate (D) is shown (bold, double underlined), autoinhibitory junction domain (region 300-375; continuous underlined) and calmodulin (CaM)-like domain with 4 EF-hands (region 400 – 550; italics and serrated underlined). The positions of the Myristoylation site - Glycine (G) residue at position 2 (large bold), Palmitoylation site – Cytosine (C) at positions 3 or 4 (large bold), NLS sequence within the junction domain (continuous underlined) are indicated. The subdomains (roman numerals); the position of insertion of introns (black arrows) and phases of introns (0, 1, 2 next to the arrows) are indicated.

associated with membrane targeting and protein-protein interactions as seen in a wide variety of signaling proteins involved in environmental stresses (Podell and Gribskov, 2004).

The deduced amino acid sequence has two cystiene residues at the N-terminal and either of the two may get palmitoylated as in rice CDPK2 (Martin and Busconi, 2000) where both cystiene residues are potential sites for palmitoylation. Unlike myristoylation, palmitoylation is a reversible lipid modification required for stabilizing membrane localization

(Yavlovsky *et al.,* 1999). It may provide the potential role of shuttling the CDPK proteins between cytoplasm and membranes in response to various stimuli which is reported in other signaling molecules (Michel *et al.,* 1993; Wedegaertner *et al.,* 1995). This potential role of the deduced amino acid is further supported by the fact that a bipartite NLS was detected in the junction domain and the subcellular localization was predicted to be in the nucleus. The CDPK family of proteins is multifunctional and targeted to different organelles, including nucleus (Schulz *et al.,* 2013) This is in consistency with CDPKs containing bipartite NLS that were found to be localized to the nucleus when the cells were stressed as in the case of *Arachis* (Raichaudhuri *et al.,* 2006) and *Mesembryanthemum* (Chehab *et al.,* 2004). TaCDPK3 and TaCDPK15 in wheat having bipartite NLS can also potentially migrate from plasma membrane to nucleus during signal transduction (Li *et al.,* 2008).

Intron pattern analysis

The positions of introns were identified by aligning the isolated genomic sequence with the isolated cDNA sequence. To understand which of the 12 subdomains are cleaved by the introns, each exon region was translated through ExPASY translate tool and position as well as phase of the introns were identified (Fig. 2) as reported from the CDPK subfamily of *Arabidopsis* (Harmon *et al.,* 2001). The intron pattern analysis, especially the phase 2 intron present within the EF3 domain showed the distinctive features of Group II *cdpk* subfamily. The identification and classification of CDPK proteins in rice (Asano *et al.,* 2005), wheat (Li *et al.,* 2008), canola (Zhang *et al.,* 2014) and the tree *Populus* (Zuo *et al.,* 2013) were based on the classification in *Arabidopsis* (Harmon *et al.,* 2001).

Sequence alignment of the isolated *Hevea cdpk* gene with *Arabidopsis cpk21* gene and comparison of modeled 3D protein structures

The isolated *Hevea cdpk* gene was aligned with Group II family *cdpk* gene sequences of *Arabidopsis* - *cdpk* 9, 15, 21, 22, 23, 27, 29, 31, 33 - as the position and phases of introns of the new sequence were identified by sequence comparison with the latter. The sequence showed significant similarities with these isoforms. The alignment scores of nucleotide sequences of *Hevea cdpk* with *Arabidopsis cdpk* showed maximum alignment score with *Arabidopsis cpk21* (At4g04720 NCBI: NM_116710.2) (Fig. 3). The amino acid sequence alignment showed considerable difference in the N-terminal variable domain region while in the other domains majority of regions remained consensus intermittently interspersed with non-consensus sites as well as semi-conserved substitutions.

The protein structures for the *Hevea cdpk* and *Arabidopsis cdpk21* sequence were modeled by submitting the sequences in the protein model portal where the protein structures were modeled using SWISS-MODEL and Phyre2. The Phyre2 structures were more reliable as they provide detailed view on how the protein model is designed. After close observations of each amino acid residue the most probable secondary structure predictions are made as to whether each region would form a-helix, b-sheet or a turn. Each region is the closely analyzed whether the combinations of these secondary structures would form an ordered or disordered region in the tertiary structure. Next, the fold recognition of the protein structures is made based on the templates (already reported protein structures). The templates are selected based on sequence similarity and domain

Alignment scores of *Hevea cdpk nucleotide* sequence with *Arabidopsis cpk* nucleotide sequences

Sequence 1: *Hevea*	1671 bp	Start of Pairwise alignments
Sequence 2: Cpk9	1626 bp	Sequences (1:2) Aligned. Score: 72.20
Sequence 3: Cpk15	1665 bp	Sequences (1:3) Aligned. Score: 74.23
Sequence 4: Cpk21	1596 bp	Sequences (1:4) Aligned. Score: 76.87
Sequence 5: Cpk22	1718 bp	Sequences (1:5) Aligned. Score: 61.64
Sequence 6: Cpk23	1602 bp	Sequences (1:6) Aligned. Score: 71.79
Sequence 7: Cpk27	1458 bp	Sequences (1:7) Aligned. Score: 71.19
Sequence 8: Cpk29	1686 bp	Sequences (1:8) Aligned. Score: 67.27
Sequence 9: Cpk31	1455 bp	Sequences (1:9) Aligned. Score: 71.96

Fig. 3. Alignment scores of *Hevea cdpk* sequence when aligned with Group II members of *Arabidopsis* CDPK subfamily. The sequence showing highest alignment score with *cpk21* is underlined.

architecture between query and template (Kelly and Sternberg, 2009). Out of the 50 templates identified for each structure the protein structure is made based on 6 top templates – c3q5iA, c3ku2A, c3igoA, c3hztA, c4mvfa and c3lijA. The *Hevea* CDPK protein model has 83 per cent of the query sequence (446 residues) modeled with >90 per cent confidence level. Though 29 per cent of the region has been detected as disordered; 42 per cent forms alpha helix and 13 per cent forms beta-sheet while 3 per cent forms transmembrane helix. The *Arabidopsis* CDPK21 has 86 per cent residues modeled with >90 per cent confidence level out of which 26 per cent has been detected as disordered region, 42 per cent forms alpha helix, 13 per cent forms beta-sheet and 3 per cent forms transmembrane helix. The conserved domains – kinase domain and calcium-binding sites of both structures are modeled with high confidence level. The structures were validated using the QMEAN server programme. The QMEAN score is a combination of 6 terms – C-beta interaction energy, all-atom pairwise energy, solvation energy, torsion angle energy, secondary structure agreement and solvent accessibility energy. Reliable models have a total Qualitative Model Energy Analysis (QMEAN) score value between 0 and 1 (Benkert *et al*., 2009). The *Hevea* CDPK protein has a total QMEAN score of 0.597 while that of *Arabidopsis* is 0.593, making them reliable protein models.

The modeled protein structures were superimposed using the FATCAT software to predict if they are similar. From the results, it was evident that the structures showed significant similarity (P-value = 0.00e+00) and they superimposed (Fig. 4). While it is a known fact that the proteins do not have rigid structures, there is limitation in many modeling programs that the modeled protein structures cannot be altered to fit each other. But this program allows flexibility which is an added advantage differing from other modeling programs although the above structures were superimposed without any twists. The structure alignment has 424 equivalent positions with a root-mean square deviation (RMSD) of 2.93 Å. The reliable superposition of the two structures provides an insight on the evolutionary aspect and orthologous nature of the particular proteins. These results tend to render the potential functional similarity between these two proteins.

The significant sequence similarity to the *Arabidopsis cdpk* nucleotide sequences in the Group II subfamily also confirms the classification of the isolated *Hevea cdpk*

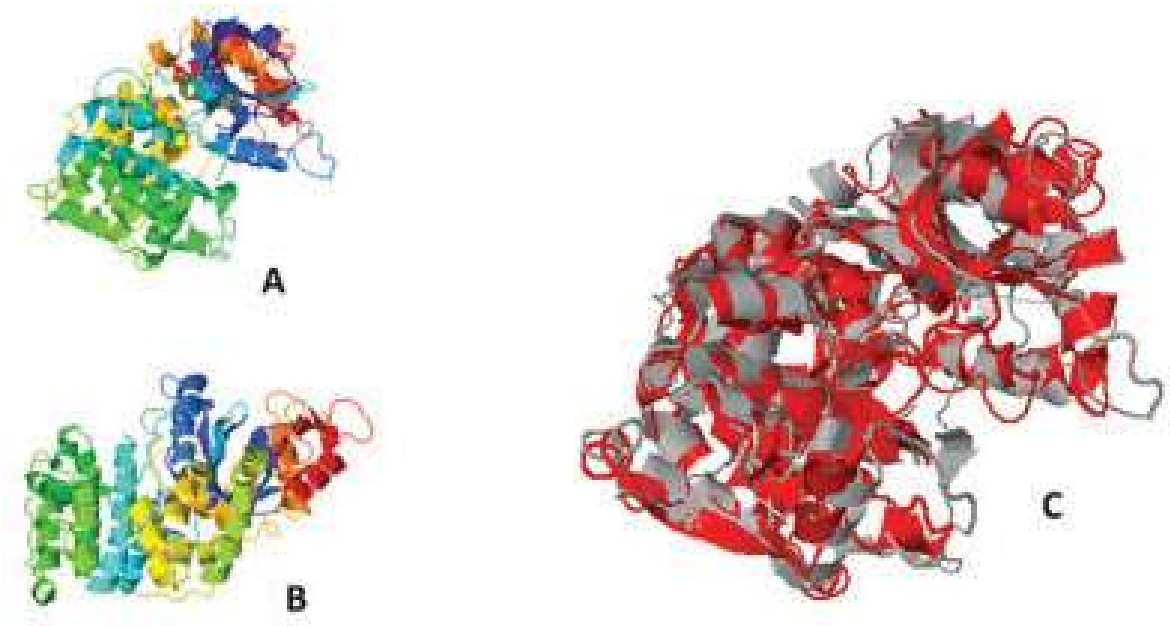

Fig. 4. 3D protein structures modeled using Swisspdbviewer. The structures are rainbow coloured with red at the N-terminal gradually changing colour sequentially to the C-terminal
A-3D protein model for *Arabidopsis* CDPK 21
B - 3D protein model for isolated *Hevea* CDPK
C-Superposition of the two models done by FATCAT software showing the similarity between the two proteins (gray – *Arabidopsis* CDPK21, red – *Hevea* CDPK)

sequence. The maximum alignment score of 76.01 was between *Hevea cdpk* nucleotide sequence and *Arabidopsis cdpk21*. Despite the sequential differences between the two amino acids differences, the modeled protein structure for *Hevea cdpk* shows significant superposition with the modeled structure for *Arabidopsis cdpk21*. These results tend to envisage the orthologous nature of the two genes. Proteins with same domain combination and architecture are likely to have a common ancestor and common functional features (Vogel *et al.,* 2004; Forslund *et al.,* 2011). *Hevea cdpk* may functionally be similar to *Arabidopsis cdpk21* which is activated during hyperosmotic stress (Franz *et al.,* 2011), thus playing a role in multiple abiotic stress response since hyperosmotic stress is closely associated with drought and temperature extremities.

CONCLUSION

In the present study, *cdpk* gene was isolated and characterized from the *H. brasiliensis* clone RRII 105. The isolated promoter region has multiple stress-responsive *cis*-elements. Intron pattern analysis of the genomic sequence helped in classifying the sequence as belonging to the Group II CDPK family. The prediction of post-translational modifications and subcellular localization, *in silico* analysis as well as the similarity of the modeled protein structure to the stress responsive protein, *Arabidopsis* CDPK 21 shows that *Hevea cdpk* may be a potential gene that could play a vital role in multiple abiotic stress signaling. Further studies will be focused on the functional evaluation as well as differential expression of this gene between the stress

susceptible clone RRII 105 and stress tolerant clone RRIM 600 to determine whether this contributes to varied levels of stress tolerance. Developing genetically modified plants is an extensive, expensive and labour-intensive process; therefore selecting the right candidate gene is very important. The present study is a crucial step in identifying potential multiple stress responsive genes for genetically engineering stress tolerant clones and thereby improving natural rubber production.

ACKNOWLEDGEMENT

The first and second authors are grateful to the Council of Scientific and Industrial Research (CSIR) and University Grants Commission (UGC), respectively for research fellowships.

REFERENCES

Asano, T., Tanaka, N., Yang, G., Hayashi, N. and Komatsu, S. (2005). Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: comprehensive analysis of the CDPKs gene family in rice. *Plant Cell Physiology*, **46**(2): 356-366

Ayadi, M., Delphine, M., Denis, F., Verdeil, J-C., Aurore, V., Emmanuel, G. and Masmoudi, K. (2004). Functional analysis of the durum wheat gene TdPIP2 and its promoter region in response to abiotic stress in rice. *Plant Physiology and Biochemistry,* **79**: 98-108.

Benkert, P., Künzli, M. and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Research*, **37**: 322-326.

Bhatnagar-Mathur, P., Vadez, V. and Sharma, K.K. (2008). Transgenic approaches for abiotic stress tolerance in plants: Retrospect and prospects. *Plant Cell Reports*, **27**(3): 411-424.

Boudsocq, M. and Sheen, J. (2013). CDPKs in immune and stress signaling. *Trends in Plant Sciences,* **18***:* 30-40.

Chehab, E.W., Patharkar, O.R., Hegeman, A.D., Taybi, T. and Cushman, J.C. (2004). *Plant Physiology,* **135**: 1-17.

Cheng, S.H., Willmann, M.R., Chen, H.C. and Sheen, J. (2002). Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiology,* **129**(2): 469-485.

Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus,* **12**:13-15.

Forslund, K., Pekkari, I., and Sonnhammer, E.L. (2011). Domain architecture conservation in orthologs. *BMC bioinformatics*, **12**(1): 326-331.

Fox, J. and Castella, J-C. (2013). Expansion of rubber (*Hevea brasiliensis*) in Mainland Southeast Asia: What are the prospects for smallholders? *The Journal of Peasant Studies,* **40**(1): 155-170.

Franz, S., Ehlert, B., Liese, A., Kurth, J., Cazalé, A.C., and Romeis, T. (2011). Calcium- dependent protein kinase CPK21 functions in abiotic stress response in *Arabidopsis thaliana. Molecular Plant,* **4**(1): 83-96.

Harmon, A.C., Gribskov, M., Gubrium, E. and Harper, J.F. (2001). The CDPK superfamily of protein kinases. *New Phytologist*, **151**(1): 175-183.

Kelly, L.A. and Sternberg, M.J.E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols*, **4**(3): 363-371.

Knight, H. and Knight, M.R. (2001). Abiotic stress signalling pathways: specificity and cross- talk. *Trends in Plant Science*, **6**(6): 262-267.

Li, A., Wang, X., Leseberg, C.H., Jia, J. and Mao, L. (2008). Biotic and abiotic stress responses through calcium-dependent protein kinase (CDPK) signaling in wheat (*Triticum aestivum* L.). *Plant Signal Behaviour,* **3**(9): 654-656.

Ludwig, A.A., Romeis, T. and Jones, J.D. (2004). CDPK mediated signalling pathways: specificity and cross-talk. *Journal of Experimental Botany,* **55**(395): 181-188.

Martín, M.L. and Busconi, L. (2000). Membrane localization of a rice calcium dependent protein kinase (CDPK) is mediated by myristoylation

and palmitoylation. *The Plant Journal*, **24**(4): 429-435.

Michel, T., Li, G. and Busconi, L. (1993). Phosphorylation and subcellular translocation of endothelial nitric oxide synthase. *Proceedings of the National Academy of Sciences USA*, **90**: 6252-6255.

Oh, S.K., Kang, H., Shin, D.H., Yang, J. and Han, K-H. (2000). Molecular cloning and characterization of a functional cDNA clone encoding isopentenyl diphosphate isomerase from *Hevea brasiliensis*. *Journal of Plant Physiology,* **157**(5): 549-557.

Parthasarathy, V.K., Chattopadhya, P.K. and Bose, T.K. (2006). Plantation crops - Volume I, Naya Udhyog, Kolkata. 345 p

Podell, S. and Gribskov, M. (2004). Predicting N-terminal myristoylation sites in plant proteins. *BMC Genomics*, **5**(1): 37-43.

Rushton, P.J., Somssich, I.E., Ringler, P. and Shen, Q.J. (2010). WRKY transcription factors. *Trends in Plant Science,* **15**: 247-258.

Saijo, Y., Hata, S., Kyozuka, J., Shimamoto, K. and Izui, K. (2000). Over expression of a single Ca2+ dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *The Plant Journal,* **23**(3): 319-327.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* - Second Edition. Cold Spring Harbour Laboratory Press, New York 11803. 121pp.

Schulz, P., Herde, M. and Romeis, T. (2013). Calcium-Dependent Protein Kinases: Hubs in Plant Stress Signaling and Development. *Plant Physiology,* **163**(2): 523-530.

Takezawa, D., Patil, S., Bhatia, A. and Poovaiah, B.W. (1996). Calcium-dependent protein kinase genes in corn roots. *Journal of Plant Physiology,* **149**(3): 329-335.

Venkatachalam, P., Thanseem, I. and Thulaseedharan, A. (1999). A rapid and efficient method for isolation of RNA from bark tissue of *Hevea brasiliensis*. *Current Science,* **77**: 635-637.

Venkatachalam, P., Jayashree, R., Rekha, K., Sushmakumari, S., Sobha, S., Jayasree, P.K., Kala, R.G. and Thulaseedharan, A. (2007). Rubber Tree (*Hevea brasiliensis* Muell. Arg). In: *Agrobacterium Protocols Volume 2*. Humana Press, India. 153pp.

Vinod, K.K., Meenattoor, J.R., Priyadarshan, P.M., Pothen, J., Chaudhuri, D., Krishnakumar, A.K., Sethuraj, M.R. and Potty, S.N. (1996). Early performance of some clones of *Hevea brasiliensis* in Tripura. *Indian Journal of Natural Rubber Research,* **9**(2): 123-129.

Vogel, C., Bashton, M., Kerrison, N.D., Chothia, C. and Teichmann, S.A. (2004). Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology,* **14**(2): 208-216.

Wedegaertner, P.B., Wilson, P.T. and Bourne, H.R. (1995). Lipid modifications of trimeric G proteins. *Journal of Biological Chemistry,* **270**: 503-506.

Witte, C.P., Keinath, N., Dubiella, U., Demoulière, R., Seal, A. and Romeis, T. (2010). Tobacco calcium-dependent protein kinases are differentially phosphorylated in vivo as part of a kinase cascade that regulates stress response. *Journal of Biological Chemistry,* **285**(13): 9740-9748.

Yalovsky, S., Rodríguez-Concepción, M., and Gruissem, W. (1999). Lipid modifications of proteins slipping in and out of membranes. *Trends in Plant Science,* **4**: 439-445.

Zhang, H., Liu, W.Z., Zhang, Y., Deng, M., Niu, F., Yang, B. and Jiang, Y.Q. (2014). Identification, expression and interaction analyses of calcium-dependent protein kinase (CPK) genes in canola (*Brassica napus* L.). *BMC genomics,* **15**(1): 211.

Zhu, J., Chen, X., Chan, W., Tiang, W. and Zhang, Z. (2010). Molecular characterization of HbCDPK1, an ethephon-induced calcium-dependent protein kinase gene of *Hevea brasiliensis*. *Bioscience Biotechnology and Biochemistry*, **74**(11): 2183-2188.

Zuo, R., Hu, R., Chai, G., Xu, M., Qi, G., Kong, Y. and Zhou, G. (2013). Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (*Populus trichocarpa*). *Molecular Biology Reports*, **40**(3): 2645-62.