In silico characterization of WRKY33 TF from Sinapis alba

VINOD KUMAR JANGID1, SHIKHA DIXIT2, LALIT DEV TIWARI3, INDRA SINGH4, ATMAKURI RAMAKRISHNA RAO5 and ANITA GROVER6

ICAR-National Institute for Plant Biotechnology, Pusa, New Delhi 110 012, India

Received: 11 June 2019; Accepted: 05 July 2019

ABSTRACT

The WRKY family of transcription factors modulates the host defense mechanisms in response to various environmental stresses. The role of WRKY33 in plant defense and its crosstalk with defense hormone was well established in Arabidopsis but very few information was noted in Sinapis alba. The present study was carried out in 2017, in which computational approaches to characterize the structural and functional features of SaWRKY33 transcription factor was used. Full length WRKY33 coding sequence (1509 bp) from S. alba has been cloned, sequenced and identified as AtWRKY33 homolog. The expression of SaWRKY33 was scored higher in fungal pathogen challenged and jasmonate-treated samples while lower expression was noticed in salicylate-treated samples. Phylogenetic classification, sequence alignment and MEME-based motif scanning demonstrated the remarkable sequential conservation in the WRKY domains and SaWRKY33 clusters with Crambe abyssinica exhibiting the monophyletic origin and paralelyfie evolution from their wild relatives. STRING data showed SaWRKY33 were interacted with MKS1, MPK3, SIB1, and those are involved in plant defense responses against diverse stress conditions. The homology-based modeling of SaWRKY33 functional WRKY domains showed acceptable Ramachandran statistics and satisfies all the necessary energy parameters. The Hex Docking server-based analysis of DNA-protein interaction showed that WRKY domain binds to the W-box through WRKYGQK along with few conserved amino acid residues in the flanking sequences and zinc finger motifs.

Key words: Docking, Plant defense, Sinapis alba, Transcription factors, WRKY33, W-Box

Throughout the course of their entire life, plants encounter multifarious environmental stresses which affect their growth, development and physico-biological processes and ultimately yield. Majority of plants are invulnerable to different pathogens due to presence of highly complex plant defense mechanisms and intricate plant pathogen interactions (Birkenbihl et al. 2012). Plants perceive external signals via their membrane anchored receptors and these signals then transmitted to the nucleus in the form of transcription factors, and ultimately initiating expression of defense responsive genes (Lippok et al. 2007). Transcription factors bind sequence-specifically to the cis-elements of the target genes and regulate the expression of stress-responsive genes in a highly dynamic manner (Yamasaki et al. 2012).

WRKY TFs family is one of the largest plant-specific transcriptional regulator gene family which is integral parts of various signaling pathways and harmonize various plant processes (Zheng et al. 2006, Agarwal et al. 2011, Mao et al. 2011, Lai et al. 2011). The WRKY TFs comprise about 60 amino acids conserved region, DNA binding domain possessing an invariant heptapeptide WRKYGQK signature sequence and a zinc-finger motif. The WRKY proteins bind to the W-box sequences-specifically ((T)TGACY, where Y is C or T) at the promoter regions of the downstream target genes (Rushton et al. 1996).

The WRKY33 is one of the members performing vital roles in modulating and fine-tuning of plant processes and signaling cross-talks which are involved in plant defense mechanism against diverse biotic and abiotic stresses (Zheng et al. 2006, Birkenbihl et al. 2012). Birkenbihl et al. (2012) experimentally confirmed that AtWRKY33 plays as a positive regulator in plant defense against necrotrophic pathogens such as Botrytis cinerea and Alternaria brassicicola along with induced expression of the JA-responsive plant defensin PDF1.2 gene. In this study, our emphasis is on detailed characterization of cloned Sinapis alba WRKY33 gene by deduced amino acid sequence analysis, expression study, in silico analysis including domain prediction, protein interactive network analysis and homology-based docking.
of WRKY33 protein to the DNA W-box.

MATERIALS AND METHODS

Plant materials and treatments: In the present study, white mustard (Sinapis alba L.) plants were grown in the growth chamber under suitable growth conditions (22°C, 16/8 h light/dark photoperiod) in 2017. Alternaria brassicaceae fungal pathogen was collected from the field and cultured on oat meal agar medium plates and kept for incubation in BOD for 10 days at 26°C temperature. 2 mM salicylic acid and 100 μM Methyl Jasmonate working solution was sprayed on four-week-old plants separately. For the disease assay, fungal spore suspension (5×10⁵ spores/ml of distilled water) was inoculated directly onto leaf surface and sterile distilled water as control treatment. Leaf samples were harvested from treated and control plants at different time intervals and stored in -80°C till further use.

Cloning and expression profiling of SaWRKY33 gene: Total RNA was extracted from 100 mg of leaf tissues using TRIzol® reagent as per manufacturer’s protocol. 1 μg of total RNA was reverse transcribed into cDNA using oligo(dT) primers with the help of SuperScript III First-Strand synthesis kit. Full-length SaWRKY33 coding sequence was PCR amplified using gene specific primer sets (W33_F - CTATATGATGCTGTCCTTC; W33_R - CTTTATCTCCAGACAAGACG) and the purified insert was cloned into pGEM®-T Easy vector. Colony PCR was conducted for screening positive colonies and the presence of insert was confirmed by EcoRI digestion of the plasmid DNA. For transcript profiling, semi-quantitative PCR was set with optimized parameters for individual target genes. For sqPCR, 20 μL reaction volume containing 1X Taq Buffer, 0.2 Units DNA Taq polymerase, 0.2 mM dNTP, 10 picomole of each primer set (qW33_F - TCAGATGCTGCACAACAACA; qW33_R - AACGATCGAAAAAACGAGGA) and PCR reactions were following as 94°C for 3 min, 30 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 35 sec and final extension at 72°C for 5 min. The amplification products were checked on 1% agarose gel and α-tubulin was used as internal control (Tub_F - CTTTATGTATGGCTGCTTCTTCC; W33_R - TGGACATCGTCTTGTTCTTCC). PCRs were conducted for multiple sequence alignments and unrooted phylogenetic tree was constructed using MEGA7 software (http://megasoftware.net/) following Neighbour-Joining method and 1500 bootstrap replicates. The functional WRKY domains were identified within the WRKY33 protein using ExPASy-Prosite scan (http://prosite.expasy.org/scanprosite/). MEME Suite 4.12.0 (http://meme-suite.org/tools/meme) was used to discover the conserved motifs present within the WRKY33 amino acid sequences with optimum motif width between 6-40 residues and maximum number of motifs up to 8. The STRING version 11.0 servers (http://string-db.org/) was employed for predicting the SaWRKY33 protein interaction network with selecting maximum 10 interactors for first shell of interactions and 5 for second shell of interactions. The prediction was based on Arabidopsis datasets due to unavailability of Sinapis datasets and interactive scores were evaluated at high confidence level (0.700).

Structural modeling of WRKY domains and their interaction with DNA W-box: Based on homology approach, WRKY domains of SaWRKY33 protein was modeled using Discovery Studio 4.1. Using BLASTp tool against Protein Data Bank (PDB; http://www.rcsb.org/pdb/), the searched template 2AYD (CTD of AtWRKY1) was found to be suitable for protein modeling and 2LEX (complex of CTD of AtWRKY4 and W-box DNA) for docking purpose with >65% sequence homology. Using PROCHECK module of PDBSum server (http://www.ebi.ac.uk/pdbsum/), models were validated and evaluated for stereo-chemically quality by Ramachandran plot. The stabilities of models were further improved by energy minimization employing implementation of CHARMM force field. HEX 6.1 standalone tool was employed for study of molecular interaction between WRKY domains of SaWRKY33 with DNA W-box using 2LEX as a template and Shape + Electro parameters. Further, DS Studio 4.1 was used for visualization and interaction analysis of the docked complexes.

RESULTS AND DISCUSSION

Molecular cloning and insight into SaWRKY33 sequence: The response of WRKY33 gene in plant defense against diverse stresses has been unravelled in model plants such as Arabidopsis (Lippok et al. 2007), but scanty information has been found in S. alba. Considering this, full length SaWRKY33 coding sequence was cloned and sequence characterized. The cloned SaWRKY33 sequence was confirmed through colony PCR and EcoRI restriction digestion by releasing of 1509 bp insert (Fig S1A-B) and sequence was submitted to GenBank database with accession number MG001179.1 with 75.09% sequence similarity with AtWRKY33. Our ProtParam results revealed that SaWRKY33 encodes a protein of 502 amino acid residues with molecular mass of 55.56 KDa, C₃₅₆₅H₃₅₂O₁₅₈₅.N₇₂ atomic composition and predicted pI of 8.10 (Table 1). The SaWRKY33 protein was shown to have hydrophilic and soluble nature based on low Grand Average Hydropathicity (GRAVY) score -1.030 and unstable nature with a half-life of less than 5 h and exhibiting instability index of 56.37
Multiple sequence alignment and phylogenetic analysis: The maximum conservation of amino acid residues was found in both C-terminal (CTD, Fig 1) and N-terminal (NTD, Fig S2) WRKY domains of the SaWRKY33 protein from various plant species with few amino acid substitutions of very similar or different chemical properties. Within the functional WRKY domains, least disturbed and most conserved core residues elucidate their evolutionary significant role in phylogenetic origin and their vital role in defense and other plant physiological activities which mediated through their interaction with specific ligand molecules for triggering responses to various environmental stresses (Brand et al. 2013). Phylogenetic analysis results showed that SaWRKY33 (AVK51629.1) exhibiting the close relationship with Crambe abyssinica (AVK51635.1) with monophyletic origin while paraphyletic origin with other mustard family members like AtWRKY33 (NP_181381.2) and BjWRKY33 (AVK51639.1) (Fig 2A). The list of plant species with their accession number and family name is given in Table S1.

Expression analysis of SaWRKY33 genes under biotic stresses: The AtWRKY33 gene showed differential expression patterns to diverse stresses and variety of hormonal treatments which were analyzed based on microarray and mutant study (Sham et al. 2017). Our sqPCR results revealed that SaWRKY33 gene displaying higher expression in fungal pathogen challenged and JA treated leaf samples while comparatively lower expression in SA treated samples (Fig S1C). Birkenbihl et al. (2012) reported higher transcript levels in fungal pathogen challenged Arabidopsis leaf samples about 6 folds at 8 hpi and that was further increased up to 48 hpi during infection.

Table 1 Physico-chemical properties of SaWRKY33 protein

| Organism/Sequence | No. of AAs | Molecular weight | pI | +ve AAs | -ve AAs | EC | li | GRAVY* | Most abundant amino acid (%) | Least abundant amino acid (%) |
|-------------------|------------|------------------|----|---------|---------|----|----|--------|-----------------------------|------------------------------|
| SaWRKY33          | 502        | 55565.16         | 8.10 | 54      | 52      | 51340 | 56.37 | 43.90 | -1.030 | Serine (15.5 %) | Trp (1.0 %), Cys (1.2 %), Met (1.4 %) |

(Table 1). Agarwal et al. (2011) correlated the protein thermostability with positive and high aliphatic index (Ai) which may be showing responses to diverse biotic and abiotic stresses. In our result, the Ai of SaWRKY33 protein was found to be 43.90 (Table 1). Mao et al. (2011) demonstrated the potential MAPK phosphorylation site was serine at N-terminus of WRKY33 protein in response to pathogen invasion. From the results of SaWRKY33 amino acid composition, tryptophan, cysteine and methionine were the least abundant amino acids while serine was the most abundant (15.5%) (Table 1). CELLO v.2.5 server predicted SaWRKY33 protein localized in nucleus with a highest reliability index of 4.287.

Expression analysis of SaWRKY33 genes under biotic stresses: The AtWRKY33 gene showed differential expression patterns to diverse stresses and variety of hormonal treatments which were analyzed based on microarray and mutant study (Sham et al. 2017). Our sqPCR results revealed that SaWRKY33 gene displaying higher expression in fungal pathogen challenged and JA treated leaf samples while comparatively lower expression in SA treated samples (Fig S1C). Birkenbihl et al. (2012) reported higher transcript levels in fungal pathogen challenged Arabidopsis leaf samples about 6 folds at 8 hpi and that was further increased up to 48 hpi during infection.

Prediction and analysis of functional domains and motifs: At least one WRKY DBD region which encompassing the highly conserved WRKYGQK signature sequence along with zinc finger motif possessed by the WRKY TF superfamily members which in turn participate in sequence specific binding to the cis-elements (Yamasaki et al. 2012). We identified two highly conserved and functional sites within the WRKY DBDs based on ExPASy-Prosite scan of the SaWRKY33 amino acid sequences. The NTD of SaWRKY33 stretches from Arg172 to Pro236 residues with the core signature sequences existed from Glu178 to Pro236 whereas the CTD stretches from Ile326 to Pro397 having signature sequences from Ser332 to Pro397 (Fig S3). Hence, SaWRKY33 protein can be classified into group I of WRKY gene superfamily based on two WRKY DBDs and C-X4-C-X22-23-H-X-H-type zinc finger motifs.
The MEME motif search results revealed the more uniformly distribution of conserved motifs across the diverse plant members which participate in forming functional WRKY domains. Our MEME results clearly demonstrated that two fundamental motifs are present in all the plant members which encompasses in the functional WRKY signature sequences and they revealing functional and structural conservation (Fig 2A). Due to substitution of one or more amino acid residues, plant members showed variation in motif numbers which reflect the sequence divergence and forming separate clusters. The closely related plant members showing more similar patterns of motif distribution. The C-terminal of SaWRKY33 represented by motif 1 (p-value 1.9e–51; Fig. 2B) and N-terminal by Motif 2 (P-value 4.9e–53; Fig 2C) in which motif 1 shows least P-values, therefore C-terminal exhibited more conserved residues.

**Protein interactive network study:** Since, SaWRKY33 have been shown great extent of conservation with AtWRKY33, therefore, they participating most of common protein interacting partners in their functional associative protein networks. Mao et al. (2011) reported two pathogen-responsive MAPks (MPK3 and MPK6), which are involved in induction of the antimicrobial compound camalex in biosynthesis via WRKY33 phosphorylation and regulation of PAD3 gene transcription. Our STRING-based results predicted various protein partners of SaWRKY33 which are involved in diverse biological functions and forming protein interactive networks (Fig 3). The interaction of SaWRKY33 protein and MKS1 (MAP Kinase Substrate 1) was predicted with maximum interactive score of 0.997 (Fig 3; Table S2). MKS1 mainly participates in plant defense regulation as it functions like MPK4 adaptor protein which in turn affects WRKY protein activities (Andreasson et al. 2005). Other predicted protein partners which include MPK3, MPK4, SIB1 protein (Sigma factor binding protein 1), SIB2, ACS6 (ACC synthase 6), STZ (Cys2/His2-type zinc-finger proteins), SZF1 (Zinc finger CCCH domain-containing protein 47), MYB51 and ERF6 (Ethylene responsive element binding factor 6) in the first shell of interaction and MKK2, MKK6, MEK1, MKP2 and VIP1 proteins in the second shell of interaction. Majority of the interacting partners for SaWRKY33 protein belong to the plant defense-related proteins which are involved in plant defense activities against biotic and abiotic stresses. It was demonstrated that the DNA-binding activity of WRKY33 is enhanced through specific recognition and modulation of the WRKY CTD mainly by three VQ proteins, viz. VQ23 (SIB1), VQ16 (SIB2) and VQ21 (MKS1) (Lai et al. 2011, Cheng et al. 2012). Li et al. (2012) revealed the role of MPK3 and MPK6 in the regulation of ACS2 and ACS6 gene expression and their protein stability post-translationally via direct phosphorylation of WRKY33 and its binding to the promoter regions. Under osmotic stress conditions, MPK3 and MPK6 phosphorylate ERF6 which in turn activates the expression of the stress-related transcription factor genes such as STZ, MYB51 and WRKY33 (Dubois et al. 2013).

**Homology modeling of WRKY domains and docking analysis of SaWRKY33:** To unravel interaction patterns of WRKY domains with other proteins and DNA W-box, WRKY domains were modeled using homology modeling-based approach. Best 3D models were selected based on associated electrostatic energies, percent coverage and their stereo quality which are very crucial for determining protein structure stability and model reliability. Our validated results showed that 99.1% amino acid residues were found in most favored regions, while 0.9% in allowed regions in Ramachandran plot which shows quality and reliability of
protein domains (Fig S4A-B).

WRKY TFs have stereotypic preferential binding with DNA cis-regulatory element, W-box of the targeted genes and regulate the gene expression spatio-temporally at transcriptional level (Brand et al. 2013, Ciolkowski et al. 2008). The involvement of WRKYQGK sequence in the DNA binding was investigated using mutational experiments and identified very crucial amino acid residues for DNA binding. Those are particularly Trp, Tyr, and two Lys residues which are playing significant role in maintaining and stabilizing of correct structure of DNA-protein complex (Maeo et al. 2001, Yamasaki et al. 2005, Yamasaki et al. 2012). Our docking results of SaWRKY33 protein over DNA W-box substantiated that Arg\(^{147}\), Lys\(^{176}\), Glu\(^{178}\), Arg\(^{216}\), Thr\(^{223}\), Glu\(^{224}\), His\(^{231}\) and Lys\(^{235}\) residues of NTD are mainly involved in the interaction with T\(^{5}\), G\(^{3}\), A\(^{9}\), G\(^{22}\), T\(^{24}\) and C\(^{52}\) DNA bases (Fig S4D). Whereas, Arg\(^{225}\), Val\(^{227}\), Glu\(^{329}\), His\(^{331}\), Glu\(^{353}\), Ser\(^{370}\) and Glu\(^{379}\) residues of C-terminal are involved in interaction with T\(^{4}\), G\(^{6}\), A\(^{26}\), A\(^{28}\), G\(^{30}\) and G\(^{46}\) DNA bases (Fig S4E). The lysine residue in the conserved WRKYQGK motif favors contact with negatively charged DNA phosphate backbone, whereas glutamine residue favors the binding with nucleotide bases due to its partial negative charge (Brand et al. 2013). The two WRKY domains from group I exhibited different roles in DNA-binding process as the CTD play a crucial role in binding to DNA W-box, whereas increasing the binding affinity and specificity to target gene as determined by NTD (Ciolkowski et al. 2008).

Using computational approaches for the structural and functional interpretation of WRKY TFs provides clear understanding about sequence specific attributes which show functional conservation within members that regulates transcriptional reprogramming of stress responsive genes under environmental stress conditions. This study would provide information regarding their evolutionary divergence, ancestral origins and phylogenetic relationships within family members and even distant members, thus the possible biological roles of gene family could be elucidated. The computational approaches along with mutant and gene complementation techniques would be helpful in functional characterization of individual gene. Against various environmental stresses including biotic and abiotic stresses, these studies may provide some positive perspective for Brassica breeding program.

ACKNOWLEDGMENTS

The authors are grateful to Department of Biotechnology, Ministry of Science and Technology, Govt. of India, for providing the financial assistance in form of DBT-JRF (DBT/2014/1ARI/234).

REFERENCES

Agarwal P, Reddy M P and Chikara J. 2011. WRKY: its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Molecular Biology Reports* **38**(6): 3883–96.

Andreasson E, Jenkins T, Brodersen P, Thorgrimsen S, Petersen N H T and Zhu S. 2005. The MAP kinase substrate MKS1 is a regulator of plant defense responses. *The EMBO Journal* **24**: 2579–89.

Birkenbihl R P, Diezel C and Somssich I E. 2012. *Arabidopsis* WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection. *Plant Physiology* **159**: 266–85.

Brand L H, Fischer N M, Harter K, Kohlbacher O and Wanke D. 2013. Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and *in vitro* binding assays. *Nucleic Acids Research* **41**(21): 9764–78.

Cheng Y, Zhou Y, Yang Y, Chi Y-J, Zhou J and Chen J-Y. 2012. Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors. *Plant Physiology* **159**: 810–25.

Ciołkowski I, Wanke D, Birkenbihl R P and Somssich I E. 2008. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Molecular Biology* **68**: 81–92.

Dubois M, Skirycz A, Claeyes H, Maleux K, Dhandt S and Bodt S D. 2013. ETHYLENE RESPONSE FACTOR6 acts as a central regulator of leaf growth under water-limiting conditions in *Arabidopsis*. *Plant Physiology* **162**: 319–32.

Lai Z, Li Y, Wang F, Cheng Y, Fan B and Yu Q J. 2011. *Arabidopsis* sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *The Plant Cell* **23**: 3824–41.

Li G, Meng X, Wang R, Mao G, Han L and Liu Y. 2012. Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in *Arabidopsis*. *PLoS Genetics* **8**(6): e1002767.

Lippok B, Birkenbihl R P, Rivory G, Brümmer J, Schmelzer E and Logemann E. 2007. Expression of *AtWRKY33* encoding a pathogen- or PAMP-responsive WRKY transcription factor is regulated by a composite DNA motif containing W box elements. *Molecular Plant- Microbe Interactions* **20**(4): 420–29.

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. *Bioscience, Biotechnology, and Biochemistry* **65**(11): 2428–36.

Mao G, Meng X, Liu Y, Zheng Z, Chen Z and Zhang S. 2011.Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. *Plant Cell* **23**: 1639–53.

Rushon 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 PRL1 genes. *EMBO Journal* **15**: 5690–700.

Sham A, Moustafa K, Al-Shamsi S, Alying S, Iratni R and AbuQamar S. 2017. Microarray analysis of Arabidopsis *WRKY33* mutants in response to the necrotrophic fungus *Botrytis cinerea*. *PLoS ONE* **12**(2): e0172343.

Yamasaki K, Kigawa T, Watanabe S, Inoue M, Yamasaki T and Seki M. 2012. Structural basis for sequence-specific DNA recognition by an *Arabidopsis* WRKY transcription factor. *Journal of Biological Chemistry* **287**(10): 7683–91.

Zheng Z, Qamar S A, Chen Z and Mengiste T. 2006. *Arabidopsis* WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant Journal* **48**: 592–605.