Short scientific report *Journal of Plantation Crops, 2010, 38 (2): 144-147*

In silico prediction of function and modelling of WRKY protein in coconut

(Manuscript Received: 30-05-08, Revised: 08-04-10, Accepted: 17-05-10)

Keywords: In silico prediction, molecular modeling, WRKY

The WRKY proteins comprise a super family of eukaryote-specific group of transcription factors belonging to the zinc-finger-type class of proteins and were first isolated from plants (Ishiguro and Nakamura, 1994; Rushton *et al.*, 1995). WRKY factors play significant roles in the regulation of many plant processes, such as pathogen defense, wound response, drought stress, plant growth, embryo formation and senescence (Ulker and Somssich, 2004). The name of the WRKY family itself is derived from the most prominent feature of these proteins, the WRKY domain, constituted by about 60 amino acid residues. In this WRKY domain, a conserved WRKYGQK sequence is followed by a C2H2- or C2HC- type of zinc finger motif (Eulgem *et al.*, 2000).

WRKY proteins have been classified into three major groups based on the number of WRKY domains and on the features of their associated zinc-finger motif. Group I comprises proteins with two WRKY domains and a Cys_2His_2 (or C_2-H_2) zinc- finger motif (more precisely C-X_{4-5}-C-X_{22-23}-H-X_1-H) or $Cys_2HisCys$ (or a C_2-HC) zinc-finger motif (more precisely C-X_7-C-X_{23}-H-X_1-C), where C is Cys, H is His, and X is any amino acid). Group II (the largest group) comprises proteins with one WRKY domain and the same Cys_2His_2 zinc-finger motif as in group 1. Group III comprises proteins with one WRKY domain but a $Cys_2HisCys$ (or a C_2-HC) zinc-finger motif (more specifically C-X_{4-5}-C-X_{22-23}-H-X_1-C or C-X_7-C-X_{23}-H-X_1-C, where C is Cys, H is His, and X is any amino acid) instead of CyS_2HiS_2. Potential *WRKY* target genes have been suggested based on the general binding activity of WRKY factors to their recognized *cis*-element, TGACC/T, or W box (Eulgem *et al.*, 2000).

Structural genomics is expected to elucidate many experimentally determined proteins structures. Since solving a protein structure by NMR or crystallography remains a long and expensive effort, constructing 3-D models based on structures of homologous proteins is an alternative approach. *WRKY* sequences have been isolated from coconut (Mauro-Herrera *et al.*, 2006; Rajesh *et al.*, 2006). The functional assignment through structural exploration of these proteins would definitely have its own place for future reference in coconut studies. The present work attempts to unravel the functions controlled by WRKY proteins in coconut via homology modelling.

Coconut WRKY2 protein sequence (Uniprot Accession number Q00A39) and other sequences were retrieved from the public databases NCBI (www.ncbi.nlm.nih.gov) and EBI (www.ebi.ac.uk). Comparative modeling was based on the homologous from the PDB (www.expasy.org) and SCOP (Structural Classification of Proteins) databases.

Blast searches for the Swissprot /TrEMBL curated databases were done using WU BLAST 2.0 algorithm to make out the local alignments for the coconut WRKY2 protein. Further analysis was performed by the ExPasy server tools (www.expasy.org/tools). Sequence pattern recognition and determination of the modular architecture were done by the Interpro analysis, based on PROSITE (www.expasy.org/prosite) and Pfam databases (www.pfam.sanger.uk). Physicochemical properties of the selected proteins were determined using the PROTPARAM tools (www.expasy.ch/tools/protparam). Hydropathy plots of the deduced proteins were determined using the Kyte and Doolittle values (Kyte and Doolittle, 1982).

The secondary structure analysis was performed using the PELE programme of the SDSC Biology workbench (http://workbench.sdsc.edu/). Intrinsic disorders in the peptide sequences were identified by GLOBPLOT analysis (www.globplot.embl.de) based on Lindings' values (Linding *et al.*, 2003). Further modeling was done with the SWISSPDB modeling software (www.expasy.org/spdbv). Modeled protein was assessed with PROCHECK (www.biochem.ucl.ac.uk/~roman/procheck/procheck). Function assignments were made based on the structural homologues and similar homologues identified for the test protein.

WU BLAST 2.0 analysis of coconut WRKY2 protein revealed the highest sequence identity to WRKY proteins from *Vitis vinifera* (92 %; Uniprot A7PTK2), *Glycine max* (92 %; Uniprot A7LHH2), *Nicotiana tabaccum* (88 %; Uniprot Q9FXS1), and *Arabidopsis thaliana* (87 %; Uniprot O22176). As shown in Fig.1, the conserved WRKYGQK residues in the N-terminal region of coconut WRKY2 protein were identified using the multiple alignment (CLUSTALX). Gene family of transcription factors generally contain a highly conserved domain involved in DNA binding. This remarkable conservation makes WRKY domain binds specifically to the DNA sequence motif (T)(T) TGAC(C/T), which is known as the W box (de Pater *et al.*, 1996; Rushton *et al.*, 1996). The invariant TGAC core of the W box is essential for function and is considered to recognize the specificity of DNA binding process (Yamasaki *et al.*, 2005). The five consecutive residues RKYGQ are distinct components of the WRKY family scaffold (Duan *et al.*, 2007). Another set of conserved residues are noticed in the pattern PRGYYK. These are good candidates for the specific DNA-binding process taking into account the specific DNA sequence that all WRKY members could

Fig. 1. Alignment of sequences similar to coconut WRKY2 protein produced by ClustalX programme. β1-β4 indicate the four β-sheets. Sequence underlined red indicates the WRKY domain. Green line shows the zinc finger motif. Blue line indicates the hydrophobic residues

recognize (Duan *et al.*, 2007). Since the coconut WRKY2 protein contains only a single WRKY domain and a well defined zinc finger binding motif, they could be classified as Group II WRKY proteins.

Hydropathy plot revealed that WRKY2 proteins are highly hydrophilic with over 90% residues falling in the hydrophilic regions with negative scores (Fig. 2). The grand average hydropathicity values of WRKY2 (GRAVY): -1.198 suggests that these proteins are highly hydrated in an aqueous environment. Solution structure of WRKY DNA binding domain demonstrated the hydrophilic transcripts in *Arabidopsis thaliana*.

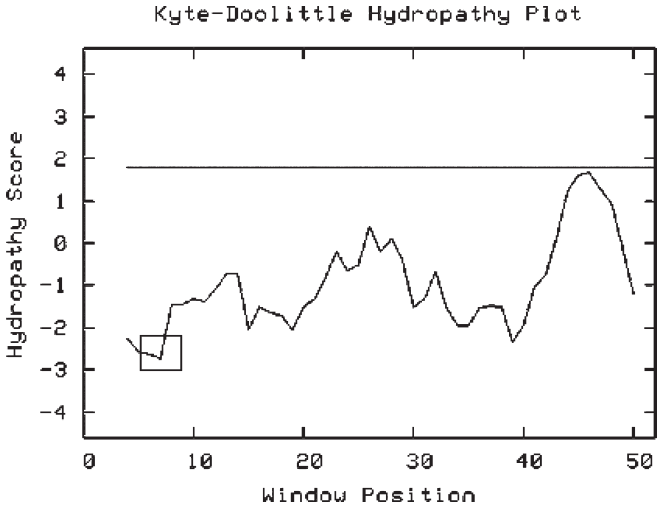

Fig. 2. Hydropathy analysis of predicted coconut WRKY2 protein based on Kyte and Doolittle values, using a seven-residue window. Those values below zero are negative and hydrophilic. Areas corresponding to highly conserved blocks are highlighted in boxes

Family and domain search also revealed the strong possibilities of this protein included in the Pf03106 cluster of WRKY DNA binding domain.

The secondary structure prediction revealed that this particular protein is of β-strands with random coils (Fig. 3). The GLOB PLOT result showed disorder region ranging 7-28 and 44-54 in the plot derived for WRKY2. The WRKYGQK conserved residues to the most N-terminal β-strand is linked in the middle of the sequence by the Gly residue, which enables extensive hydrophobic interactions involving the Trp residue and contributes to the structural stability of the β-sheet (Yamasaki *et al.*, 2005). The numerous basic residues involved in the hydrophobic interactions are likely to be suitable for maintain a single β-sheet structure.

SCOP analysis showed that WRKY2 protein shared homology to sequences that possess β-strands (PDB: 1WJ2, 2AYD). Their highly conserved nature and

```
10        20        30        40        50        60
    ....:.....x.....:.....x.....:.....x.....:.....x.....:.....x
    YQWRKYGQKPIKGSPYPRGYYKCGSLRGCPARKHVERAPDDPSMLIVTYEGEHN
    gi_115520922_gb_ABJ08844.1_ WRKY2 [Cocos nucifera]
    CCCCHCCCCCCCCCCCCCEECEEEECCCCCCCCCCCCCCCCCCCCCEEEEECCCCCBPS
    CCHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCCEEEEECCCCCD_R
    CCCCCCCCCCCCCCCCCCCCEEEECCCCCCCCCCCCCCCCCCCCCEEEEECCCCCDSC
    CCCCCCCCCCCCCCCCCCCCEEEECCCCCCCCCCCCCCCCCCCCCEEEEECCEECGGR
    HEHHHCCCCCCCCCCCCCCCEEEECCCCCCCCCCCCECCCCCCCCEEEEECCCCCGOR
    CCCCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCCCCCCCCCCEEECCCCCCCH_K
    CCCCCCCCCCCCCCCCCCCCEEEECCCCCCCCCCCCCCCCCCCCCEEEEEEEECCEK_S
    CCCCCCCCCCCCCCCCCCCCEEEECCCCCCCCCCCCCCCCCCCCCEEEEECCCCCJOI
```

Fig. 3. Secondary structure analysis of the predicted WRKY2 protein performed with PELE programme. Seven different structure predictions are shown, with the most likely structural feature at each residue indicated by H (α-helices), E (β-sheets) or C (random coils). The programmes used are denoted BPS, D_R, DSC, GGR, GOR, H_K and K_S. The 'winner-takes-all' joint prediction was given by JOI programme

stress induced structural transistion suggests that coconut WRKY2 protein shares similar hypothetical functions.

3-D structure of coconut WRKY2 protein was generated based on the secondary analysis done before and the 3-D structure thus obtained was assessed using PROCHECK. The homology model validated by PROCHECK essentially satisfied the stereo-chemical parameters with well-refined structures at similar resolutions (Morris *et al.*, 1992). The structure consisted of a four stranded, anti-parallel β-sheets with Zn binding pocket formed by conserved Cys/His residues located at one end of the sheet (Fig. 4). One residue in the C-terminal end was missing. WRKY protein's signature sequence, 'WRKYGQK', is in the β1 strand. Solution structure of an *Arabidopsis thaliana* WRKY4 protein revealed that the structure consisted of a four-stranded β-sheet, with a zinc binding pocket formed by conserved Cys/His residues located at one end of the β-sheet. High resolution crystal structure study of *Arabidopsis thaliana* WRKY1 protein revealed that it was mainly composed

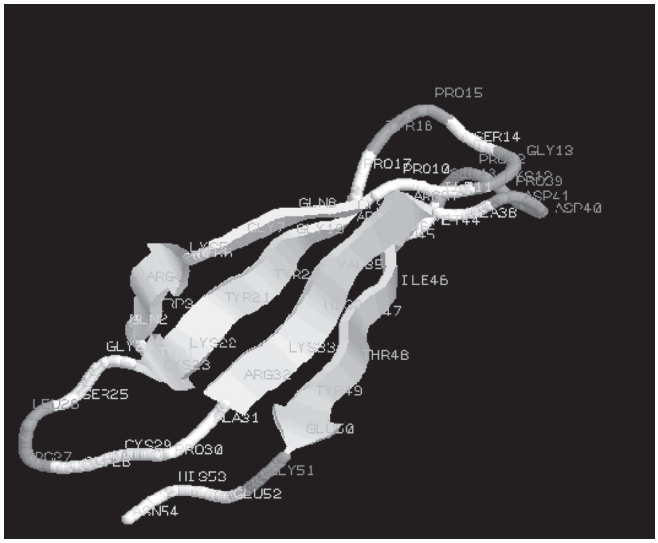

Fig. 4. Predicted 3-D structure of coconut WRKY2 protein generated using Rasmol programme

of a five-stranded anti-parallel β-sheet with the disordered N-terminal and C-terminal three residues missing from the structure (Duan *et al.*, 2007). Maeo *et al.* (2001) had pointed out that the conserved WRKYGQK region located on β2 strand of tobacco WRKY protein was important for DNA binding. Distribution of residues in the most favoured region of the Ramachandran plot for WRKY2 was found to be greater than 90% (Table 1). The statistics of Ramachandran plot signifies that WRKY2 is fairly good protein model (Laskowski *et al.*, 1993).

Table 1. Ramachandran plot statistics of coconut WRKY2 protein

	Scattered residues	
Quadrangular regions of plot	**Number**	**Percentage**
Most favoured regions	40	97.56
Additional allowed regions	1	02.44
Generously allowed region	0	
Disallowed region	0	
		100.00
Non-glycine and proline residues	41	
Proline residues	6	
Glycine residues	6	
Total no. of residues	53	

The coconut WRKY2 fragment protein can be grouped under Group II WRKY proteins based on their possessing a single WRKY domain, a prominent zinc finger motif, extreme hydrophilic and predominant β-strand and random coils of the residues. Function assignment of this protein revealed the DNA binding property due to the highly conserved motif in the N-terminal. The presence of some basic residues reveals the hydrophobic nature and the hydrophilicity suggests the globularity of the protein protecting the hydrophobic core to withstand stress. The homology model of WRKY2 protein generated in this study could provide us with structural information in determining the function of this important class of proteins in coconut.

Acknowledgement

This work was supported by a grant from Department of Biotechnology (BTISnet), New Delhi, India.

References

de Pater, S., Greco, V., Pham, K., Memelink, J. and Kijne, J. 1996. Characterization of a zinc-dependent transcriptional activator from Arabidopsis. *Nucleic Acids Res.* **24**: 4624-4631.

Duan, M.R., Nan, J., Liang, Y.H., Mao, P., Lu, L., Li, L., Wei, C., Lai, L., Li, Y. and Su X.-D. 2007. DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. *Nucleic Acids Res*. **35**: 1145-1154.

Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. 2000. The *WRKY* superfamily of plant transcription factors. *Trends Plant Sci*. **5**: 199-206.

Ishiguro, S. and Nakamura, K. 1994. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato. *Mol. Gen. Genet*. **244**: 563-571.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. 1993. PROCHECK: a program to check the stereo-chemical quality of protein structures. *J. Appl. Cryst.* **26**: 283-291.

Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. 2003. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**: 3701-3708.

Maeo K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. and Nakamura, K. 2001. Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci. Biotechnol. Biochem*. **65**: 2428-2436.

Mauro-Herrera, M., Meerow, A.W., Borrone, J.W., Kuhn, D.N. and Schnell, R.J. 2006. Ten informative markers developed from WRKY sequences in coconut (*Cocos nucifera* L.). *Mol. Ecol. Notes* **6**: 904-6.

Morris, A.S., Thanki, N. and Goodfellow, J.M. 1992. Hydration of amino acid side chains: Dependence on secondary structure. *Protein Eng.,* **5**: 717-728.

Rajesh, M.K., Bharathi, M. and Nagarajan, P. 2006. Isolation and characterization of WRKY genes in coconut (*Cocos nucifera* L.). *J. Plantn. Crops* **34**: 519-28.

Rushton, P.J., Mac Donald, H., Huttly, A.K., Lazarus, C.M. and Hooley, R. 1995. Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy 2 genes. *Plant Mol. Biol*. **29**: 691–702.

Rushton, P.J., Torres, J.T., Parniske, M., Wernert, P., Hahlbrock, K. and Somssich, I.E. 1996. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *Embo J*. **15**: 5690-5700.

Ulker, B. and Somssich, I.E. 2004. WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol*. **7**: 491-498.

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K. and Yokoyama, S. 2005. Solution structure of an *Arabidopsis* WRKY DNA binding domain. *Plant Cell* **17**: 944-56.

Bioinformatics Centre,
Central Plantation Crops Research Institute,
Kasaragod - 671124, Kerala

V. Sreesmitha,
M.K. Rajesh,
Bobby Paul,
Anitha Karun,
M.P. Ramesh Kumar,
George V. Thomas